EUROPEAN JOURNAL OF STEM EDUCATION

European Journal of STEM Education,
2026, 11(1), Article 39
eISSN: 2468-4368

lectito

Research paper

Utilizing Customized ChatGPT-Based Chatbots to Support Preservice Teachers' Understanding of Collective Mathematical Argumentation

Yuling Zhuang [1*], Xiangquan Yao[2], Trina J. Davis[1], Wisdom Y. Nudze[1]

[1] *Texas A&M University*, UNITED STATES
[2] *The Pennsylvania State University,* UNITED STATES

***Corresponding Author:** ylzhuang@tamu.edu

This paper is part of the special issue: AI and STEM: Exploring the Impact of Artificial Intelligence on Science, Technology, Engineering, and Mathematics Education

Citation: Zhuang, Y., Yao, X., Davis, T. J., & Nudze, W. Y. (2026). Utilizing customized ChatGPT-based chatbots to support preservice teachers' understanding of collective mathematical argumentation. *European Journal of STEM Education, 11*(1), Article 39. https://doi.org/10.20897/ejsteme/18918

Published: July 5, 2026

ABSTRACT

This study investigated how ChatGPT-based chatbots can support preservice mathematics teachers' (PMTs) understanding of core teaching practices, specifically collective mathematical argumentation. Drawing on a situated learning perspective and principles of practice-based teacher education, we focused on the first two phases of the Generative Role-play AI Simulation for Pedagogy (GRASP) model, in which we designed customized AI chatbots to function as both knowledge builders and situation generators. Ten PMTs enrolled in a mathematics education course at a U.S. research university engaged with these chatbots to learn foundational ideas of collective argumentation, analyze argument structure, and examine the productivity of the provided classroom scenarios. Data sources included written assignments, reflections, and pre- and post-surveys. Analyses across multiple data sources indicated that most PMTs in this study developed a robust understanding of collective mathematical argumentation and of the teacher moves that facilitate it through their interactions with the customized AI chatbots, while also demonstrating awareness of the chatbots' usefulness and inherent limitations. These findings highlight both the promise and the constraints of integrating AI tools into teacher preparation and point to design considerations for creating effective AI-supported learning environments.

Keywords: collective mathematical argumentation, generative artificial intelligence, preservice mathematics teachers, AI-based role-play simulation, mathematics teacher education

Mathematics standards and policy documents from various countries (e.g., Australia, New Zealand, the United Kingdom, and the United States) emphasize collective mathematical argumentation — where teachers and students collaboratively share ideas, construct and critique arguments, and make mathematical reasoning visible — as a critical instructional practice across all grade levels to support meaningful mathematical learning (Australian Curriculum, Assessment and Reporting Authority [ACARA], n.d.; Department for Education, 2014; Ministry of Education, 2007; National Council of Teachers of Mathematics [NCTM], 2000). Research indicates that students' participation in collective argumentation in mathematics classrooms supports their conceptual understanding (Wood, 1999), mathematical autonomy (Yackel & Cobb, 1996), and communicative competencies (Andriessen, 2006) as well as improvements in their mathematics achievement (Cross, 2009). While collective argumentation offers multiple benefits for students, it is also an important instructional tool for teachers. Through argumentation, teachers can assess and respond to students' thinking (Krummheuer, 1995), support the construction of new mathematical concepts (Yackel, 2002), and create opportunities for students to value and compare different forms of mathematical explanations (Wood, 1999).

Despite its importance for mathematics teaching and learning and national policy recommendations, argumentation remains uncommon in U.S. classrooms, and multiple studies document significant challenges for teachers in implementing it, especially for novice teachers (Wagner et al., 2014). Moreover, prior research (Bieda, 2010; Kosko et al., 2014) shows that teachers' interpretations of what it means to facilitate collective argumentation frequently diverge from the visions articulated by researchers and teacher educators. Teaching argumentation is complex and pedagogically demanding: teachers must themselves be proficient in argumentation, yet there are limited opportunities for them to learn and practice it in teacher preparation programs (Wagner et al., 2014).

Advances in generative artificial intelligence (GenAI), particularly large language models such as ChatGPT, offer new opportunities to address these persistent challenges. Recent research demonstrates that GenAI-powered simulations can support preservice teachers in rehearsing instructional practices and generating lesson planning ideas, while also requiring careful evaluation of AI-produced content and the ability to craft effective written prompts, skills that preservice teachers frequently find challenging (e.g., Gurl et al., 2024; Zhuang & Zhang, 2025a). Despite these emerging possibilities, limited research has examined how GenAI tools can support PMTs in developing a meaningful understanding of pedagogical practices. To address this gap, the present study investigated the potential of ChatGPT to support PMTs' understanding of collective mathematical argumentation by developing a Generative Role-play AI Simulation for Pedagogy (GRASP) model. The following research questions guided this study:

RQ1: How does the implementation of the GRASP model support PMTs' understanding of collective mathematical argumentation?

RQ2: What are PMTs' perceptions of the strengths and limitations of using AI chatbots to learn to support collective mathematical argumentation?

BACKGROUND AND RELATED LITERATURE

Teacher support for collective argumentation

Argumentation is recognized as one of the eight Standards for Mathematical Practice in the Common Core State Standards, which identify "constructing viable arguments and critiquing the reasoning of others" as a core expectation for all learners (National Governors Association Center for Best Practices [NGA] & Council of Chief State School Officers [CCSSO], 2010). In mathematics classrooms, teachers play a central role in creating opportunities for students to engage in argumentation, as they serve as representatives of the mathematical community and guide how ideas are introduced, justified, and negotiated (Yackel & Cobb, 1996). Since Krummheuer's (1995) application of Toulmin's model to classroom interactions, many scholars have used this framework to analyze teachers' contributions to argumentation, including how they prompt, extend, or refine students' justifications (e.g., Conner et al., 2014; Zhuang & Conner, 2022).

An argument, as described by Toulmin (1958/2003), comprises three core elements: claim (i.e., the statement whose validity is being determined), data (i.e., the foundations on which the claim is based), and warrants (i.e., the inference rule connecting the data to the claim). An argument may also include other argument components such as rebuttals (i.e., statements describing circumstances under which the warrant would not be valid), qualifiers (i.e., statements describing the certainty with which a claim is made), and backings (i.e., statements describing the relevance of the warrant). Studies have found that introducing Toulmin's model to PMTs helps them develop a clearer, more nuanced understanding of collective argumentation by offering a structured framework for identifying and analyzing the components of mathematical arguments, thereby enabling them to recognize better, facilitate, and reflect on classroom-based argumentation (Wagner et al., 2014).

Given the pedagogical demands of facilitating collective argumentation, teachers benefit from understanding how mathematical arguments are structured and how to support students' participation in argumentation (Conner et al., 2014; Yackel, 2002). However, research shows that PMTs often hold narrow or incomplete interpretations of argumentation, viewing it primarily as disagreement or debate (e.g., Wagner et al., 2014) or as students simply explaining or defending answers (Staples & Newton, 2016). These limited understandings, widely documented in prior studies, underscore the need to provide PMTs with opportunities to develop a more robust understanding of argumentation and the teacher's role in supporting it (Francisco, 2025; Yackel, 2002).

Use of GenAI in mathematics teacher education

Recent research on the integration of GenAI in mathematics teacher education has progressed from introductory explorations of perceptions and potential to more nuanced investigations of practices, pedagogical design, and professional knowledge development (Almarashdi et al., 2026). Studies have demonstrated that GenAI tools such as ChatGPT can support PMTs in lesson planning and mathematical task generation, while also underscoring the necessity of teacher judgment and critical evaluation of AI-generated outputs due to limitations in mathematical accuracy (Gurl et al., 2024; Kim et al., 2025). Studies also show the potential of using ChatGPT to support PMTs' engagement with pedagogical content knowledge, particularly through activities such as problem posing, adapting instructional tasks, and examining the quality of mathematical explanations generated by AI (Biton & Segal, 2025; Walkington, 2025). Findings consistently indicate that PMTs perceive GenAI as useful for supporting instructional preparation, especially by saving time, generating ideas, and prompting consideration of alternative instructional approaches (Uğraş et al., 2024; Zhuang & Zhang, 2025a). Despite these advances, existing research has primarily emphasized content knowledge and instructional preparation, with limited studies examining how GenAI supports mathematical teaching in practice. Even with those studies that examined how GenAI supports core teaching practices (e.g., Lee & Yeo, 2022), they primarily focused on teacher questioning strategies (e.g., Zhuang & Zhang, 2025b). As a result, we still know relatively little about how GenAI may support teachers' development of broader pedagogical knowledge and practices (e.g., facilitating meaningful mathematical discourse).

One promising approach to extend our understanding of how AI can support teachers' professional learning involves simulation-based learning environments that provide preservice teachers with structured opportunities to learn and enact core teaching practices. The potential of such environments to provide collaborative, reflective, and skill-development opportunities for preservice teachers' engagement is well documented in the review literature (Ledger & Fischetti, 2020). The promise of simulation-based experiential learning experiences includes providing accessible, safe spaces for preservice teachers to practice core teaching skills that may not be readily replicable in other learning contexts (Davis et al., 2022). Moreover, scholars maintain that early simulation-based digital tools have allowed for limited customization and may not fully reflect authentic educational settings (Son et al., 2024). In recent years, GenAI tools like ChatGPT have shown promise in supporting simulation-based and practice-based learning experiences for preservice teachers (Lee & Yeo, 2022). Since the public release of ChatGPT in November 2022, its adoption and impact have accelerated significantly, transforming how educators and students interact with content and with each other, and more specifically, how interventions can impact opportunities in teacher education (Lee et al., 2024). For example, using customized chatbots, Son et al. (2024) examined how preservice teachers engaged in simulations focused on responsive teaching practices. Chatbots were designed to serve dual roles of both a virtual student and a virtual mentor. They examined changes in preservice teachers' questioning patterns and found that their tendency to adjust varied.

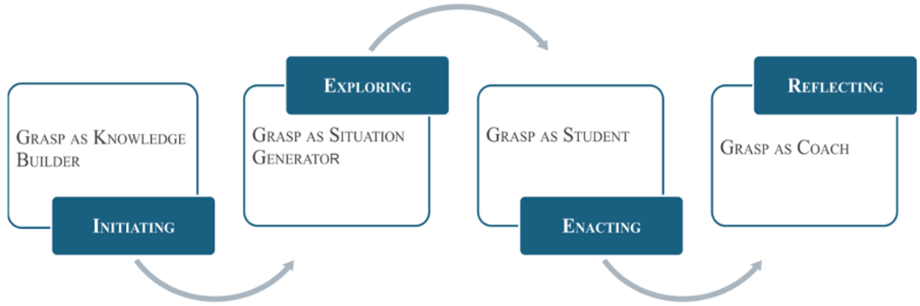
In our previous work, we developed a ChatGPT-powered chatbot for role-play simulations in which PMTs acted as teachers while ChatGPT simulated struggling middle school students, supporting their understanding of common mathematical misconceptions and providing opportunities to practice responsive teaching skills (Zhuang & Zhang, 2025b). Participants with limited teaching experience reported that engaging with AI-simulated students offered a low-risk way to rehearse authentic classroom exchanges, describing the experience as "a good way to get practice, the closest you can get to a student" (Zhuang & Zhang, 2025a). Building on this work, we extended ChatGPT's role in the current project so that it can also serve as a knowledge builder and scenario generator, providing interactive, research-informed, and practice-based learning experiences to support PMTs in developing a deep understanding of collective argumentation.

DESIGN A GENERATIVE ROLE-PLAY AI SIMULATION FOR PEDAGOGY MODEL

Grounded in a situated perspective on learning, the development of the GRASP model was informed by the view that learning occurs through engagement in authentic activities and contexts (Lave & Wenger, 1991). It was also informed by prior research on practice-based teacher education, core teaching practices, and high-leverage mathematics practices (Grossman et al., 2009; Lampert et al., 2013), as well as the importance of bridging the gap between theoretical learning and practical application in teacher education (Ensor, 2001). Through a structured four-phase learning cycle, the GRASP model (**Figure 1**) aimed to create a dynamic and interactive environment in which PMTs engage in guided trial-and-error experiences to develop and refine their understanding of core teaching practices. In addition, GenAI-based simulations offered personalized training and feedback tailored to individual learning needs, providing a flexible and low-risk space for PMTs to experiment, rehearse, and strengthen their pedagogical content knowledge.

Figure 1

Generative role-play AI simulation for pedagogy (GRASP) model



The Initiating phase focused on supporting PMTs in developing a foundational conceptual understanding of the targeted teaching practice. During this phase, PMTs engaged with the GRASP model which served as a knowledge builder, designed to introduce key concepts, pedagogical frameworks, and illustrative examples. The GRASP model scaffolded early sense-making through guided inquiry and responsive explanation. From a situated learning perspective, this phase supported PMTs in developing an initial understanding of instructional practice by providing representations of how the specific teaching practice unfolds within simulated teaching contexts. The Exploring phase emphasized developing PMTs' skills in analyzing and interpreting teaching practices within simulated classroom contexts. During this phase, the GRASP model functioned as a situation generator, presenting GenAI-generated classroom scenarios for PMTs to examine and to identify effective instructional moves, interpret student thinking, and notice opportunities to strengthen engagement and learning. This phase aligns with situated learning theory's emphasis on the idea that learning is inherently embedded in the contexts in which knowledge is used. We argue that these two phases are most powerful when combined, as the foundational knowledge developed in the Initiating phase can be contextualized through classroom examples examined in Exploring, thereby supporting PMTs in making stronger connections between pedagogical concepts and their enactment in simulated instruction activities, thereby fostering deeper understanding. Grounded in situated learning theory, the Enacting phase emphasizes PMTs' development of pedagogical content knowledge through participation in simulated interactive teaching scenarios. The Enacting phase supported PMTs in practicing teaching in realistic simulations. In this phase, PMTs took on the role of the teacher while GRASP simulated diverse learners who reasoned, questioned, and sometimes struggled. These role-play interactions enabled PMTs to rehearse instructional moves, respond to emerging student ideas, and make pedagogical decisions in real time. The Reflecting phase focused on refining teaching skills through targeted feedback and structured reflection. GRASP acted as a coaching agent by analyzing the PMTs' instructional moves, offering individualized feedback grounded in established teacher-support frameworks, and highlighting alternative strategies that could strengthen future enactments. These two phases typically unfold together within a single chat session: PMTs first engage in simulated teaching during enactment, then immediately receive a reflection report that analyzes their instructional moves and provides feedback on their teaching practice. The joining of the Enacting and Reflecting phases mirrors authentic professional learning cycles, in which participation in practice and sensemaking about that participation occur in interaction and mutually reinforce PMTs' development of practice-based expertise (Grossman et al., 2009; McDonald et al., 2013).

This paper examines how we support PMTs' understanding of collective mathematical argumentation. Specifically, we designed two AI chatbots that served different learning purposes. The first chatbot helped PMTs develop key ideas of collective mathematical argumentation and introduced Toulmin's model of argumentation. The second chatbot demonstrated the teacher's role in supporting collective mathematical argumentation. Both AI chatbots were designed to act as knowledge builders and situation generators, aligned with the initiating and exploring phases of the GRASP model, to support PMTs' understanding of argumentation practices through exemplified examples in a simulated interactive learning environment.

Guided by the prompt structure, the first AI chatbot provides concise and academically grounded explanations of collective mathematical argumentation and Toulmin's argument model. It uses a professorial tone and draws on course readings, academic papers, and policy documents embedded in the custom GPT to cite sources that PMTs typically read in methods courses. This design supports PMTs' understanding of collective argumentation in ways that are aligned with how researchers and teacher educators conceptualize this practice, addressing prior findings (e.g., Wagner et al., 2014) that preservice teachers often lack a robust grasp of what collective argumentation looks like in classroom practice. Next, to bridge theoretical knowledge with practice, the AI chatbot provides transcript examples of collective mathematical argumentation episodes that illustrate what this practice looks like in a mathematics classroom. Following this interaction, the AI chatbot presents examples of productive and unproductive collective argumentation episodes and explains the rationale for classifying each as productive or unproductive.

To support PMTs' pedagogical understanding of collective mathematical argumentation, a second AI chatbot was developed to help them examine how teachers support it in classroom settings. This chatbot was designed to make visible how teachers' instructional actions shape the development of students' mathematical arguments. It generates example classroom transcripts that enact a range of research-based teacher moves, such as eliciting student thinking, pressing for reasoning, and revoicing, in ways consistent with the literature on teacher support for collective argumentation (Conner et al., 2014). By illustrating these moves within simulated classroom episodes, the chatbot enables PMTs to analyze how particular teacher moves facilitate the development of collective argumentation in classroom contexts.

We utilized the GRASP model to enact these two phases because it can flexibly generate a wide range of realistic classroom scenarios for PMTs to practice and analyze, tailored to their interests, grade levels, and specific questions. Unlike traditional lecture-based instruction, in which students often engage passively with slides, AI-mediated interactions invite PMTs to take a more active role by posing questions, exploring alternative scenarios, and revisiting episodes at their own pace and in line with their prior knowledge. In addition, our prior work on using GenAI in mathematics teacher education suggests that prompt quality is an important factor in how effectively new users can engage with AI tools (Zhuang & Zhang, 2025a). Building on this insight, in both the Initiating and Exploring phases, we provided structured prompts that PMTs could use to interact with AI chatbots in the GRASP model, while also inviting them to pose free-form questions that emerged from their own curiosities, uncertainties about collective argumentation and teacher moves, and ongoing interactions with the AI chatbots. This combination of scaffolded and open-ended prompting was intended to lower barriers for novice AI users and to position PMTs as active inquirers into the foundational knowledge and practice of supporting collective mathematical argumentation.

Before implementing the GRASP model with PMTs, our research team followed an iterative design and testing process to ensure the customized AI chatbots aligned with the goals of the Initiating and Exploring phases described above. We began by writing detailed prompts for a custom GPT that incorporated relevant academic literature and policy documents on collective mathematical argumentation. The research team then conducted multiple rounds of internal testing, during which team members interacted with the chatbots as PMTs, using both structured prompts and open-ended questions.

Following each round of testing, we reviewed the chatbot outputs using three criteria. First, we examined whether the chatbots' explanations and definitions of argumentation aligned with established definitions and theoretical perspectives in the literature. Second, we evaluated the authenticity of the AI-generated classroom transcripts to determine whether they represented typical classroom experiences, considering the language, tone, and mathematical reasoning, as well as the flow of discussion, to be plausible for middle school students at the relevant grade levels. Finally, we analyzed whether the AI chatbots could appropriately respond to users' questions. Next, we iteratively revised the prompts and adjusted the chatbot instructions to reduce ambiguity and limit unproductive or misleading responses. This iterative process continued until the chatbots consistently produced outputs that aligned with our design expectations and provided PMTs with more coherent, concise, research-informed support for understanding collective mathematical argumentation.

METHODOLOGY

Participants and context of the study

This study was reviewed and approved by the Institutional Review Board (IRB). Participation was voluntary, and informed consent was obtained from all participants prior to data collection. The participants in this study were 10 of the 11 PMTs (91%) enrolled in a mathematics education course in Spring 2025 at a public research university located in the northeastern United States. These PMTs were in their senior year and planned to have their student teaching in the following semester. Six of them were female, and four were male. Six of the PMTs were White, three were Asian, and one identified as other/multiracial. Nine of the PMTs were 18 to 22 years old. The primary focus of this course was the interaction among pedagogy, content, and technology in mathematics teaching and learning related to secondary mathematics curricula. The class met weekly for three hours with homework assignment(s) after each class session. The PMTs were not introduced to the notion of collective mathematical argumentation and Toulmin's model prior to the study. During the last few weeks of the course, GenAI was introduced to PMTs as a tool for mathematics problem-solving, lesson planning, and teacher professional learning.

Preservice teachers' responses to the background survey items indicated that they frequently used GenAI tools, while 60% reported rarely or never using them to support their mathematics teaching. Among the 40% who had used GenAI to support teaching, most reported using it primarily to generate or better understand mathematical activities. The results from the pre-survey also showed that PMTs believed "it is important to support [their] [future] students to engage in classroom discussions where students and teachers make a mathematical claim and provide evidence to support it". PMTs' responses were predominantly *Strongly Agree* for this item.

Implementation of the GRASP model

After introducing GenAI as a tool for problem-solving and lesson planning, we implemented the GRASP model to support PMTs' learning about collective mathematical argumentation. PMTs interacted with the two customized AI chatbots described above. To facilitate meaningful engagement, PMTs were provided with a set of structured prompts to guide their initial interactions, while also being encouraged to pose free-form questions that emerged from their curiosities, uncertainties, and evolving understandings. Throughout the activity, the course instructor was available to address technical issues and provide support as needed, but the primary learning interactions occurred directly between PMTs and the two customized AI chatbots.

PMTs completed two homework assignments after interacting with the two AI chatbots corresponding to the Initiating and Exploring phases of the GRASP model. In the first assignment, PMTs analyzed two short transcripts of classroom interactions. For each transcript, they determined whether it represented an episode of collective mathematical argumentation. When they identified an episode as argumentative, they used Toulmin's model to label the argument components (e.g., data, claims, warrants, rebuttals) and to evaluate the episode's productivity. In the second assignment, PMTs reflected on their experiences interacting with the two customized AI chatbots, including what they learned about collective mathematical argumentation, which features of the chatbots they found useful, and what challenges they encountered during the interactions.

Data collection and analysis

To address the research questions, we collected multiple forms of data to examine how the GRASP model supported PMTs' understanding of collective mathematical argumentation and how PMTs perceived the strengths and limitations of the AI chatbots. These data sources included written assignments, reflections, and responses to pre- and post-surveys.

The research team employed a deductive qualitative approach (Elo & Kyngäs, 2008) using the rubrics shown in **Tables 1** and **2** to code the written assignment, in which PMTs analyzed two classroom transcripts. Each response was first coded as either correct or incorrect to determine whether the participant accurately identified the transcript as an episode of collective mathematical argumentation. Drawing on prior research on the definition of collective argumentation (e.g., Staples et al., 2017), we examined participants' written justifications along two dimensions: the collective and the structural. **Table 1** operationalizes a three-point rubric (1 = Limited, 2 = Basic, 3 = Advanced) to reflect the extent to which participants (a) recognized and explained how ideas and reasoning were built, connected, and advanced toward shared conclusions and (b) identified argument components using Toulmin's framework.

Table 1

Evaluating participants' understanding of collective mathematical argumentation

Coding Criteria	Definition	Rubric
Collective aspects of argumentation	We examined whether students demonstrated an understanding of the collaborative nature of argumentation, including recognizing when a group of students collectively engaged with mathematical ideas, contributed to each other's reasoning, and worked toward a shared mathematical conclusion.	Limited Understanding: No attention to the collective aspect. Basic Understanding: The student notes co-construction of arguments but provides limited detail on how the collective reasoning unfolded (e.g., simply citing speaker turns or shared answers). Advanced Understanding: Student identifies and explains how participants built on one another's ideas, refined reasoning together, and collectively progressed toward a mathematical understanding.
Structural aspects using Toulmin's Model	We assessed whether students could accurately identify key components of an argument, specifically, claims paired with appropriate warrants, and any evidence of rebuttals in alignment with Toulmin's (1958/2003) model of argumentation.	Limited Understanding: Fails to identify argument elements or mislabels non-argumentative statements. Shows confusion about the structure of an argument. Basic Understanding: Identifies at least one correct claim-warrant pair but does not fully account for the argument structure. Justification for identified elements may be vague or incomplete. May overlook or inaccurately represent rebuttals. Advanced Understanding: Accurately identifies multiple claim-warrant pairs. Provides coherent and well-reasoned justification for each component.

Guided by dimensions of argument validity adapted from Habermas's (1984) theory of validity claims — truth, rightness, sincerity — together with an added dimension of communicative effectiveness developed for mathematical argumentation in its initial version (Zhuang & Conner, 2020), we developed an updated coding framework shown in Table 2. We coded the number of validity dimensions each participant addressed. Responses were then scored using a parallel three-point rubric: Limited (one dimension), Basic (two dimensions), and Advanced (three or more dimensions).

Table 2

Validity dimensions for analyzing the productivity of argumentation episodes

Validity Dimensions	Definition
Truth	Argumentation leads to accurate mathematical conclusions, supported by logically valid claims and appropriately justified warrants. The truth of an argument was judged by the expert's perspective according to shared mathematical theorems, axioms, and principles in the given mathematical classroom community.
Rightness	Argumentation aligns with the social and sociomathematical norms of a given classroom context.
Sincerity	Participants build on one another's ideas not only to reach consensus but also to express and negotiate their own understandings, intentions, and uncertainties. This emphasis on sharing personal perspectives fosters mutual recognition and deeper engagement, rather than mere speculation or disconnection.
Communicative	The use of mathematical expressions (e.g., mathematical language, visual representations, symbolic notation, etc.) that are understandable in the mathematical community.

Next, using descriptive statistics, we analyzed data from the pre- and post-surveys and reflection exercise responses—PMTs enrolled in the course completed pre- and post-survey questionnaires. The surveys included nine Likert-type items on a seven-point scale, ranging from 1 = strongly disagree to 7 = strongly agree. Illustratively, survey items included: "I know how to help students build on each other's ideas during a math discussion" or "GenAI tools can be helpful for practicing my skills in facilitating mathematics classroom discussions" (see Table 3). We calculated mean differences and confidence intervals (Gravetter & Wallnau, 2021) for pre- and post-survey item responses to determine whether there were changes in the PMTs' perceptions of their ability to facilitate mathematical argumentation in classrooms and the usefulness of GenAI tools in supporting their development of these pedagogical skills.

Table 3

GRASP survey pre- and post-items

Question	Item
Question 1	I feel confident facilitating whole class discussions where students and teachers make a mathematical claim and provide evidence to support it.
Question 2	I believe it is important to support my [future] students to engage in classroom discussions where students and teachers make a mathematical claim and provide evidence to support it.
Question 3	I know how to help students build on each other's ideas during a math discussion.
Question 4	I have strategies to guide a class toward a shared mathematical understanding through discussion.
Question 5	I am comfortable allowing students to disagree with one another during math discussions.
Question 6	I can recognize when students are making mathematical arguments or critiquing others' reasoning.
Question 7	GenAI tools can be helpful in supporting whole classroom discussion where students and teachers make a mathematical claim and provide evidence to support it.
Question 8	GenAI tools can be helpful for practicing teaching.
Question 9	GenAI tools can be helpful for practicing my skills in facilitating mathematics classroom discussions.

Additionally, thematic analysis was employed to examine the written reflections, in which PMTs were prompted to reflect on their interactions with the two customized AI chatbots and described what they learned, what they found useful, and what challenges they experienced. The thematic analysis followed a primarily inductive and iterative coding process (Braun & Clarke, 2021). The goal of the analysis was to identify salient themes that captured participants' self-reported learning from interactions with the customized chatbots, the features they perceived as useful, and the perceived limitations of customized AI chatbots. For written assignments and reflections, two research team members independently coded the written responses, and any disagreements were resolved through collaborative discussion until consensus was reached.

RESULTS

Understanding the concepts of collective mathematical argumentation

Results from our analysis of PMTs' written assignments showed that following interactions with the first AI chatbot, participants generally demonstrated a well-developed conceptual understanding of collective mathematical argumentation. Only two of the ten participants misclassified episodes: one labeled a collective argumentation episode as 'none' and a non-collective episode as 'collective argumentation'; the other labeled a non-collective episode as 'collective argumentation'. Our analysis of PMTs' justifications revealed that, on the collective dimension, 50% demonstrated an advanced understanding of the first classroom interaction transcript, 10% a basic understanding, and 20% a limited understanding; one response (10%) was not scored due to incorrect classification. For the second transcript, 60% demonstrated advanced understanding, 20% demonstrated limited understanding, and 20% were not scored due to incorrect classification. Across responses, PMTs consistently articulated collective dimensions in their justifications. For example, PMT3 justified why the first transcript counted as collective argumentation by noting that students "are collectively working on a factorization problem by proposing diverse values for the blank and presenting reasons… reasoning is going back and forth." Similarly, PMT6 highlighted the joint construction of reasoning, explaining that because "they're sharing reasons out loud… everyone's building the reasoning together instead of one person doing all the work." In contrast, two PMTs misclassified the episode. For example, PMT1 argued that the episode was not collective argumentation because "there is little interaction between the students" and the warrant was not clearly articulated, reflecting a focus on the productivity or quality of the argumentation rather than on its definitional features.

For the structural dimension, which applies only to the collective argumentation episode, 60% of participants demonstrated an advanced understanding, 20% a basic understanding, and 10% a limited understanding; one response (10%) was unscorable due to incorrect classification. Taken together, these results indicate that the majority of PMTs achieved at least a basic understanding of collective argumentation across both dimensions. Participants with advanced understanding accurately identified multiple claim–warrant pairs and articulated how these components functioned within the episode. For example, PMT7 identified two distinct student claims ("6," "24") and paired them with the students' arithmetic reasoning as grounds, noting that S1 justified 6 by stating, "4×2 is 8 and $4 + 2$ is 6," and S2 justified 24 by explaining, "$4 + 4$ is 8 and 4×4 is 16." In contrast, PMT5 demonstrated limited understanding, as he was unable to identify warrants, stating, "Warrant: This step is mostly missing."

Our analysis of PMTs' evaluation of productive argumentation showed that 70% demonstrated advanced understanding, referencing three or more validity dimensions from Table 2. One participant (10%) demonstrated basic understanding by referencing two dimensions, while another (10%) showed limited understanding by referencing only one dimension. Most participants evaluated the truth of argumentation episodes by considering whether students recognized incorrect answers and why (e.g., "students also realize that they have made errors and explore why they made those errors as well," PMT8), and by noting how the conversation supported students' mathematical conceptual understanding (e.g., "connect[s] the concepts of addition and multiplication to factoring," PMT2). PMTs also attended to the teacher's role in creating a context for collective argumentation (Rightness), noting practices such as encouraging students to explain their reasoning and guiding self-correction. They further highlighted how teachers use incorrect answers as resources to support argumentation: "The teacher overall not criticizing the students for making mistakes, she takes that as a learning opportunity and turns those mistakes into valuable lessons so that the student can fully understand what they are doing wrong and why they did it wrong as well" (PMT8). Moreover, when participants judged productivity, they frequently flagged co-construction (sincerity), explicit mentions of how students build on each other's ideas, revising thinking based on feedback, and working together toward a shared solution. In contrast, only PMT2 focused on whether students use formal math vocabulary (communicative) by highlighting that "they (students) don't use formal math vocabulary."

The pre-survey means for Questions 1 through 8 ranged from 5.3 to 6.6. To examine any differences in PMTs' pre- and post-survey responses, we analyzed matched pairs of the survey items using descriptive statistics. Mean differences were calculated to determine if there were any differences in the pre- and post-survey item responses from PMTs related to their self-efficacy of facilitating classroom conversations and their beliefs on the use of GenAI in supporting them to learn how to facilitate classroom conversations (see Table 4). Notably, Question 9 (GenAI tools can be helpful for practicing my skills in facilitating mathematics classroom discussions) was the only item that showed a significant increase in the mean from the pre-survey (Mean = 4.5) to the post-survey (Mean = 6.0). Confidence intervals for the paired samples were used to estimate the mean difference between the pre- and post-item responses by analyzing the interval of the difference. The 95% confidence interval for the mean difference in Question 9 did not include zero, indicating the difference was statistically significant ($p < 0.05$). Also, the mean difference for Question 8 (GenAI tools can be helpful for practicing teaching) suggests a moderate change that perhaps a larger study could potentially detect as statistically significant. We acknowledge that this was a pilot study with a smaller sample size, and the results of the analysis of pre- and post-survey data are not generalizable. From a practical standpoint, the results from the survey data helped to gauge the preservice teachers' self-reported perceptions of the usefulness of GenAI tools in practicing their facilitation of mathematics classroom discussions. These preliminary results can inform design decisions in our future work.

Table 4

Summary of Mean Differences and Confidence Intervals

Item	Pre Mean	Post Mean	Mean Diff.	95% CI Lower	95% CI Upper
Question 1	5.5	5.8	0.30	-0.53	1.13
Question 2	6.6	6.3	-0.30	-1.31	0.71
Question 3	5.5	5.8	0.30	-0.53	1.13
Question 4	5.3	5.5	0.20	-0.61	1.01
Question 5	5.7	5.7	0.00	-0.89	0.89
Question 6	5.8	6.0	0.20	-0.25	0.65
Question 7	5.4	5.8	0.40	-0.44	1.24
Question 8	5.3	5.9	0.60	-0.24	1.44
Question 9	4.5	6.0	1.50	0.42	2.58

Findings from the thematic analysis further validate, enrich, and augment the above results derived from the written assignments and survey data. Indeed, our thematic analysis of PMTs' reflections revealed multiple layers of learning supported by their interaction with customized AI chatbots. The following themes capture the core areas of learning as articulated by the PMTs in their reflections.

Expanded understanding of mathematical argumentation

A recurring theme across PMTs' reflections was a broadened understanding of mathematical argumentation. Many PMTs came to see argumentation as more than just arriving at correct answers; they began to appreciate it as a process involving reasoning, explanation, and collaborative sense-making. One PMT summarized this shift by stating, "Mathematical argumentation isn't just about getting the right answer. It's about explaining why something works, using logic, and sometimes going back and forth with others to make sure the reasoning makes sense" (PMT6). Another remarked, "I realized that these arguments in mathematical classrooms are not just about having the correct answer… it's about justifying, connecting, and communicating ideas with evidence" (PMT8). Several PMTs also noted that exposure to contrasting examples helped them identify features of productive and unproductive classroom argumentation. One PMT commented, "I especially liked how the chatbot pointed out which arguments were productive, and which weren't" (PMT2).

Ability to analyze the structure of arguments

PMTs also reported learning how to analyze the structure of mathematical arguments using Toulmin's model. They learned to identify claims, data, warrants, rebuttals, and qualifiers in classroom dialogue. This analytical lens helped them better understand how students build and support their mathematical ideas. As one PMT reflected, "I was introduced to Toulmin's model… which breaks down arguments into different aspects like claim, data, warrant, rebuttal, and qualifier" (PMT8). Such insights suggest that chatbot interaction deepened PMTs' ability to analyze and support reasoning in discourse.

Insight into teacher moves that support argumentation

Many PMTs came to a clearer understanding of the teacher's role in supporting mathematical argumentation. They learned how teachers could guide student thinking through open-ended questioning, validating contributions, and using student errors productively. One PMT noted, "Teachers should act more as facilitators… asking open-ended questions, encouraging student contributions, and using student errors as opportunities for deeper reasoning" (PMT8). Another PMT wrote, "Teachers can ask good questions, clear up confusion, and keep the conversation focused without giving away the answer" (PMT6). Through chatbot interactions, PMTs were exposed to facilitation strategies that position students as active participants in constructing mathematical meaning.

Development of pedagogical vision

Many PMTs began to reflect on how they could apply what they were learning in their own future classrooms. They started to see themselves not just as learners of content, but as facilitators of discourse-oriented instruction. Several PMTs described gaining clarity on how to use argumentation strategies in practice. One stated, "I have a clearer understanding of how I can apply and use mathematical argumentation in a classroom setting, especially being a teacher" (PMT8). Another remarked, "The activity did help make me realize the bigger picture of how argumentation is being done in math class and how my role can be as a future teacher" (PMT3). These reflections highlight the pedagogical potential of AI chatbots to foster professional vision.

Perceived useful features of customized AI chatbots

PMTs identified several chatbot features as especially useful in supporting their understanding of mathematical argumentation and teacher facilitation of mathematical argumentation. These included access to realistic classroom dialogues, tailored explanations, interactive responsiveness, focused content, and structured content delivery. Together, these features suggest that well-designed, domain-specific AI chatbots can play a significant role in mathematics teacher preparation by bridging the gap between theory and practice.

Concrete examples and realistic classroom dialogues

Many PMTs highlighted the usefulness of chatbot-generated examples of classroom conversations. These realistic dialogues helped them visualize how mathematical argumentation and teacher questioning strategies play out in practice. One PMT noted, "The part I found most useful was the chatbot's ability to create real classroom conversations" (PMT2). PMTs appreciated that these examples modeled realistic student-teacher interactions and allowed them to analyze argument components using Toulmin's model. As one PMT shared, "My favorite feature of the chatbot was the example dialogues it would provide if asked. This is great to see how mathematical argumentation can work in a classroom" (PMT9). Another added, "The chatbot was able to apply Toulmin's model to a real classroom dialogue so that I could see clearly how to identify claims, data, warrants, and rebuttals" (PMT8). These examples served as powerful pedagogical tools, enabling PMTs to transfer theoretical understanding into instructional practices.

Clear definitions and explanations of complex concepts

PMTs consistently found the AI chatbot's ability to provide definitions and conceptual explanations helpful, especially for challenging ideas like Toulmin's model of argumentation, productive versus unproductive argumentation, and teacher facilitation moves. The ability to request clarification or ask follow-up questions further enhanced their understanding. One PMT explained, "I asked them to make the definition and to explain Toulmin's method in a simpler manner… they added so many more details and gave definitions" (PMT8). Such features empowered PMTs to explore complex concepts at their own pace.

Responsiveness to student prompts

The interactivity of the chatbots, particularly their responsiveness to specific prompts, was seen as a key strength. PMTs appreciated the flexibility to ask follow-up questions, request examples, or reword a response to meet their needs. One PMT reflected, "You can continue to ask for more clarification. This is super helpful when you do not initially understand something" (PMT4). Another said, "They can break down big or confusing questions into smaller parts… they give you just enough to get the idea without making it overwhelming" (PMT6). Another also mentioned, "I could ask them any questions I had, and they would help explain" (PMT8). This adaptive responsiveness supported differentiated learning pathways and reinforced agency in student learning.

Tailored focus on mathematical argumentation

PMTs valued that the chatbots were intentionally designed to focus on specific pedagogical content, namely, mathematical argumentation. Unlike general-purpose AI tools, these chatbots were aligned with the course goals and terminology, making their responses more relevant and meaningful. One PMT emphasized, "The most helpful features of the Chatbots were how they were specifically designed around mathematical argumentation and stayed on topic" (PMT10). Another noted, "It seems the Collective Mathematical Argumentation Chatbot is better at curating specific examples and scenarios …, whereas the Teacher Support Collective Mathematical Argumentation Chatbot relies more on generalized scenarios where it applies fundamental research concepts and conclusions" (PMT1). This alignment enhanced chatbots' effectiveness as learning tools in teacher preparation.

Structured content delivery

Some PMTs noted that structured outputs such as summary tables helped them organize and internalize key ideas. These features made abstract concepts more concrete and supported review and study. As one PMT shared, "They provided a summary table… it was extremely beneficial in the sense that it fully broke down and simplified everything I read above" (PMT8). Clear formatting and structured content delivery were found to be helpful for comprehension and retention.

Perceived Limitations of AI Chatbots

Data analysis revealed four key limitations identified by PMTs during their engagement with the two customized AI chatbots designed to support their learning of mathematical argumentation and teacher moves. Drawing on illustrative quotes from their written reflections, the limitations discussed here include information overload, lack of visual representation, lack of human feel, and insufficient context in AI-generated classroom dialogues.

Information overload: "Too much at once."

A frequently noted limitation was the overwhelming amount of information in a single chatbot response. Rather than breaking down complex ideas into smaller, digestible segments, the chatbot often delivered extended explanations that PMTs found difficult to process. This led to cognitive overload and disengagement. For example, one PMT remarked, "The amount of information was daunting at first to work through" (PMT5). At the same time, another reflected, "Sometimes the chatbot gave a lot of information quickly, so I had to pause and reread to really understand it" (PMT6). A third PMT noted the repetitiveness in the responses: "It's hard to retain my attention since it just seems like the same points repeated" (PMT9). Another PMT noted, "The chatbot gave me too much information, I'd rather it gives me a little bit and let me think about what I want before asking it to give me more" (PMT10). These insights suggest a need for progressive disclosure features or the ability to control the level of detail shown at one time.

Lack of visual representation

Several PMTs expressed a desire for visual aids to accompany the chatbot's textual information. Although the chatbots offered detailed definitions and examples, the absence of diagrams, illustrations, or video-based content made it more difficult for PMTs to visualize key concepts. As one PMT noted, "Sometimes a diagram would be nice to help map out different characteristics of mathematical argumentation" (PMT4). Another commented, "If the chats could provide video supports and examples… it would be awesome if they could create an AI video of what it would look like in a classroom" (PMT8). These responses indicate that visual and multimodal representations could enhance comprehension and make abstract ideas more accessible.

Lack of human feeling

Some PMTs found the chatbot responses to be impersonal or lacking the nuance and relational warmth of human interaction. Despite their instructional value, responses were sometimes described as overly mechanical or detached from authentic classroom dynamics. One PMT reflected, "It just felt very generic and vague and not specific. Also, it doesn't plan into real classroom scenarios as much as I thought it would" (PMT8). This PMT added, "The biggest challenge I had was the overall lack of realness and personalization that the chatbots have sometimes". Another stated, "It definitely lacks that human feel" (PMT2). These reflections suggest that chatbots should adopt more conversational, context-aware styles to better simulate responsive, human-like interactions.

Lack of context in AI-generated classroom transcripts

A few PMTs identified a lack of context in the AI-generated classroom transcripts as a barrier to meaningful analysis. Without a clear introduction to the problem or prior dialogue, PMTs found it challenging to make sense of the argumentative structure being modeled. As one PMT explained, "Sometimes the transcripts started in the middle of a conversation, so I didn't have the full context of the students' reasoning" (PMT2). This highlights the importance of providing framing information, such as the mathematical task, prior exchanges, and the roles of students in the transcript, to support interpretation and learning.

These limitations, including information overload, lack of visual representation, lack of human feel, and insufficient contextualization, identify critical areas for refining the design and implementation of AI chatbots in mathematics teacher education. Addressing these challenges can enhance the usability, engagement, and instructional value of such tools, enabling PMTs to explore mathematical argumentation and responsive pedagogical practices more effectively.

DISCUSSION

The purpose of this study was to examine whether customized ChatGPT-based chatbots could support PMTs' conceptual understanding of mathematical teaching practice with a focus on argumentation. Our analysis drew on data capturing both participants' ability to analyze classroom episodes and their perceptions of the learning support provided by AI chatbots. Specifically, we examined PMTs' understanding of collective mathematical argumentation through their written assignments, and we used survey responses and written reflections to explore their perceptions of this experience. Overall, the results indicated that the GRASP model could serve as a promising AI-supported pedagogical tool for scaffolding PMTs' conceptual understanding of collective mathematical argumentation by positioning them both as learners and future teachers. This dual perspective underscores the potential of AI tools to be integrated into teacher preparation courses as scaffolds that support PMTs' conceptual understanding of complex pedagogical practices.

Results from our analysis of PMTs' written assignments suggested that, after interacting with the GRASP model through the Initiating and Exploring phases, over half of the participants demonstrated at least a basic understanding of the collective dimension of argumentation, and 50% reached an advanced level. This finding suggests that AI-supported activities helped PMTs view argumentation as a collaborative process rather than merely as disagreements, broadening the narrow conceptions of argumentation documented in prior research (e.g., Staples & Newton, 2016; Wagner et al., 2014). At the same time, not all PMTs developed a robust understanding of collective argumentation through interaction with AI chatbots alone. Future studies should further examine factors that shape how PMTs learn with AI chatbots, such as their motivation to engage with text-based explanations and how the ways AI responses are represented and communicated influence what PMTs take up from these interactions.

Prior research has shown that teachers often struggle to identify the structural features of arguments (Bieda, 2010; Conner et al., 2014) and that Toulmin's model can support this aspect of understanding when teachers analyze authentic classroom transcripts (Wagner et al., 2014). However, methods courses may not always provide PMTs with sufficient opportunities to diagram argumentation episodes, given time constraints, the instructional effort required, and limited access to classroom episodes for practice. The dual design of AI chatbots as both knowledge builders and situation generators may help address this limitation by providing flexible access to multiple classroom-based scenarios. The results of this study suggest that interactions with AI chatbots could support many PMTs in correctly identifying key components of mathematical arguments.

In this study, PMTs demonstrated a relatively comprehensive understanding of productive argumentation: 70% attended to at least three validity dimensions. This result suggests that the interactions with the AI chatbots may have helped PMTs connect validity dimensions to rightness and sincerity, dimensions that prior research indicates are often underemphasized or insufficiently articulated in PMTs' interpretations of argumentation (Bieda, 2010; Wagner et al., 2014). Because this study focused only on the initiating and exploring phases of the GRASP model, future work should examine how PMTs apply validity dimensions during enactment and reflection phases. Furthermore, only one PMT addressed the communicative dimension, even though prior work highlights the importance of mathematical language and representations for the clarity and rigor of arguments (Zhuang & Conner, 2020). Future refinements of the chatbot designs could more explicitly model how teachers and students use mathematical language and notation to influence the process of argumentation.

Results from the analysis of the survey responses showed changes in PMTs' perspectives on the usefulness of GenAI tools in practicing skills in facilitating mathematics classroom discussions (Question 9). This result is encouraging and supports our other findings. The PMTs' reported experiences in the survey and written responses resonate strongly with emerging scholarship on the role of AI-supported tools in teacher preparation, particularly in helping novices bridge the persistent gap between pedagogical knowledge and the realities of classroom practice through simulated experiences (Lee & Yeo, 2022; Zhuang & Zhang, 2025a).

Prior studies have shown that representations of practice, such as transcripts, rehearsals, and simulated teaching, can support teacher learning by making complex interactions visible and analyzable (e.g., Grossman et al., 2009; Lampert et al., 2013). The chatbots' ability to generate realistic classroom dialogues appears to function as a new form of such representation, enabling PMTs to analyze and make sense of the structure of mathematical arguments in ways aligned with established research on mathematical argumentation, such as the widely used Toulmin model. The perceived usefulness of domain-specific design aligns with work showing that AI tools are most impactful when pedagogically targeted rather than generic (Schleiss et al., 2023), suggesting that intentional alignment with disciplinary and instructional goals is critical in the development of educational AI. In the GRASP model, this domain specificity was achieved by embedding key features of collective mathematical argumentation into AI-mediated learning activities, including attention to argument structure, collective discourse, and research-based teacher moves. In this way, GRASP positions AI chatbots as practice-based scaffolds for learning to teach argumentation rather than as general-purpose support tools. Finally, participants' appreciation for structured outputs, such as summary tables, aligns with research on knowledge organization and cognitive scaffolding that support teachers in internalizing complex pedagogical constructs (Bransford et al., 2000). Collectively, these findings contribute to a growing body of evidence that well-designed AI chatbots can serve as powerful mediational tools in mathematics teacher preparation, offering responsive, practice-oriented support that complements and extends traditional coursework.

The limitations identified by PMTs in this study echo important concerns raised in broader research on AI-supported learning tools and digital pedagogies (Giannakos et al., 2025; Kasneci et al., 2023). Consistent with cognitive load theory, PMTs noted that the chatbots often provided overly dense or repetitive responses, leading to information overload and difficulty processing key ideas, an issue well documented when digital tools fail to segment or scaffold complex content (Martin et al., 2025; Sweller, 2020). The absence of visual or multimodal representations similarly reflects established findings that visual supports, diagrams, and demonstrations enhance learners' conceptualization of abstract disciplinary constructs (Mayer, 2021), suggesting that text-only chatbot environments may underutilize opportunities for deeper comprehension. PMTs' critiques of the lack of a "human feel" align with prior studies that emphasize the importance of relational warmth, authenticity, and contextual sensitivity in teacher learning environments (Huynh & Aichner, 2025; Kishnani, 2025). AI-generated responses that feel generic or decontextualized may limit engagement and reduce perceived pedagogical relevance. Together, these limitations highlight critical design considerations for future development of AI chatbots in mathematics teacher education. Future designs of AI chatbots should address the need for more adaptive scaffolding, multimodal supports, context-rich representations of practice, and more naturalistic conversational features to better mirror authentic classroom interactions.

CONCLUSION AND LIMITATIONS

Previous research has shown the potential of GenAI tools to support lesson planning, strengthen content knowledge, and facilitate practice-based teaching (e.g., Gurl et al., 2024; Zhuang & Zhang, 2025b). The findings of this study extend this work by demonstrating that GenAI tools can also be leveraged to support teachers' understanding of complex pedagogical constructs. In particular, the use of custom GPT functionality enabled us, as teacher educators, to design AI chatbots that embed research-informed pedagogical ideas, ensuring that the concepts presented align with what we consider essential for PMTs' learning. Moreover, the AI chatbots provided an interactive, personalized, and risk-free learning environment. This approach differs from traditional methods courses, which typically rely on instructor lectures, shared video exemplars, and transcript analyses. The results of this study showed that the dual-role design of the GRASP model, which combines a knowledge builder that introduces key concepts of argumentation with a situation generator that produces corresponding classroom scenarios, served as a valuable resource for addressing challenges in teachers' understanding of argumentation identified in prior research (e.g., Conner et al., 2014; Kosko et al., 2014; Wagner et al., 2014). At the same time, the participants identified important limitations, including information overload, limited visual support, insufficient contextualization, and a lack of human-like interaction. These findings highlight both the promise and the constraints of integrating AI tools into teacher preparation. They offer insights into the design of AI chatbots as they explore how such tools can complement existing representations of practice. Rather than replacing traditional instruction, we view the GRASP model as an additional teaching tool that can broaden PMTs' opportunities to explore and make sense of instructional practices.

This study was conducted as an exploratory pilot case study with a relatively small sample size and without a comparison group; therefore, future research should employ larger samples and comparative or experimental designs, including control groups using traditional instructional approaches, to more rigorously examine the effects of AI tools on learning mathematical teaching practices. In addition, this study focuses on the learning outcomes associated with PMTs' interactions with AI tools rather than the underlying learning processes. We suggest that future research examine the chat histories of PMTs' interactions with the AI chatbots to better understand how they engaged with the AI-generated explanations and scenarios, where misunderstandings emerged, and what additional supports might enhance their learning through AI.

Acknowledgement

The authors thank the participants for their contributions to this study.

Funding

This research was supported by an internal grant from the College of Education at the Pennsylvania State University.

Ethical statement

This study was reviewed and approved by the Institutional Review Board of Pennsylvania State University (IRB Protocol No. 00026345). All participants provided informed consent before participation.

Competing interests

The authors declare that they have no competing interests.

Author contributions

Yuling Zhuang contributed to the research concept and design, data analysis and interpretation, writing of the article, and critical revision of the article. Xiangquan Yao contributed to the research concept and design, data collection, data analysis and interpretation, writing of the article, and critical revision of the article. Trina J. Davis contributed to the research concept and design, data analysis and interpretation, writing of the article, and critical revision of the article. Wisdom Y. Nudze contributed to data analysis and interpretation, as well as reference checking and formatting. All authors contributed to the final approval of the article.

Data availability

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

AI disclosure

The authors stated that Generative AI tools (e.g., ChatGPT by OpenAI) were used to improve the clarity of the manuscript in English.

Biographical sketch

Dr. Yuling Zhuang is an Assistant Professor of Mathematics Education in the Department of Teaching, Learning, and Culture at Texas A&M University, USA. Her research focuses on supporting meaningful mathematics learning by empowering teachers to facilitate mathematical argumentation and proof, designing interventions that advance the teaching and learning of argumentation and proof, and leveraging argumentation as a bridge for STEM integration. Her recent work extends this agenda by examining how AI-simulated role-play can support preservice teachers' practice-based learning and help students critically validate AI-generated mathematical responses through argumentative and proving practices.

Dr. Xiangquan Yao is an Associate Professor of Mathematics Education in the Department of Curriculum and Instruction at the Pennsylvania State University, USA. His research focuses on mathematical thinking and processes in dynamic and interactive mathematics learning environments, the development of technological, mathematical, and pedagogical understanding for teaching mathematics with technology, and the teaching and learning of mathematical modeling. His recent scholarship explores the design and use of AI-supported learning environments to support preservice mathematics teachers in learning and enacting core teaching practices.

Dr. Trina J. Davis is an Associate Professor of Learning Design and Technology and Program Chair of STEM Education in the Department of Teaching, Learning, and Culture at Texas A&M University, USA. Her research focuses on practice-based learning for preservice teachers, and various aspects of technology integration in teaching and learning, with an emphasis in mathematics and broader STEM education. Her work is centered in the design and evaluation of technology enhanced learning experiences, situated within interactive and immersive environments (e.g., AI-, virtual world-, VR-based simulations).

Wisdom Y. Nudze is a doctoral candidate in the Curriculum and Instruction program in the Department of Teaching, Learning and Culture at Texas A&M University, USA. His primary area of emphasis is Learning Design and Technology. His key research interests include (1) artificial intelligence (AI) in education, (2) virtual and augmented reality (VR/AR) in education, (3) digital equity and digital divide, and (4) preservice teacher preparation for technology integration in classrooms.

Disclaimer/Publisher's Note

REFERENCES

Almarashdi, H. S., Abu Khurma, O., AlArabi, K., Abulibdeh, E., & Yousef, J. (2026). AI-enhanced STEM education: A bibliometric study of research trends toward achieving sustainable development goals. *European Journal of STEM Education*, *11*(1), Article 16. https://doi.org/10.20897/ejsteme/18190

Andriessen, J. (2006). Arguing to learn. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 443–459). Cambridge University Press.

Australian Curriculum Assessment and Reporting Authority [ACARA]. (n.d.). *The Australian curriculum: Mathematics*. https://www.australiancurriculum.edu.au/f-10-curriculum/mathematics/

Bieda, K. N. (2010). Enacting proof-related tasks in middle school mathematics: Challenges and opportunities. *Journal for Research in Mathematics Education*, *41*(4), 351–382. https://doi.org/10.5951/jresematheduc.41.4.0351

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.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. National Academy Press.

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. Sage.

Conner, A., Singletary, L. M., Smith, R. C., Wagner, P. A., & Francisco, R. T. (2014). Teacher support for collective argumentation: A framework for examining how teachers support students' engagement in mathematical activities. *Educational Studies in Mathematics*, *86*(3), 401–429. https://doi.org/10.1007/s10649-014-9532-8

Cross, D. I. (2009). Creating optimal mathematics learning environments: Combining argumentation and writing to enhance achievement. *International Journal of Science and Mathematics Education*, *7*(5), 905–930. https://doi.org/10.1007/s10763-008-9144-9

Davis, T. J., Merchant, Z., & Kwok, O. M. (2022). An examination of practice-based virtual simulations and preservice mathematics teaching efficacy and outcome expectancy. *Education Sciences*, *12*, Article 262. https://doi.org/10.3390/educsci12040262

Department for Education. (2014). *National curriculum in England: Mathematics programmes of study*. Her Majesty's Stationery Office.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Ensor, P. (2001). From preservice mathematics teacher education to beginning teaching: A study in recontextualizing. *Journal for Research in Mathematics Education*, *32*(3), 296–320. https://doi.org/10.2307/749829

Francisco, J. M. (2025). Teacher support of collective argumentation in mathematics classrooms: Insights from an afterschool program. *Journal of Mathematics Teacher Education*, *28*(1), 211–237. https://doi.org/10.1007/s10857-024-09651-5

Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., ... & Rienties, B. (2025). The promise and challenges of generative AI in education. *Behaviour & Information Technology*, *44*(11), 2518–2544. https://doi.org/10.1080/0144929X.2024.2394886

Gravetter, F. J., & Wallnau, L. B. (2021). *Statistics for the behavioral sciences* (11th ed.). Cengage Learning.

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice*, *15*(2), 273–289. https://doi.org/10.1080/13540600902875340

Gurl, T. J., Markinson, M. P., & Artzt, A. F. (2024). Using ChatGPT as a lesson planning assistant with preservice secondary mathematics teachers. *Digital Experiences in Mathematics Education*, 1–26. https://doi.org/10.1007/s40751-024-00162-9

Habermas, J. (1984). *Theory of communicative action: Volume one: Reason and the rationalization of society* (T. McCarthy, Trans.). Beacon Press.

Huynh, M. T., & Aichner, T. (2025). In generative artificial intelligence we trust: Unpacking determinants and outcomes for cognitive trust. *AI & Society*, 1–21. https://doi.org/10.1007/s00146-025-02378-8

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, Y. R., Park, M. S., & Joung, E. (2025). Exploring the integration of artificial intelligence in math education: Preservice teachers' experiences and reflections on problem-posing activities with ChatGPT. *School Science and Mathematics*, 1–15. https://doi.org/10.1111/ssm.18336

Kishnani, D. (2025). *The uncanny valley: An empirical study on human perceptions of AI-generated text and images* [Doctoral dissertation, Massachusetts Institute of Technology].

Kosko, K. W., Rougee, A., & Herbst, P. (2014). What actions do teachers envision when asked to facilitate mathematical argumentation in the classroom? *Mathematics Education Research Journal*, *26*(3), 459–476. https://doi.org/10.1007/s13394-013-0116-1

Krummheuer, G. (1995). The ethnography of argumentation. In P. Cobb & H. Bauersfeld (Eds.), *The emergence of mathematical meaning: Interaction in classroom cultures* (pp. 229–269). Erlbaum.

Lampert, M., Franke, M. L., Kazemi, E., Ghousseini, H., Turrou, A. C., Beasley, H., Cunard, A., & Crowe, K. (2013). Keeping it complex: Using rehearsals to support novice teacher learning of ambitious teaching. *Journal of Teacher Education*, *64*(3), 226–243. https://doi.org/10.1177/0022487112473837

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Ledger, S., & Fischetti, J. (2020). Micro-teaching 2.0: Technology as the classroom. *Australasian Journal of Educational Technology*, *36*, 37–54. https://doi.org/10.14742/ajet.4561

Lee, D., Son, T., & Yeo, S. (2024). Impacts of interacting with an AI chatbot on preservice teachers' responsive teaching skills in math education. *Journal of Computer Assisted Learning*, *41*(1), Article e13091. https://doi.org/10.1111/jcal.13091

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

Martin, A. J., Collie, R. J., Kennett, R., Liu, D., Ginns, P., Sudimantara, L. B., ... & Rüschenpöhler, L. G. (2025). Integrating generative AI and load reduction instruction to individualize and optimize students' learning. *Learning and Individual Differences*, *121*, Article 102723. https://doi.org/10.1016/j.lindif.2025.102723

Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press.

McDonald, M., Kazemi, E., & Kavanagh, S. S. (2013). Core practices and pedagogies of teacher education: A call for a common language and collective activity. *Journal of Teacher Education*, *64*(5), 378–386. https://doi.org/10.1177/0022487113493807

Ministry of Education. (2007). *The New Zealand curriculum*.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. http://www.corestandards.org/Math/

Schleiss, J., Laupichler, M. C., Raupach, T., & Stober, S. (2023). AI course design planning framework: Developing domain-specific AI education courses. *Education Sciences*, *13*(9), Article 954. https://doi.org/10.3390/educsci13090954

Son, T., Yeo, S., & Lee, D. (2024). Exploring elementary preservice teachers' responsive teaching in mathematics through an artificial intelligence-based chatbot. *Teaching and Teacher Education*, *146*, Article 104640. https://doi.org/10.1016/j.tate.2024.104640

Staples, M., & Newton, J. (2016). Teachers' contextualization of argumentation in the mathematics classroom. *Theory Into Practice*, *55*(4), 294–301. https://doi.org/10.1080/00405841.2016.1208070

Staples, M., Newton, J., Kosko, K. W., Conner, A., Cirillo, M., Bieda, K., Yopp, D., Zaslavsky, O., Hummer, J., Strachota, S., Singh, R., An, T., Going, T., & Zhuang, Y. (2017). Using artifacts to explore conceptions and consequences of argumentation, justification, and proof [White Paper]. Working Group of the 39th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Available at https://www.researchgate.net/profile/Karl-Kosko/publication/317267228_ White_Paper_Using_Artifacts_to_Explore_Conceptions_and_Consequences_of_Argumentation_Justification_and_Proof/links/592f0edb0f7e9beee752bcd0/White-Paper-Using-Artifacts-to-Explore-Concep

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Toulmin, S. E. (2003). *The uses of argument* (Updated ed.). Cambridge University Press. (Original work published 1958)

Uğraş, H., Uğraş, M., Papadakis, S., & Kalogiannakis, M. (2024). Innovative early childhood STEM education with ChatGPT: Teacher perspectives. *Technology, Knowledge and Learning*, *30*, 809–831. https://doi.org/10.1007/s10758-024-09804-8

Wagner, P. A., Smith, R. C., Conner, A., Singletary, L. M., & Francisco, R. T. (2014). Using Toulmin's model to develop prospective secondary mathematics teachers' conceptions of collective argumentation. *Mathematics Teacher Educator*, *3*(1), 8–26. https://doi.org/10.5951/mathteaceduc.3.1.0008

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

Wood, T. (1999). Creating a context for argument in mathematics class. *Journal for Research in Mathematics Education*, *30*(2), 171–191. https://doi.org/10.2307/749609

Yackel, E. (2002). What we can learn from analyzing the teacher's role in collective argumentation. *The Journal of Mathematical Behavior*, *21*(4), 423–440. https://doi.org/10.1016/S0732-3123(02)00143-8

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, *27*, 458–477. https://doi.org/10.5951/jresematheduc.27.4.0458

Zhuang, Y., & Conner, A. (2020). Teacher questioning strategies in supporting validity of collective argumentation: Explanation adapted from Habermas' communicative theory. In A. I. Sacristán, J. C. Cortés-Zavala, & P. M. Ruiz-Arias (Eds.), *Mathematics education across cultures: Proceedings of the 42nd Meeting of the North American Chapter of* *the International Group for the Psychology of Mathematics Education* (pp. 2288–2296). Cinvestav/AMIUTEM/PME-NA. https://doi.org/10.51272/pmena.42.2020

Zhuang, Y., & Conner, A. (2022). Teachers' use of rational questioning strategies to promote student participation in collective argumentation. *Educational Studies in Mathematics*, *111*(2), 345–365. https://doi.org/10.1007/s10649-022-10160-6

Zhuang, Y., & Zhang, S. (2025a). Pre-service mathematics teachers' perceptions of using GenAI for practicing teacher questioning: A semester-long study. *Eurasia Journal of Mathematics, Science and Technology Education*, *21*(9), Article em2689.

Zhuang, Y., & Zhang, S. (2025b). Integrating ChatGPT in mathematics teacher education: AI-based simulation role-playing to support practice-based teaching. *International Journal of Artificial Intelligence in Education*, *35*(6), 3873–3895. https://doi.org/10.1007/s40593-025-00519-0

