PROSPECTIVE TEACHERS' INSTRUMENTAL GENESIS IN DESIGNING AI-GENERATED MATHEMATICAL APPLICATIONS

Maria Kattou, Constantinos Christou, & Demetra Pitta-Pantazi

University of Cyprus

The integration of AI into mathematics education is shifting the teacher's role from user of digital tools to designer of AI-generated apps. This study examines the criteria that prospective teachers consider when designing mathematical apps and the alignment between their design intentions and app evaluations. Forty-two prospective primary teachers (PT) participated in an AI-integrated mathematics education course, documenting their prompts, apps, and evaluations. Exploratory factor analysis identified three dimensions of criteria: pedagogical tool quality, learning experience, and assessment features. The results indicated alignment between PTs' design intentions and their app evaluations. AI appears to support the instrumental genesis by embedding pedagogical schemes directly into app design.

INTRODUCTION

The integration of artificial intelligence (AI) in mathematics education is reshaping teaching practices (Yeo et al., 2024). As AI tools become more accessible, teachers are transitioning from implementers of existing applications (apps) to designers of interactive AI-generated apps that provide adaptive learning experiences appropriate for students' needs. This shift represents a fundamental reconceptualization of the teacher's role requiring multiple competencies, including mathematical content knowledge, pedagogical expertise, understanding of students' learning, and technological proficiency. The teacher's role as designer is not merely an adaptation to new technological capabilities; rather, it primarily entails developing AI-generated pedagogical approaches. Thus, teachers enrich their pedagogical knowledge and skills through iterative cycles of creating, testing, and improving learning materials. In this process, teachers reflect on their knowledge of how students construct mathematical understanding, identifying which representations support conceptual development in ways that existing materials and tools cannot.

Despite the potential of AI tools, significant questions remain about how teachers conceive of the designer's role and which criteria they consider during the design and evaluation processes. Until now, most research on AI has focused on algorithmic capabilities or student learning outcomes rather than on the evolving approaches that they develop with AI (e.g., Lee & Yeo, 2022). Additionally, to the best of our knowledge, no previous research has investigated teachers' abilities as designers of AI-generated mathematical apps. This gap is particularly pronounced in mathematics

2026. In H. Krzywacki, M. Hähkiöniemi, E. Haataja, M.S. Hannula & P. Hästö (Eds.). *Proceedings of the 49th Conference of the International Group for the Psychology of Mathematics Education: Research Reports*, pp. 763-770. PME.

teaching because teachers must emphasize visual representations, procedural fluency, and conceptual understanding.

This study addresses this gap by investigating prospective teachers' (PTs') ability to design AI-generated mathematical apps and proposes a framework of assessment criteria for such apps. These criteria are used to identify the pedagogical characteristics that PTs incorporate into their designs, as reflected in their prompts, and in the evaluations of the AI-generated apps they developed.

THEORETICAL FRAMEWORK

From tool users to tool designers: A shifting paradigm in mathematics education

The potential of AI in mathematics education is multifaceted, offering educators innovative tools and resources to enrich their pedagogical practices (Yeo et al., 2024). Empirical research on the integration of AI in mathematics education remains limited (Lee & Yeo, 2022). Early applications of AI in mathematics education include chatbots, provide explanations, support problem-solving, and prepare for assessment (e.g., Taani & Alabidi, 2024).

Beyond these advancements, mathematics education research witnesses a significant paradigm shift. Whereas earlier technology integration focused on how digital tools support students' mathematical learning, research will now increasingly examine how teachers purposefully design technology-enhanced activities to foster mathematical learning (Clark-Wilson et al., 2020). The emergence of AI technology has accelerated this shift, enabling teachers' transition from orchestrators of existing digital apps to designers of customized mathematical apps aligned with their pedagogical vision.

This transformation requires reconceptualizing the theoretical frameworks that guide our understanding of technology integration in mathematics education. The theory of instrumental genesis (Artigue, 2002), which has predominantly examined how students transform technological artefacts into cognitive instruments, must now be extended to account for the teacher's role as designer of AI-generated mathematical apps.

Instrumental genesis: Expanding the framework

According to the instrumental genesis theory, a technological object becomes an instrument through two interrelated processes (Drijvers et al., 2010). Instrumentation describes how the artefact shapes users' thinking, practices, and problem-solving approaches. Instrumentalization refers to how users adapt and customize the artefact to their needs and purposes. Through these bidirectional processes, users develop utilization schemes that integrate domain-specific knowledge with technical proficiency and purposeful application.

While substantial research (e.g., Kieran & Drijvers, 2006) has examined students' instrumental genesis with mathematical software, teachers' instrumental genesis in designing mathematical apps remains largely unexplored. When designing such AI-generated apps, teachers must make explicit decisions about the mathematical content,

representations, and students' thinking and understanding, while embedding their technological and pedagogical assumptions into the tool's structure. This process entails an iterative cycle in which teachers communicate their pedagogical intentions, evaluate AI-generated outputs, identify limitations, refine their prompts, and collaboratively co-design mathematical apps with AI.

In this iterative process, AI functions not merely as a content generator but as a reflective partner in pedagogical design. Teachers translate their pedagogical knowledge into explicit design criteria through their prompts. The AI's outputs may align with or diverge from teachers' expectations; however, through reflection and prompt refinement, teachers can modify the content, the pedagogical affordances, and functionalities of the app. Through this dialogue, the artefact, which is the AI-generated app, gradually evolves into a calibrated instrument aligned with specific learning goals.

Criteria for evaluating the AI-generated mathematical apps

As teachers assume designer roles, establishing quality criteria for AI-generated apps becomes essential. Such assessment criteria are helpful for app designers during the development process, guiding the creation of more pedagogical applications that enhance learning (Harrison & Lee, 2018). Assessment criteria for digital apps have been widely proposed (e.g., Bokhove & Drijvers, 2010; Cayton-Hodges et al., 2015). Drawing on instrumental genesis theory, Bokhove and Drijvers (2010) proposed the need for developing an "evaluation instrument". They proposed an initial set of 27 criteria for digital algebra tools, emphasizing the integration of tool techniques with conceptual understanding and the inclusion of feedback mechanisms to support students' learning. Subsequent studies confirmed the persistence of core dimensions related to mathematical content, such as mathematical accuracy and richness (Cayton-Hodges et al., 2015). Features of the tool's surface, such as representations, potential for interaction, degree of adaptability, and opportunities for user experimentation, have also been highlighted (Cayton-Hodges et al., 2015). Another convergence across studies was the tool's ability to provide feedback and scaffold students' learning, offering potential for differentiation (e.g., Cayton-Hodges et al., 2015).

PURPOSE OF THE STUDY

This theoretical background reveals critical gaps. While previous studies provided criteria for assessing mathematical software and existing apps, we lack an empirical understanding of how teachers navigate AI to design mathematical apps appropriate for their teaching. Hence, this study investigates PTs' design of AI-generated mathematical apps by examining: (1) the criteria PTs consider during the design process, and (2) the alignment between PTs' design intentions, as expressed in their prompts, and their evaluation of the resulting AI-generated apps.

METHOD

Participants

The present study presents preliminary results of a project conducted in Cyprus in the fall semester of the 2025-2026 academic year, involving 42 prospective primary school teachers (PT) (males = 13, females = 29; age range = 19-23 years). Most participants were in their second year of study (39), whereas two were in their third year and one was in their fourth and final year.

Procedure

As part of their studies, PTs were required to attend a course in the Didactics of Mathematics. The course aimed to familiarize PTs with the fundamental principles of mathematics teaching and learning, the objectives and content of primary school mathematics, instructional methods and resources, and contemporary assessment approaches for students' abilities. The course lasted 13 weeks, delivered by two of the three authors, three times per week, with each session lasting 75 minutes. At the beginning of the semester, participants provided written consent to participate in the research and granted the researchers access to all collected data for research purposes.

During this course, AI was systematically used as a teaching tool to enhance students' understanding and engagement. PTs were invited to explore mathematical concepts and teaching approaches. Furthermore, AI was used as an instructional resource to create lesson plans, define teaching objectives, and develop engaging activities. In this framework, AI tools enabled PTs to create AI-generated mathematical apps by providing prompts regarding the app's aim, characteristics, and functionalities.

Data collection and analysis

During this course, the PTs were introduced to the unit on fractions. This included the fraction subconstructs, curricular development of fraction concepts, teaching considerations, operations, and problems with fractions. PTs had the opportunity to use online apps and GeoGebra lessons to discuss their affordances and limitations, and how these activities could be employed in the classroom. Based on their experiences, PTs were asked to use AI tools to develop an AI-generated mathematical app to teach elementary school students any fraction concept or operation. PTs began their experimentation during the course session and were given 3 days to submit their app, along with the prompts they used during the design process with the AI tools.

Furthermore, a framework with 15 assessment criteria for mathematical apps (see Table 1) was developed based on existing literature review (e.g., Bokhove & Drijvers, 2010; Cayton-Hodges et al., 2015). This framework had a dual role in the data analysis. First, the criteria were used to evaluate the prompts by coding the characteristics mentioned or implied by the PTs during the design process. Each characteristic was scored dichotomously (0 or 1). Secondly, these criteria were used to develop a questionnaire that PTs completed to evaluate the apps they designed. The questionnaire

included the statements on a Likert scale (1 = Strongly Disagree – 5 = Strongly Agree) and was completed after the PTs had submitted their prompts and apps. The questionnaire's internal reliability was examined using Cronbach's alpha (α = .85). Given the small sample size, the alpha coefficient is considered indicative of preliminary reliability.

To answer the first research question, that is, to uncover the underlying structure of PTs' assessment criteria, we used an exploratory factor analysis (EFA). For the second research question, we examined the alignment between PTs' prompts and their app evaluations by computing the mean scores of the identified factors across prompts and comparing them with the EFA results.

RESULTS

To address the first research question, we conducted an EFA on PTs' evaluation ratings of their AI-generated apps to uncover the underlying structure of their assessment criteria. Prior to the analysis, the assumptions for EFA were examined and met (KMO = .65; Bartlett's $\chi^2(105) = 216.23$, $p < .001$). We applied principal axis factoring for extraction and Varimax rotation for interpretability. Table 1 presents the three factors that accounted for 55.02% of the total variance (Eigenvalues: 5.95, 2.23, and 1.32, respectively). All retained items demonstrated loadings above .47 on their primary factor, supporting the adequacy of the factor structure. Factor 1 (items 1-7; α = .83) measured *Pedagogical tool quality* (23.29% of variance). Factor 2 (items 8-12; α = .84) captured *Learning experience* (19.29% of variance). Factor 3 (items 13-15; α = .72) reflected *Assessment features* (12.44% of variance). All Cronbach's alpha values indicated acceptable reliability.

Table 1: Factor Loadings for Exploratory Factor Analysis of the Questionnaire.

	Items	Factor		
		1	2	3
1	The app is dynamic.	**.76**	.20	-.03
2	The app incorporates representations of concepts.	**.71**	.17	.03
3	The app supports practice of mathematical concepts.	**.69**	.39	.05
4	The app is user-friendly.	**.66**	.29	.02
5	The app can be connected to the math curriculum.	**.53**	.46	.22
6	The app promotes exploration of mathematical concepts.	**.48**	.09	.13
7	The app promotes inquiry-based learning.	**.47**	.22	-.29
8	The app offers interaction with students.	.18	**.75**	.07
9	The app offers understanding of mathematical concepts.	.60	.73	.06
10	The app increases students' interest.	.39	**.69**	.16
11	The app connects mathematical concepts with real-life.	.34	**.56**	-.01
12	The app encourages reflection on math concepts.	.17	**.54**	.27

13	The app allows differentiation of learning.	.16	-.03	**.94**
14	The app enables the assessment of students.	.09	.17	**.69**
15	The app provides feedback to students.	-.21	.32	**.51**
	Eigenvalues	5.95	2.23	1.32
	% of Variance	23.29	19.29	12.44

For the second research question, we examined the alignment between PTs' prompts and their app evaluations by computing the mean factor scores for their prompts. PTs emphasized Pedagogical tool quality (M = .35, SD = .22), followed by Learning experience (M = .26, SD = .18) and Assessment features (M = .17, SD = .24) through their prompts, indicating consistency between design intentions and evaluations.

To further illustrate this alignment, indicative prompts are presented below, demonstrating how PTs' design intentions aligned with the three assessment criteria for AI-generated apps. For Pedagogical tool quality, prompts emphasized dynamic applications with multiple representations, engaging activities supporting exploration, inquiry, and practice, and alignment with the mathematics curriculum. For example, PTs requested: *"Design a dynamic app to teach the concept of fractions. It should include multiple representations"*, *"Include a 'New Exercise' button, allowing children to practice on their own,"* and *"The app should be more engaging, by allowing students to create equivalent fractions and check if they are correct"*. For the Learning experience, prompts emphasized fostering students' engagement and conceptual understanding through meaningful, motivating contexts. For example, they stated: *"It should have a game-like style and be engaging for all students,"* and *"The application should include images, interactive exercises, and real-life examples."*

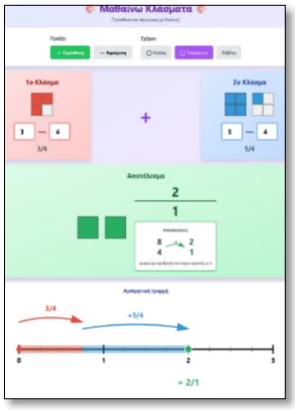
For Assessment features, prompts emphasized differentiation, feedback, and learning evaluation, with PTs focusing on accommodating varying student proficiency, providing scaffolded feedback. For example, they requested: *"The app should include three levels of student proficiency"*, *"When a student initially provides an incorrect answer, the app should offer hints without revealing the correct solution. If errors persist, the app should present the correct answer along with an explanation"* and *"For each correct answer, the student earns 10 points. When the student accumulates 200 points, positive feedback should be displayed"*.

Figure 1 shows screenshots of the resulting AI-generated apps, demonstrating how prompts translated into features across the three dimensions of assessment criteria.

Figure 1: Indicative mathematical apps.

Fractions on a number line

Equivalent fractions



Addition of fractions

DISCUSSION

AI is shifting teachers' roles (Yeo et al., 2024) from implementers of existing mathematical apps to designers of interactive AI-generated apps. Hence, AI provides teachers with innovative tools to enhance their pedagogical practices (Yeo et al., 2024). This study aimed to investigate (a) the criteria that PTs consider when designing AI-generated mathematical apps, and (b) the relationship between PTs' design intentions and their evaluation of the apps they generated with AI tools.

In response to the first research question, an EFA identified three dimensions of criteria for AI-generated apps: Pedagogical tool quality, Learning experience, and Assessment features. Although no prior study has examined the criteria teachers consider when designing apps, the results align with assessment criteria for existing mathematical apps. For instance, Bokhove and Drijvers (2010) identified similar dimensions: tools, general characteristics, and assessment criteria, respectively. Cayton-Hodges et al. (2015) referred to the criteria for adaptability of applications, richness of interaction, and feedback/scaffolding as features of digital tools.

As PTs designed their apps, they considered pedagogical principles related to students' mathematical learning, and technological features such as usability and adaptability. PTs placed the greatest emphasis on representations, interaction, and alignment with the curriculum, reflecting the quality of the pedagogical tool. They also considered engagement and conceptual understanding through meaningful, motivating contexts, addressing users' learning experience. Further, they considered assessment features, such as differentiation, scaffolded feedback, and reinforcement of learning. Through this process, PTs externalized their underlying pedagogical assumptions as they crafted prompts and communicated with AI to design their apps.

The consistency between PTs' design intentions and the dimensions emphasized in their evaluations suggests that PTs' instrumental genesis unfolds during the design process, as pedagogical schemes are embedded in the AI-generated app. Through an iterative cycle, PTs communicated their pedagogical intentions, evaluated AI-

generated outputs, refined their prompts, and ultimately co-designed the mathematical app. It appears that the app shaped PTs' thinking and practices (instrumentation), while PTs adapted it according to their pedagogical assumptions (instrumentalization). While instrumental genesis (Artigue, 2002) describes how users transform existing artefacts into instruments through instrumentalization and instrumentation, our study reveals a distinct process. Now, with AI, PTs can create their own technological artefacts. In this study, instrumental genesis begins not with appropriation but with construction as pedagogical schemes are embedded directly into the artefact's design. PTs can now design technological artefacts that are then transformed into cognitive instruments.

While the findings provide valuable insights, they should be interpreted with certain limitations in mind. The sample was relatively small, which may have influenced the outcomes. Future research could examine larger, more diverse samples, including in-service teachers, and explore the evolution of prompts and apps across multiple design iterations to better understand how teachers' intentions and apps develop over time. AI-generated apps could also be used to assess teachers' knowledge for teaching.

References

Artigue, M. (2002). Learning mathematics in a CAS environment: The genesis of a reflection about instrumentation and the dialectics between technical and conceptual work. *International Journal of Computers for Mathematical Learning, 7,* 245–274.

Bokhove, C. & Drijvers, P. (2010). Digital Tools for Algebra Education: Criteria and Evaluation. *International Journal of Computers for Mathematical Learning, 15,* 45–62.

Cayton-Hodges, G. A., Feng, G., & Pan, X. (2015). Tablet-based math assessment: What can we learn from math apps?". *Educational Technology & Society, 18*(2), 3–20.

Clark-Wilson, A., Robutti, O., & Thomas, M. (2020). Teaching with digital technology. *ZDM Mathematics Education, 52*(7), 1223–1242.

Drijvers, P., Doorman, M., Boon, P., Reed, H., & Gravemeijer, K. (2010). The teacher and the tool: instrumental orchestrations in the technology-rich mathematics classroom. *Educational Studies in Mathematics, 75,* 213–234.

Harrison, T.R. & Lee, H.S. (2018). iPads in the mathematics classroom: Developing criteria for selecting appropriate learning apps. *International Journal of Education in Mathematics, Science and Technology, 6*(2), 155–172.

Kieran, C. & Drijvers, P. (2006). The co-emergence of machine techniques, paper-and-pencil techniques, and theoretical reflection: A study of CAS use in secondary school algebra. *International Journal of Computers for Mathematical Learning, 11,* 205–263.

Lee, D. & Yeo, S. (2022). Developing an AI-based chatbot for practicing responsive teaching in mathematics. *Computers & Education, 191,* 1–17.

Taani, O. & Alabidi, S. (2024). ChatGPT in education: benefits and challenges of ChatGPT for mathematics and science teaching practices. *International Journal of Mathematical Education in Science and Technology, 56,* 1–30.

Yeo, S., Moon, J., & Kim, D. (2024). Transforming mathematics education with AI: Innovations, implementations, and insights. *The Math. Education, 63*(2), 386–392.