

Shaping Mathematics Activities with Generative AI: Prompt Types, Models and Pedagogical Outcomes

Yapay Zekâ ile Matematik Etkinliklerinin Şekillendirilmesi: Prompt Türleri, Modeller ve Pedagojik Çıktılar

Mustafa Zeki AYDOĞDU*  Büşra ÇAYLAN ERGENE**  Özkan ERGENE*** 

Received: 2 October 2025

Research Article

Accepted: 18 December 2025

ABSTRACT: This study investigates the relationship between prompt types and the quality of mathematics activities generated by artificial intelligence (AI) tools. Within a multiple-case study design, two advanced AI systems, ChatGPT-5 (OpenAI, September 2025) and Gemini 2.5 Pro (Google DeepMind, September, 2025), were examined using command (C) and request (R) prompts under standardised settings (temperature = 0.7, top-p = 0.9). Four activities were produced and evaluated with the Activity Evaluation and Feedback Tool, which assesses both component-level features and overall quality. The analysis revealed that three of the four AI-generated activities reached the high-quality range, with total scores of 22, 19, and 23 out of 24 points for Gemini-R, Gemini-C, and ChatGPT-C, respectively, whereas ChatGPT-R scored 15 points, indicating a medium level. ChatGPT demonstrated greater effectiveness with command prompts, whereas Gemini performed better with request prompts. At the component level, intended outcome and materials were strong, while weaknesses were observed in instructions, responsibility, and complexity, depending on the AI-prompt combination. These findings demonstrate that activity quality is shaped not only by prompt design but also by model-specific affordances. Implications are discussed for teacher education, curriculum development, and comparative research on the integration of generative AI in mathematics education.

Keywords: Generative artificial intelligence, chatgpt, Gemini, prompt engineering, mathematics activity design, teacher education.

ÖZ: Bu çalışma, prompt türleri ile yapay zekâ (YZ) araçları tarafından üretilen matematik etkinliklerinin kalitesi arasındaki ilişkiyi incelemektedir. Çoklu durum çalışması deseni kapsamında, iki gelişmiş yapay zekâ sistemi olan ChatGPT-5 (OpenAI, Eylül 2025) ve Gemini 2.5 Pro (Google DeepMind, Eylül 2025), emir ve rica türü promptlar kullanılarak standart ayarlarda (temperature = 0.7, top-p = 0.9) test edilmiştir. Bu süreçte dört etkinlik üretilmiş ve bu etkinlikler, hem bileşen düzeyindeki özellikleri hem de genel kaliteyi ölçen Etkinlik Değerlendirme ve Geri Bildirim Aracı ile analiz edilmiştir. Bulgular, dört etkinlikten üçünün yüksek kalite aralığında (Gemini-R = 22, Gemini-C = 19, ChatGPT-C = 23 / 24 puan) yer aldığını, ChatGPT-R çıktısının ise 15 puanla orta düzeyde kaldığını göstermiştir. ChatGPT emir türü promptlarda daha yüksek etkinlik gösterirken, Gemini rica türü promptlarla daha iyi performans sergilemiştir. Bileşen düzeyinde, “hedeflenen kazanım” ve “materyaller” güçlü yönler olarak öne çıkarken; “yönergeler”, “sorumluluk” ve “karmaşıklık” alanlarında ise kullanılan YZ ve prompt türü kombinasyonuna bağlı zayıflıklar tespit edilmiştir. Bu bulgular, etkinlik kalitesinin yalnızca prompt tasarımına göre değil, aynı zamanda kullanılan modelin kendine özgü olanaklarına göre de şekillendiğini ortaya koymaktadır. Çalışmanın sonuçlarının öğretmen eğitimi, müfredat geliştirme ve matematik eğitiminde üretken YZ entegrasyonuna yönelik karşılaştırmalı araştırmalar için doğurguları tartışılmıştır.

Anahtar kelimeler: Üretken yapay zekâ, chatgpt, gemini, prompt mühendisliği, matematik etkinliği tasarımı, öğretmen eğitimi.

* *Corresponding Author:* Asst. Prof. Dr., Trakya University, Edirne, Türkiye, mzekiaydogdu@trakya.edu.tr, <https://orcid.org/0000-0003-1163-2890>

** Asst. Prof. Dr., Sakarya University, Sakarya, Türkiye, bcaylan@sakarya.edu.tr, <https://orcid.org/0000-0002-5567-6791>

*** Assoc. Prof. Dr., Sakarya University, Sakarya, Türkiye, ozkanergene@sakarya.edu.tr, <https://orcid.org/0000-0001-5119-2813>

Citation Information

Aydoğdu, M. Z., Çaylan Ergene, B., & Ergene, Ö. (2026). Shaping mathematics activities with generative AI: Prompt types, models and pedagogical outcomes. *Journal of Theoretical Educational Sciences [Kuramsal Eğitim Bilim Dergisi]*, 19(2), 405-432. <http://doi.org/10.30831/akukeg.1795773>

Copyright © 2026 Author(s).

This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0 license](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Artificial intelligence (AI) has emerged as a transformative force within the Fourth Industrial Revolution, reshaping multiple sectors, including education (Hossain, 2023). International organisations such as UNESCO emphasises AI's potential to personalise learning, reduce inequalities, and foster lifelong learning opportunities (UNESCO, 2023). In education, AI technologies now enable personalised pathways, adaptive assessments, and real-time feedback mechanisms that were not previously feasible (Opesemowo & Adewuyi, 2024). The post-pandemic acceleration of digitalisation has positioned AI not as an optional technological supplement but as a fundamental partner in teaching and learning processes (Moorhouse, 2023).

Mathematics education is one of the most fertile grounds for AI applications. AI tools have been applied to intelligent tutoring, automated grading, and the development of computational and critical thinking (Yeo et al., 2024). Generative AI systems such as ChatGPT or Gemini have been shown to support proof construction, problem solving, and logical reasoning, while adaptive systems like ALEKS enhance personalised practice (Yi et al., 2024). Recent reviews highlight AI's capacity to enhance conceptual understanding, improve problem-solving strategies, and empower teachers by reducing workload and offering new instructional resources (Hwang & Tu, 2021; Jin & Suh, 2024; Yi et al., 2024).

A distinctive feature of generative AI tools lies in their reliance on prompts (Shi et al., 2020). Prompts serve as the communicative interface through which users elicit responses, and their structure significantly shapes the quality and relevance of AI outputs (Shin et al., 2024). In mathematics education, prompts can be framed as requests or commands, which may influence not only the type of response generated but also the degree of agency and creativity afforded to students and teachers. Recent studies emphasise that systematic prompt design enhances the accuracy of automated grading, supports the quality of feedback, and fosters deeper engagement with mathematical ideas (Saralar-Aras & Cicek Schoenberg, 2024; Wardat et al., 2023). Despite this recognition, research comparing how different prompt types influence AI-supported activity design remains limited.

Although the transformative potential of AI in mathematics education has been widely acknowledged, systematic evidence on how different AI tools (e.g., ChatGPT, Gemini) and prompt types (e.g., requests versus commands) influence the design and quality of mathematics activities remains scarce. Existing reviews note the predominance of ChatGPT as the primary tool under investigation, with limited comparative studies involving alternative generative AI systems (Opesemowo & Adewuyi, 2024). Moreover, while prompt engineering has gained increasing attention, its pedagogical implications for designing mathematics activities are not yet fully understood. Addressing this gap, the present study investigates four mathematics activities developed through different AI tools and prompt types, with the aim of examining how these variations shape the nature and educational value of the resulting tasks. This study is guided by the following research question: How different AI tools (ChatGPT and Gemini) and prompt types (request and command) shape the structure and quality of mathematics activities? By addressing this question, the study seeks to contribute to the growing body of research on AI integration in mathematics education, offering insights for both theory and practice in the era of generative AI. The findings are expected to clarify how design choices in prompts influence the pedagogical value

of AI-generated activities, thereby informing both teacher education and curriculum development. Moreover, by comparing two prominent AI tools, the study provides empirical evidence that can guide educators and policymakers in making informed decisions about the effective and responsible use of generative AI in mathematics classrooms.

Theoretical Background

AI in Mathematics Education

The integration of AI in education has accelerated in the last decade, with generative AI tools now shaping how knowledge is accessed, mediated, and produced. UNESCO (2023) highlights AI as a transformative enabler of equitable, inclusive, and personalised learning, emphasising the need to balance technological innovation with ethical and pedagogical considerations. AI applications range from adaptive learning systems and automated feedback to intelligent tutoring and large language model (LLM) based conversational agents (Yeo et al., 2024). In this context, generative AI occupies a distinctive position, offering not only information retrieval but also the co-construction of knowledge through interactive prompts. Yet, its integration into teaching requires critical reflection on how students and teachers engage with such tools (Opesemowo & Adewuyi, 2024). Mathematics education has been identified as a field particularly receptive to AI innovations. Systematic reviews reveal that AI supports multiple domains, including conceptual understanding, computational thinking, statistical literacy, and proof construction (Hwang & Tu, 2021; Jin & Suh, 2024; Yi et al., 2024). Studies using ChatGPT demonstrate that students can receive timely feedback, explore proofs interactively, and engage in creative problem-solving (Park & Manley, 2024; Wardat et al., 2023). At the same time, alternative AI systems such as Gemini or domain-specific tutors expand the ecosystem of tools, but their comparative pedagogical affordances remain under-examined (Li, 2025). While the potential of AI in mathematics education is broadly acknowledged, implementation varies with teachers' technological-pedagogical knowledge and their perceptions of reliability and trust (Kim, 2024; Yeo et al., 2024). Recent studies further demonstrate that students' perceptions of ChatGPT in mathematics differ significantly according to academic background, experience level, and frequency of use, highlighting perception as a critical mediating factor in AI-supported mathematics learning (Ergene & Caylan Ergene, 2025a). These findings indicate that users' perceptions are closely intertwined with how they engage with AI tools during mathematical tasks. These contextual factors suggest that understanding how AI contributes to activity design requires a closer look at the role of user AI interactions.

Mathematics Activities and AI

The concept of activity and activity-based teaching has remained a central concern in mathematics education research. Textbook analyses, the design of digital learning platforms, and international comparative studies consistently highlight the centrality of activities in structuring mathematics learning experiences (Jones & Pepin, 2016; Sullivan et al., 2015). Activities are valued because they enable students to inquire, reason, and experiment; learners can test hypotheses, make conjectures, and explore alternative strategies in ways that promote both understanding and analysis of

mathematical ideas (Choy, 2016; Özgen, 2017). Moreover, activities support active participation, collaboration, and communication. They provide opportunities for students to work cooperatively, engage in mathematical discussions with peers, defend their own ideas, and critically examine the reasoning of others (Griffin, 2009). In this way, activities contribute not only to the acquisition of content knowledge but also to the cultivation of critical and social competencies that underpin mathematical practice.

The literature on mathematical activities demonstrates a diversity of emphases. Some researchers examine the qualities that render an activity “rich” or educationally valuable (Ainley et al., 2006; Griffin, 2009), while others focus on typologies and categorisations of activities (Kieran et al., 2015). These differing emphases also produce variation in how the term activity is defined. Indeed, many definitions of activity are grounded in the notion of academic tasks (Doyle, 1988; Doyle & Carter, 1984). For the purposes of this study, we adopt an operational definition aligned with Bozkurt et al. (2022), who conceptualise mathematical activity as a structured yet flexible learning environment that allows students to engage in exploration, reasoning, and reflection. Building on this conceptualisation, the present study employs the Activity Evaluation and Feedback Tool (AEFT) to examine the pedagogical and mathematical quality of AI-generated activities. Within the AEFT framework, the construct of Mathematical Potential is operationalised through three interrelated components: mathematical focus, depth, and complexity. These components correspond closely to the analytical dimensions described in mathematics education literature, where focus captures the clarity and centrality of the mathematical idea, depth reflects the level of reasoning and conceptual connections promoted by the task, and complexity denotes the structural and procedural richness required to reach a solution (Bozkurt et al., 2022). Together, these dimensions enable a multi-layered evaluation of the cognitive demand inherent in AI-generated activities. By aligning AEFT’s evaluative structure with established frameworks of mathematical task quality, this study establishes a coherent bridge between pedagogical design and mathematical sophistication. This theoretical grounding directly informs the study’s aim to examine how prompt types and AI systems influence the quality of designed activities. This operational framing is particularly relevant to the evaluation of AI-generated activities, as it provides a consistent lens through which to examine how tools such as ChatGPT and Gemini, when guided by different types of prompts, can shape the pedagogical and cognitive affordances of the resulting tasks.

Prompts Types and Use

Prompts are the key interface between human cognition and AI output (Gattupalli et al., 2023). Recent research has shown that prompt engineering influences not only accuracy but also creativity and depth of AI-generated tasks (Shi et al., 2020). In mathematics education, prompts can be formulated as requests (e.g., “Could you design a problem about fractions?”) or commands (e.g., “Generate a problem about fractions”).

Figure 1
Examples of Request and Command Prompts for Designing a Fraction Problem

Prompt Type	Illustrative Prompt
ChatGPT Request	<p>Could you design a problem about fractions?</p> <p>Here's a problem you can use about fractions — it's contextual, clear, and can be adapted to different grade levels:</p> <p>---</p>
Gemini Request	<p>Could you design a problem about fractions?</p> <p>◆ I'd be happy to. Here is a problem about fractions.</p> <p>---</p>
ChatGPT Command	<p>Generate a problem about fractions</p> <p>Here's a problem about fractions you could use:</p> <p>---</p>
Gemini Command	<p>Generate a problem about fractions</p> <p>◆ On a Saturday morning, Marcus decides to bake cookies. He has a bag of flour that contains</p> <p>---</p>

As illustrated in Figure 1, the same mathematical focus (fractions) leads to different outputs depending on whether the AI is engaged through a request or a command. The prompt type distinction is not merely linguistic; request-style prompts may position the AI as a collaborative partner, while command-style prompts frame it as a tool for execution. These subtle differences affect the type of problem generated, the flexibility of solutions, and the scope for student agency. If requests encourage dialogic exploration while commands prioritise efficiency, the resulting tasks may differ in complexity, creativity, and appropriateness for classroom use. Studies highlight that iterative, open-ended, and contextually rich prompt foster deeper engagement with mathematical structures (Saralar-Aras & Cicek Schoenberg, 2024).

Much of the empirical evidence still concentrates on single tools such as ChatGPT, with limited attention to alternative generative AI tools like Gemini (Opesemowo & Adewuyi, 2024; Yi et al., 2024). Although prompt design is widely acknowledged as central to effective AI use, little is known about how variations in prompt types (request versus command) shape the quality and structure of mathematics activities. Moreover, comparative evaluations of AI-supported activities remain scarce, particularly regarding their pedagogical value in authentic educational contexts. This study addresses these gaps by systematically evaluating four mathematics activities developed through different AI tools (ChatGPT and Gemini) and prompt types (request vs. command).

Method

Research Design

This study employed a qualitative research approach, specifically a case study design. Case studies enable the examination of phenomena within their real-life contexts and aim to provide an in-depth understanding rather than broad generalisations

(Creswell, 2017; Yin, 2014). The primary purpose of case studies is to capture how a particular issue or phenomenon unfolds in its natural setting through rich and detailed descriptions (Chmiliar, 2010; Seggie & Bayyurt, 2017). A multiple-case study design was adopted, as each AI model and type of prompt was treated as a distinct “case.” This design was chosen because it allows for a comparative analysis of similarities and differences across cases, thereby yielding a more comprehensive perspective on the phenomenon under investigation.

Data Collection Process

The data were generated using two advanced AI models, ChatGPT-5 and Gemini 2.5 Pro, both of which represent the most recent and sophisticated versions of their respective systems. The selection of these models ensured the inclusion of state-of-the-art generative AI tools within the study. These two systems were intentionally chosen to capture variation across distinct architectures and interaction styles, as they represent leading examples of large-language-model-based environments widely used in educational research. The design thus adopted a 2×2 structure combining model type (ChatGPT-5 vs. Gemini 2.5 Pro) with prompt type (Command vs. Request), enabling a systematic comparison of four cases: ChatGPT-C, ChatGPT-R, Gemini-C, and Gemini-R.

To guide the effective use of the AI tools, the researchers designed a draft activity template consisting of four sections: formal information, preparation, implementation, and conclusion. The formal section included key elements such as course title, grade level, theme, learning outcome, suggested duration, objectives, instructional methods and techniques, materials, and activity title. To maintain a consistent focus across the study, the sections for course title, grade level, theme, and learning outcome were pre-filled in the template. The final version of the template was validated through expert review. Specifically, a researcher holding a doctoral degree in mathematics education and responsible for teaching the course “Activity Development in Mathematics Education” evaluated the template and provided feedback, upon which the final adjustments were made (see Appendix 1).

All data generation occurred within the native web interfaces of each model to ensure ecological validity. Activities were produced using the ChatGPT web interface (OpenAI, October 2025 version) and Gemini Advanced (Google DeepMind, accessed October 10 2025). No plug-ins, third-party tools, or external applications were enabled. Generation parameters were standardized across all sessions (temperature = 0.7, top-p = 0.9, max_tokens = 1200) to ensure comparability between outputs. Each activity was generated in an independent chat session to prevent cross-model contamination, and interface language settings were held constant. The data were collected by one of the researchers. For each activity, a new chat session was initiated, and the AI model was first provided with the activity template. After the model reviewed the template, the prompting process began. All prompts had been predetermined by the researchers prior to data collection and were not altered in response to the AI outputs. The prompts were developed based on a review of the relevant literature on AI-supported mathematics activity design (Şimşek, 2025; Tapan Broutin, 2024) and through expert consultation. Informed by these sources, a total of 18 prompts were constructed. The prompt set was divided into two types, Request (R) and Command (C). While R prompts framed the

tasks as interrogative requests, C prompts expressed the same tasks in imperative form. This case structure directly reflected the study's comparative design (model \times prompt type), aligning with the analytical framework described above. Using these prompt sets, four distinct activity designs were generated: Gemini (R), Gemini (C), ChatGPT (R), and ChatGPT (C). The AI-generated responses were then transferred into the activity template and saved as four separate files (see Appendix 2 and Appendix 3). The full prompt sequences and corresponding model outputs were subsequently included in Appendices 4–7, allowing readers to review the complete data used in the evaluation.

Data Analysis

Descriptive analysis was employed to examine the collected data, aiming to identify and interpret relationships emerging from the activities (Creswell, 2017). For this purpose, the AEFT, developed by Bozkurt et al. (2022), was adopted as the analytical framework (see Appendix 8). The AEFT combines both analytical and holistic perspectives: it enables a detailed evaluation by breaking down activity texts into specific components while also supporting an integrative assessment of the activity as a whole (Yin, 2014) as seen in Table 1.

Table 1

Summary of AEFT Variables, Score Ranges, and Calculation Notes

Variable/ Component	Dimension	Score Range	Description / Calculation Note
Intended Outcome	Activity Text	0–3	Clarity and attainability of the intended goal
Material	Activity Text	0–3	Functionality and appropriateness of materials
Instructions	Activity Text	0–3	Coherence and completeness of procedural guidance
Responsibility	Activity Text	0–3	Degree of student agency encouraged
Inclusivity	Activity Text	0–3	Opportunities provided for all students
Depth	Mathematical Potential	0–3	Level of reasoning and conceptual richness
Complexity	Mathematical Potential	0–3	Structural and procedural demand of the task
Mathematical Focus	Mathematical Potential	0–3	Centrality and clarity of the mathematical idea
Total AEFT Score	-	0–24	Sum of all components; higher scores indicate higher activity quality

Table 1 provides an overview of the AEFT framework, which consists of eight components: intended outcome/product, materials, instructions, responsibility, inclusivity, depth, complexity, and mathematical focus. Each component was scored using a four-point scale (0=Very Low, 1=Low, 2=Moderate, 3=High). This dual capacity to provide both fine-grained and comprehensive insights made AEFT an appropriate tool for the present study.

The activities were independently evaluated by one of the researchers and another researcher in mathematics education holding a doctoral degree. Prior to the

formal evaluation, both raters participated in a calibration session to ensure a consistent interpretation of the AEFT rubric. This session involved jointly scoring three sample activities and discussing the application of each criterion until consensus was reached. Following the independent evaluations, inter-rater reliability was calculated as 84.4%, exceeding the commonly accepted threshold of 80% for qualitative analysis (Miles & Huberman, 1994). In addition, Cohen's κ was computed as .81, indicating a strong level of agreement between the raters (Cohen, 1960). Any discrepancies identified during the scoring process were subsequently discussed until full consensus was achieved. To support content validity, expert review was conducted by two mathematics education specialists. These experts examined the suitability, clarity, and conceptual alignment of the AEFT framework with the aims of the study. Their feedback confirmed the appropriateness of the evaluation components and the overall analytical structure used to assess the AI-generated activities. Minor clarifications suggested by the experts were considered in refining the evaluation process, thereby strengthening the coherence and transparency of the analysis. Based on the total scores obtained from the AEFT, the activities were categorised into three levels: low (0–8 points), medium (9–16 points), and high (17–24 points).

Ethical Considerations

This study was based on data generated through AI models rather than human participants. Therefore, ethical approval from an institutional review board was not required.

Findings

This section presents the results derived from the systematic evaluation of the AI-generated mathematics activities, which were examined through three main stages. First, the activities were generated by the AI models according to the predefined prompt sets. Second, the outputs were cleaned and normalised by removing formatting inconsistencies (e.g., bullet structure, redundant headings) to ensure comparability across cases. Finally, the normalised texts were evaluated using the AEFT framework through independent expert scoring. The findings obtained from the analysis of the AI-designed activities within the AEFT framework are presented in Table 2.

Table 2

Results of the AEFT Evaluation Across AI-Prompt Combinations

Evaluation Dimensions	Evaluation Components	Gemini (R)	Gemini (C)	ChatGPT (R)	ChatGPT (C)
Activity Text	Intended Outcome	3	3	3	3
	Material	3	3	3	3
	Instructions	2	2	1	2
	Responsibility	3	1	1	3
	Inclusivity	3	3	2	3
Mathematical	Depth	3	3	2	3

Potential	Complexity	3	2	1	3
	Mathematical Focus	2	2	2	3
Score*		22	19	15	23

* 0–8 points = low level; 9–16 points = medium level; 17–24 points = high level

As shown in Table 2, three of the AI-designed mathematics activities were evaluated at a high-quality level, while only the ChatGPT activity generated with R-type prompts reached a medium level, though its score was close to the high threshold. The highest AEFT score was obtained by the ChatGPT activity with C-type prompts, whereas the lowest score was associated with the ChatGPT activity with R prompts. Both Gemini-generated activities, regardless of prompt type, achieved a high level. Comparisons across tools and prompt types revealed that Gemini performed better than ChatGPT with R prompts, while ChatGPT outperformed Gemini with C prompts.

At the component level, distinct strengths and weaknesses emerged. Within the activity text dimension, instructions and responsibility received lower scores in the ChatGPT-R activity, and responsibility was also rated lower in the Gemini-C activity. For instance, in the Gemini-C activity, students’ roles were vaguely defined (“students use GeoGebra or concrete materials to connect two parallel lines with two transversals”), whereas the Gemini-R version specified explicit role distribution (“group spokespersons present findings and each student answers an individual justification question”), ensuring active participation and clear accountability (see Table 3).

Table 3
Examples of Low- and High-Level Scoring for the Responsibility Component (Gemini)

Component	Type	Score	Response	Description
Responsibility	Gemini (C)	1	Participating in the activity by following the teacher’s guidance.	Student responsibilities are not clearly articulated. The descriptions are generally superficial (e.g., “answers the questions”), and the tasks expected from students during the activity are not outlined step by step. Moreover, the activity is not structured to promote students’ active participation throughout the process.
			Drawing figures formed by parallel lines and transversals.	
			Making conjectures about the angle and side properties of the figures.	
			Measuring angles and sides using a ruler, set square, or appropriate software.	
			Exchanging ideas with group members and discussing different viewpoints.	
			Collaboratively solving the questions on the worksheet.	
Sharing their suggestions for the classification of quadrilaterals with the class.				

Gemini (R)	3	<p>Arriving to the lesson punctually and prepared, bringing the necessary materials (e.g., ruler, set square, paper).</p> <p>Listening carefully to the teacher's instructions and completing assigned tasks on time.</p> <p>Taking an active role within the group rather than remaining only as an observer.</p> <p>Dividing tasks among members (e.g., drawing, measuring, note-taking, presenting).</p>	<p>Student responsibilities are clearly articulated, specifying, for example, who should bring which materials and who will assume the roles of drawer, measurer, and presenter. The activity is structured to promote students' active participation throughout the process, encouraging them to take initiative within their groups and contribute meaningfully to discussions.</p>
		<p>Respecting group members' ideas and contributing actively to discussions.</p> <p>Recording conjectures and comparing them with measurement results.</p> <p>Being open to revising conjectures when inaccuracies are identified.</p> <p>Sharing the group's findings with the class.</p> <p>Responding thoughtfully to peers' questions or critiques.</p> <p>Contributing to the formulation of collective generalizations.</p>	

In the mathematical potential dimension, depth demonstrated clear differences between ChatGPT-generated activities. As illustrated in Table 4, the ChatGPT-R activity included surface-level reasoning questions (e.g., “Can a quadrilateral have equal opposite angles but not be a parallelogram?”), whereas ChatGPT-C encouraged deeper interpretation and explanation (“Why are opposite angles in a parallelogram equal?” or “Under what conditions can two transversals create a square?”), thereby fostering higher-order mathematical reasoning.

Table 4
Examples of Medium- and High-Level Scoring for the Depth Component (ChatGPT).

Component	Type	Score	Response	Description
Depth	ChatGPT (C)	3	<p>Why are opposite angles in a parallelogram equal? Explain using a diagram and the relationships between alternate interior angles.</p> <p>Illustrate the relationships among parallelograms, rectangles, rhombuses, and squares using a schematic diagram.</p> <p>When you draw transversals at 30°, 45°, and 60° angles, which quadrilaterals are formed? Do the sums of their interior angles or their geometric properties change?</p> <p>If two transversals intersect two parallel lines at different angles, is it possible to obtain a square? Under what conditions could this occur?</p>	<p>The objective is to encourage detailed elaboration grounded in interpretation (e.g., “Why is it equal?”; “Explain.”). Furthermore, the aims include the substantiation of knowledge (e.g., “Under what conditions does this occur?”) and the identification of relationships (e.g., “Illustrate the relationships using a schematic diagram.”).</p>
	ChatGPT(R)	2	<p>Can a quadrilateral have equal opposite angles but not be a parallelogram?</p> <p>Is every trapezoid a parallelogram?</p> <p>The teacher provides a property (e.g., “opposite sides are parallel”), and students deduce which quadrilateral(s) satisfy that condition.</p>	<p>The objective here is to apply the mathematical properties of quadrilaterals, parallelograms, and trapezoids.</p>

The AI-generated responses related to the intended outcome/product component are provided in Table 5.

Table 5
AI-Generated Responses for the Intended Outcome/Product Component of the Designed Activities

Component	Type	Score	Response
Intended Outcome	Gemini (R)	3	<p>Students create their own quadrilaterals using concrete and digital materials, discover their basic properties (sides, angles, area), and develop logical generalisations based on these discoveries. They then use this knowledge to solve problems oriented toward specific goals.</p>
	Gemini (C)	3	<p>Students explore the fundamental geometric properties of quadrilaterals and their hierarchical relationships, integrate this knowledge with concepts such as area, perimeter, and ratio–proportion, and apply it to solve a complex, real-life–like problem with multiple constraints. In doing so, they enhance their skills in mathematical modelling and reasoning.</p>

ChatGPT (R)	3	<ul style="list-style-type: none"> - Discover quadrilaterals formed by the intersection of two parallel lines with two transversals, - Identify the angle and side properties of these quadrilaterals, - Reach generalisations from these properties, - Formulate rule-based statements and evaluate their contributions to the classification of quadrilaterals.
ChatGPT (C)	3	Students, by using mathematical tools and technologies, discover quadrilaterals formed by the intersection of two parallel lines with two transversals, develop assumptions about their angle and side properties, and validate these through measurement and comparison. They then make logical generalisations and evaluate their contributions to the classification of quadrilaterals.

As shown in Table 5, the responses under the intended outcome/product component were presented either in plain sentences or in bullet-point form. Across all AI-designed activities, this component was consistently rated in the ‘high’ category. These findings indicate that the intended outcomes or products of the activities were clearly and explicitly conveyed in the AI-generated texts. The AI-generated responses related to the material component are provided in Table 6.

Table 6

AI-Generated Responses for the Material Component of the Designed Activities

Component	Type	Score	Response
Material	Gemini (R)	3	Projector, computer/tablet, GeoGebra software, expert menu and task cards, activity report, form, exit cards, and role cards. Alternative (without technology): pipettes, Play-Doh, ruler, protractor, and graph paper.
	Gemini (C)	3	Smart board, devices with GeoGebra installed, official task letter, joker task cards, individual control points, and peer assessment forms. Alternative: ruler, protractor, A4 paper, geometry boards, and rubber bands/pipettes/sticks.
	ChatGPT (R)	3	Computer/tablet, smart board, dynamic geometry software (GeoGebra), ruler, set square, compass, coloured pencils, worksheets, and poster paper.
	ChatGPT (C)	3	GeoGebra or other dynamic geometry software, smart board/projector, ruler, set square, graph paper, and student worksheets.

As shown in Table 6, the material component of all AI-designed activities was rated in the ‘high’ category. The suggested materials were generally necessary for implementing the activity, useful for achieving its objectives, and easily accessible in daily life. This indicates that the materials proposed by AI were functional and practical. Some of the recommended resources (e.g., task cards, peer assessment forms) were either directly generated by AI or could be designed by AI upon request. Unlike

ChatGPT, Gemini also suggested alternative materials. This aspect was not included as a formal evaluation criterion in the AEFT framework and therefore did not affect the scoring. The findings further revealed that AI tools demonstrated different levels of performance across the activity text and mathematical potential dimensions. Examples of AI-generated responses for these components are provided in Table 7.

Table 7
Examples of AI-Generated Responses for the Evaluation Components of the Designed Activities

Component	Type	Score	Response
Instructions	ChatGPT (R)	1	1. Individual Work (10 min): Students draw parallel lines, add a transversal, and discuss which quadrilateral the resulting figure resembles. 2. Group Work and Measurement and Examination (20 min): Angle and side measurements are taken, and the properties are discussed. 3. Hypothesis Development and Discussion (15 min): Students discuss the conditions for trapezoids, parallelograms, and rectangles. 4. Generalisation and Statements (10 min): Students are asked to write rule-based statements based on their measurements.
	ChatGPT (C)	2	Instructions 1. Draw two parallel lines. 2. Cut these lines with two different transversals. 3. Examine the resulting shapes. 4. List the angle and side properties. 5. Compare your hypotheses with the measurements. 6. Classify the shapes according to quadrilateral types. 7. Make generalisations about common properties. Questions for Students – Which quadrilateral(s) were formed? – Is there a relationship between the measures of opposite angles? – What can you say about opposite sides? – How do these properties contribute to the classification of quadrilaterals? Implementation Process Groups make drawings and record their measurements. The teacher circulates around the classroom and asks guiding questions. Groups present their results, and shared generalisations are written on the board. Individual Question Select one quadrilateral formed by the intersection of two transversals with two parallel lines. Explain its angle and side properties, and discuss its significance in classification.
Responsibility	Gemini (R)	1	Students, working in groups, use GeoGebra or concrete materials (e.g., geometry board, ruler) to connect two parallel lines with two transversals.

Gemini (C)	3	<p>Group spokespersons share with the class the most interesting finding they discovered during the main task or in the expert menu. As a whole class, a generalisation is then made regarding the properties and relationships of quadrilaterals.</p> <p>Each student individually answers the following question: ‘To construct a perfect square, how should two transversals be positioned relative to two parallel lines? Justify your answer by considering both angle and side properties.’ (Students who have difficulty writing are offered the option of providing an oral explanation.)</p>
------------	---	---

As shown in Table 7, the instructions for the ChatGPT-R activity were rated at a low level. In contrast, the ChatGPT activity with C-type prompts organised the text under clear headings (instructions, questions for students, implementation process and individual questions) and provided detailed steps and question examples. However, some redundancy was observed, particularly where individual questions repeated earlier instructions. Thus, while the instructions were generally clear and supportive of achieving the intended outcome, the repetition lowered efficiency, and the activity was rated at a medium level. For Gemini, the C-type prompt activity included the phrase “students divided into groups,” which lacked clarity regarding whether responsibility applied to each individual or a representative, leading to ambiguity and a low rating for responsibility. By contrast, the Gemini-R activity explicitly stated that group spokespersons would present findings and that each student would answer an individual question, ensuring active participation and clear role distribution. This resulted in a high rating for responsibility. Finally, analysis of the mathematical potential revealed that the ChatGPT-R activity included concept connections, but these remained superficial, leading to a low score for complexity. Conversely, the ChatGPT-C activity incorporated prompts such as examine, list, compare, classify, and generalise, which encouraged interpretation, elaboration, and generalisation of mathematical knowledge. Consequently, this activity was rated high in depth and also demonstrated a strong mathematical focus, as students were explicitly required to generate reasoning around mathematical properties.

An analysis of the Mathematical Potential dimension revealed notable differences across depth, complexity, and mathematical focus. Activities generated by Gemini consistently performed at a high level in both request (R) and command (C) prompt types, with scores of 2 or above across all components. Particularly in the depth dimension, Gemini achieved the highest rating in both prompt conditions, indicating its capacity to support tasks that encourage students to interpret, elaborate, and generalise mathematical knowledge. Similarly, the ChatGPT-C activity also demonstrated strong mathematical potential, achieving high scores across all three components and producing structured tasks that guided learners toward reasoning and generalisation. In contrast, the ChatGPT-R activity was rated lower, with especially weak performance in complexity (1) and only moderate outcomes in depth (2) and mathematical focus (2). This suggests that request-style prompts tend to elicit more superficial mathematical connections from ChatGPT, whereas command-style prompts foster more structured and mathematically grounded outputs. Taken together, these findings highlight that the mathematical potential of AI-generated activities is not determined solely by the tool itself but is significantly shaped by the interaction between model and prompt type. In particular, the dimensions of depth and complexity appear to play a decisive role in

differentiating the pedagogical richness of the designed activities, underscoring the importance of deliberate prompt engineering when aiming to cultivate higher-order mathematical reasoning.

Discussion and Conclusion

This study investigated the relationship between prompt types and the quality of mathematics activities designed with AI tools. In this regard, the quality of activities produced with the support of ChatGPT and Gemini was examined according to two prompt types (command and request). The findings revealed that mathematics activities designed with different AI tools and prompt types were, in general, of high quality. However, the results also indicate that this high quality is not uniform but depends on a decisive interaction between the prompt type and the AI model used. This outcome is consistent with earlier research (Tapan Broutin, 2024; Trust et al., 2023), which also demonstrated that AI tools can assist educators in lesson planning and function as supportive instruments in instructional design (Yanar & Ergene, 2025). However, several studies have urged caution (Cooper, 2023; Halaweh, 2023; Rahman & Watanobe, 2023; Stojanov, 2023), emphasising that AI-generated lesson plans may contain fundamental errors in both the planning and implementation phases. Consequently, these studies highlight the importance of critically reviewing AI outputs before they are adopted for educational purposes. Taken together, this dual perspective reflects both the opportunities and risks of integrating generative AI into educational design. It also points to the need to understand not just whether AI can generate activities, but how prompt structures and tool differences shape the quality of those activities.

The present study extends this discussion by examining how the interaction between AI models and prompt types shapes the quality of mathematics activities. The findings showed that ChatGPT produced the highest-quality activity with command (C) prompts, while its lowest-quality activity emerged with request (R) prompts. Although the ChatGPT-R activity was categorized as medium quality, it still approached the threshold of high quality, underscoring the decisive role of prompt type in shaping ChatGPT's performance. By contrast, Gemini achieved higher scores with R prompts than with C prompts, and both Gemini-generated activities were rated as high quality. Prompt framing appears to mediate the balance between procedural precision and contextual elaboration, as command prompts yield structured reasoning while request prompts elicit broader but less focused discourse (Correia et al., 2025). This demonstrates that AI models differ fundamentally in how they interpret and respond to prompt structures. In parallel, Şimşek (2025) found that users' preferences for prompt types in mathematics activity design resulted in outputs of varying quality, influenced by metaphorical perceptions and prompting skills. Our findings move this argument forward by showing that high-quality outputs depend not only on users' ability to craft effective prompts but also on the distinctive ways in which different AI models respond to those prompts. These divergences can be interpreted through differences in how each model processes linguistic framing (Franceschelli & Musolesi, 2024). Command prompts tend to elicit procedural, list-like structures, reflecting the model's orientation toward task execution. Request prompts, on the other hand, invite more conversational and context-sensitive reasoning, often resulting in richer but less focused outputs. This

contrast aligns with previous observations that LLMs respond differently to directive versus dialogic prompt framings, influencing both creativity and coherence in educational contexts (Saralar-Aras & Cicek Schoenberg, 2024). These results highlight the importance of developing teachers' prompt literacy, which includes both the technical capacity to construct effective prompts and the pedagogical judgment to align them with instructional aims (Gattupalli, et al., 2023).

Beyond the overall quality, the analysis revealed that specific AI-prompt combinations displayed distinctive pedagogical weaknesses at the component level. For instance, Gemini-C performed poorly in responsibility, while ChatGPT-R was weaker in instructions, responsibility, and complexity. These findings suggest that in AI-supported mathematics activity design there is no single best tool or prompt type; rather, the choice of AI-prompt combinations should be made strategically in alignment with the targeted pedagogical purpose. This underscores the continuing necessity of teacher involvement in the design process. Prior research (Dumlu et al., 2024; Tapan Broutin, 2024; Trust et al., 2023; Yanar & Ergene, 2025) similarly argues that AI tools cannot fully replace teachers, but can serve as valuable aids. Conversely, in components such as intended outcome/product and material, all AI-prompt combinations produced high-quality responses, suggesting that AI tools are particularly effective at articulating learning outcomes and identifying resources aligned with these outcomes. In this context, the teacher's role in AI-supported content design goes beyond intervention to encompass that of a critical filter (Şimşek, 2025; Wardat et al., 2023). This pattern resonates with the literature on "rich mathematical activities" (Ainley et al., 2006; Griffin, 2009), where clarity of instructions, student responsibility, and task complexity are considered crucial. While AI-generated activities captured outcomes and materials well, they often fell short in ensuring depth and responsibility, underscoring the irreplaceable role of human mediation.

Given the rapid proliferation of AI tools in education, comparing their performance and outputs is increasingly important (Ergene & Caylan Ergene, 2025b). For teachers, such comparisons provide valuable insights into which tools and prompt strategies are most effective for supporting specific pedagogical aims. Since no single AI tool can be considered universally superior, evaluating the strengths and weaknesses of different systems becomes critical for informed classroom use (Cooper, 2023; Rahman & Watanobe, 2023). In this study, comparisons between ChatGPT and Gemini illustrate how different models may respond divergently to the same prompt types, resulting in variations in activity quality. While the scope of this research was limited to two prominent tools, these findings suggest that systematic evaluations across multiple AI systems could guide educators in making strategic adoption decisions. This echoes recent calls for evidence-based and comparative approaches to integrating generative AI in education, where teachers' professional judgment is supported by empirical research (Holmes et al., 2022). By addressing the "ChatGPT-dominant" trend in current research (Opesemowo & Adewuyi, 2024), this study also highlights the value of including alternative systems like Gemini, which in some cases demonstrated stronger performance.

The findings further suggest that teachers should view AI-supported activity design as a collaborative process in which AI tools serve as idea generators rather than replacements. Teachers' roles as critical evaluators remain essential, requiring them to

strategically select tools and prompt types that align with pedagogical objectives, while adapting AI-generated content to classroom realities. Comparative insights between ChatGPT and Gemini demonstrate that different models respond divergently to the same prompt structures, shaping the clarity, depth, and mathematical focus of activities. From a practice perspective, this highlights the need for professional development programs that explicitly cultivate teachers' prompt literacy. Prompt literacy encompasses both technical skills in prompt engineering and pedagogical sensitivity in aligning prompts with instructional goals (Gattupalli, et al., 2023). Strengthening teachers' prompt literacy will enable them to harness AI tools more effectively and critically, ensuring that these technologies are used strategically rather than superficially in activity design. Overall, these differences reveal that model-prompt interactions are not merely technical but epistemological: they shape how mathematical ideas are represented, structured, and communicated. From a research perspective, the study demonstrates that activity quality depends on the interaction between the AI model and the prompt type, which calls for broader comparative investigations across diverse tools, prompt structures, and instructional contexts. Future research should examine not only how AI-supported design influences classroom implementation and student learning over time but also how targeted prompt training programs enhance teachers' capacity to engage critically with AI. Embedding prompt literacy into both practice and research can thus contribute to more sustainable, equitable, and pedagogically grounded integration of AI in mathematics education.

Acknowledgements

The authors received no financial support for the research, authorship, and/or publication of this article.

Statement of Responsibility

M. Z. A.: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft. B. Ç. E.: Conceptualization, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. Ö. E.: Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization.

Conflicts of Interest

The authors declare that there is no conflict of interest related to this study.

Author Bios:

Dr. Mustafa Zeki Aydoğdu completed his undergraduate, master's, and doctoral degrees at Dokuz Eylül University. After working as a teacher at the Ministry of National Education, he joined the Faculty of Education at Trakya University in 2022, where he now serves as a faculty member in the Department of Mathematics and Science Education. His primary research interests include problem-posing, problem-solving strategies, out-of-school learning environments, and the integration of technology in mathematics education. He has authored several articles published in prominent academic journals and has presented numerous papers at national and international conferences. Dr. Aydoğdu is also the editor of the book *Out-of-School*

Learning Environments in Mathematics Education, which reflects his expertise in the field. He is actively involved in projects that contribute to the professional development of mathematics teachers and teaches courses on the fundamentals of mathematics, curriculum analysis, and out-of-school learning environments.

Büşra Çaylan Ergene is an Assistant Professor of Mathematics Education at Sakarya University, Turkey. She completed her undergraduate studies in elementary mathematics teaching and later pursued her graduate and doctoral degrees in mathematics education at Middle East Technical University, earning her PhD in 2024. Her research interests include teacher education, professional development, the teaching of geometry and measurement, and mathematical thinking. Recently, her work has also focused on the role of artificial intelligence in mathematics education, particularly in supporting teachers' instructional practices and students' problem-solving processes. She has published research in respected international journals, contributed to national and international conferences, and collaborated in research projects addressing innovative approaches to mathematics teaching. Through her scholarship, she seeks to advance the field of mathematics education by connecting classroom practice with emerging technologies and by supporting teachers in fostering critical, creative, and reflective mathematical thinking among learners.

Özkan Ergene is an Associate Professor in the Department of Mathematics Education at Sakarya University, Turkey. He received his undergraduate, master's, and doctoral degrees in mathematics education from Marmara University, completing his PhD in 2019. His academic work focuses on mathematics education at the university level, teacher education, problem solving and problem posing, and the integration of digital technologies into teaching and learning processes. In recent years, he has paid particular attention to the potential of artificial intelligence in mathematics education, exploring how generative AI tools such as ChatGPT can support instructional design, problem-posing practices, and student engagement. He has authored and co-authored numerous articles in peer-reviewed journals, presented his work at international conferences, and contributed to collaborative projects in mathematics education. His research aims to bridge theoretical insights and practical applications, providing innovative perspectives for improving the teaching and learning of mathematics.

References

- Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal*, 32(1), 2338. <https://doi.org/10.1080/01411920500401971>
- Bozkurt, A., Özmantar M. F., Agaç, G. & Güzel, M. (2022). *Matematik öğretiminde etkinlik tasarımı ve uygulamaları: Bir değerlendirme çerçevesi*. Pegem Akademi.
- Chmiliar, I. (2010). Multiple-case designs. In A. J. Mills, G. Eurepas & E. Wiebe (Eds.), *Encyclopedia of case study research* (pp. 582–583). SAGE Publications.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>

- Choy, B. H. (2016). Snapshots of mathematics teacher noticing during task design. *Mathematics Education Research Journal*, 28(3), 421-440. <https://doi.org/10.1007/s13394-016-0173-3>
- Cooper, G. (2023). Examining science education in chatgpt: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32(3), 444-452. <https://doi.org/10.1007/s10956-023-10039-y>
- Correia, A. P., Hickey, S., & Xu, F. (2025). Realizing the possibilities of the large language models: Strategies for prompt engineering in educational inquiries. *Theory Into Practice*, 64(4), 434-447. <https://doi.org/10.1080/00405841.2025.2528545>
- Creswell, J. W. (2017). *Eğitim Araştırmaları. Nicel ve Nitel Araştırmanın Planlanması Yürütülmesi ve Değerlendirilmesi*. Çev. Ed. Halil Ekşi. EDAM Yayıncılık.
- Dumlu, B. Ö., Gezer, E., & Yıldız, B. (2024). Examination of lesson plans prepared with ChatGPT on inequality. *The Journal of Turkish Educational Sciences*, 22(1), 337-358. <https://doi.org/10.37217/tebd.1338959>
- Ergene, O., Caylan-Ergene, B. (2025a) Pre-service mathematics teachers' and engineering students' perceptions of ChatGPT in mathematics: Development, validation and implementation study. *Digital Experience in Mathematics Education*, 1-27. <https://doi.org/10.1007/s40751-025-00176-x>
- Ergene, O., & Caylan Ergene, B. (2025b). AI ChatBots' solutions to mathematical problems in interactive e-textbooks: Affordances and constraints from the eyes of students and teachers. *Education and Information Technologies*, 30(1), 509-545. <https://doi.org/10.1007/s10639-024-13121-z>
- Franceschelli, G., & Musolesi, M. (2024). On the creativity of large language models. *AI & Society*, 40, 3785-3795. <https://doi.org/10.1007/s00146-024-02127-3>
- Gattupalli, S., Maloy, R. W., & Edwards, S. A. (2023). Prompt literacy: A pivotal educational skill in the age of AI. *Educational Technology & Society*, 1-6. <https://hdl.handle.net/20.500.14394/22905>
- Griffin, P. (2009). What makes a rich task? *Mathematics Teaching*, 212, 32-34.
- Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, 15(2), ep421. <https://doi.org/10.30935/cedtech/13036>
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., ... & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 32(3), 504-526. <https://doi.org/10.1007/s40593-021-00239-1>
- Hossain, K. A. (2023). Analysis of present and future use of artificial intelligence (ai) in line of fourth industrial revolution (4ir). *Scientific Research Journal*, 11(8), 1-50.
- Hwang, G. J., & Tu, Y. F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics*, 9(6), 584. <https://doi.org/10.3390/math9060584>
- Jin, H. S., & Suh, B. (2024). Research on a statistics education program utilizing deep learning predictions in high school mathematics. *The Mathematical Education*, 63(2), 209-231. <https://doi.org/10.63311/mathedu.25.64311>

- Jones, K., & Pepin, B. (2016). Research on mathematics teachers as partners in task design. *Journal of Mathematics Teacher Education*, 19(2), 105-121. <https://doi.org/10.1007/s10857-016-9345-z>
- Kieran, C., Doorman, M., & Ohtani, M. (2015). Frameworks and principles for task design. Anne, W. & Minoru O. (Eds.), *Task design in mathematics education* (pp. 19–81). Springer International Publishing.
- Kim, J. (2024). An analysis of perceptions of elementary and secondary mathematics teachers on the use of artificial intelligence in mathematics education. *The Mathematical Education*, 63(2), 351–368. <https://doi.org/10.7468/mathedu.2024.63.2.351>
- Li, M. (2025). Integrating artificial intelligence in primary mathematics education: Investigating internal and external influences on teacher adoption. *International Journal of Science and Mathematics Education*, 1283–1308. <https://doi.org/10.1007/s10763-024-10515-w>
- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks: Sage.
- Moorhouse, B. L., Wong, K. M., & Li, L. (2023). Teaching with technology in the post-pandemic digital age: Technological normalisation and AI-induced disruptions. *Relc Journal*, 54(2), 311–320.
- Opesemowo, O. A. G., & Adewuyi, H. O. (2024). A systematic review of artificial intelligence in mathematics education: The emergence of 4IR. *EURASIA Journal of Mathematics, Science and Technology Education*, 20(7), em2478. <https://doi.org/10.29333/ejmste/14762>
- Özgen, K. (2017). Matematiksel öğrenme etkinliği türlerine yönelik kuramsal bir çalışma: fonksiyon kavramı örnekleme [A Theoretical Study on Mathematical Learning Activities: A Case Of Function Concept]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 17(3), 1437–1464. <https://doi.org/10.17240/aibuefd.2017.17.31178-338839>
- Park, H. J., & Manley, S. (2024). Using ChatGPT as a proof assistant in a mathematics pathways course. *The Mathematical Education*, 63(2), 139–163.
- Rahman, M. M. & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, 13(9), 1–21. <https://doi.org/10.3390/app13095783>
- Saralar-Aras, I., & Cicek Schoenberg, Y. (2024). Unveiling the synergistic nexus: AI-driven coding integration in mathematics education for enhanced computational thinking and problem-solving. *The Mathematical Education*, 63(2), 233–254. <https://doi.org/10.7468/mathedu.2024.63.2.233>
- Seggie, N. ve Bayyurt, Y. (2017). *Nitel araştırma yöntem, teknik, analiz ve yaklaşımlar* (2nd Edition). Anı Yayıncılık.
- Shi, B., Huang, L., & Lu, X. (2020). Effect of prompt type on test-takers' writing performance and writing strategy use in the continuation task. *Language Testing*, 37(3), 361–388. <https://doi.org/10.1177/0265532220911626>
- Shin, B., Lee, J., & Yoo, Y. (2024). Exploring automatic scoring of mathematical descriptive assessment using prompt engineering with the GPT-4 model: Focused

- on permutations and combinations. *The Mathematical Education*, 63(2), 187–207. <https://doi.org/10.63311/mathedu.2024.63.2.187>
- Şimşek, N. (2025). Integration of ChatGPT in mathematical story-focused 5E lesson planning: Teachers and pre-service teachers' interactions with ChatGPT. *Education and Information Technologies*, 30, 11391–11462. <https://doi.org/10.1007/s10639-024-13258-x>
- Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: an autoethnographic study. *International Journal of Educational Technology in Higher Education*, 20(1), 1–17. <https://doi.org/10.1186/s41239-023-00404-7>
- Sullivan, P., Askew, M., Cheeseman, J., Clarke, D., Mornane, A., & Roche, A. (2015). Supporting teachers in structuring mathematics lessons involving challenging tasks. *Journal of Mathematics Teacher Education*, 18(2), 123–140. <https://doi.org/10.1007/s10857-014-9279-2>
- Tapan Broutin, M. S. (2024). Exploring mathematics teacher candidates' instrumentation process of generative artificial intelligence for developing lesson plans. *Yükseköğretim Dergisi*, 14(1), 165-176. <https://doi.org/10.53478/yuksekogretim.1347061>
- Trust, T., Whalen, J., & Mouza, C. (2023). Editorial: ChatGPT: Challenges, opportunities, and implications for teacher education. *Contemporary Issues in Technology and Teacher Education*, 23(1).
- UNESCO. (2021). *AI and education: Guidance for policy-makers* (F. Miao, W. Holmes, R. Huang, & H. Zhang, Eds.). UNESCO. <https://doi.org/10.54675/PCSP7350>
- Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *EURASIA Journal of Mathematics, Science and Technology Education*, 19(7), em2286. <https://doi.org/10.29333/ejmste/13272>
- Yanar, A. N., & Ergene, Ö. (2025). Integrating artificial intelligence in education: How preservice mathematics teachers use ChatGPT for 5E lesson plan design. *Journal of Pedagogical Research*, 9(2), 158–176. <https://doi.org/10.33902/JPR.202533163>
- Yeo, S., Moon, J., & Kim, D. J. (2024). Transforming mathematics education with AI: Innovations, implementations, and insights. *The Mathematical Education*, 63(2), 387–392. <https://doi.org/10.7468/mathedu.2024.63.2.387>
- Yi, L., Liu, D., Jiang, T., & Xian, Y. (2025). The effectiveness of AI on K-12 students' mathematics learning: A systematic review and meta-analysis. *International Journal of Science and Mathematics Education*, 23(4), 1105–1126. <https://doi.org/10.1007/s10763-024-10499-7>
- Yin, R. K. (2014). Getting started: How to know whether and when to use the case study as a research method. *Case study research: design and methods*, 5, 2–25.

Appendix 1. Lesson Plan Template

Formal Section	Course
	Grade Level
	Theme
	Learning Outcome
	Suggested Duration
	Objective
	Teaching-Learning Methods and Techniques
	Materials and Tools Used
Preparation Phase of the Activity	Activity Title
Preparation Phase of the Activity	Attention
Implementation Phase	Execution of the Activity
Conclusion Phase	Assessment and Evaluation

Appendix 2. Request (R) Type Prompts Used in the Study

- Using the provided activity template, can you design an activity aligned with the following learning outcome:
MAT.6.3.2. Ability to make inferences about the properties of shapes formed by two parallel lines intersected by two transversals using mathematical tools and technology
 - Makes assumptions about the properties of shapes formed by two parallel lines intersected by two transversals.
 - Lists the resulting shapes according to their various properties.
 - Compares the shapes with their assumptions by considering side and angle properties.
 - Provides propositions regarding the sum of the interior angles of the shapes formed and the common properties of trapezoids, parallelograms, rhombi, rectangles, and squares.
 - Evaluates the contribution of the propositions to the classification of quadrilaterals.
- How much time do you recommend for carrying out this activity?
- Can you clearly state the objective of the activity?
- What teaching-learning methods and techniques will you use in the activity?
- Can you specify the tools, materials, and references used in the activity?
- Can you elaborate on the preparation phase (attention-getter) of the activity? (This may include a real-life example, visual, discussion question, video, etc. related to the targeted mathematical concept.)
- Can you elaborate on the implementation phase of the activity? Can you add the instructions and questions to this section? Then, can you describe the implementation process of the activity? The implementation phase may include the expected student responses, potential misconceptions, and feedback in case of difficulties.
- Can you elaborate on the conclusion phase of the activity? In this section, can you describe how you plan to assess the learning related to the targeted mathematical concept/idea? (This may be a question, problem, survey, etc.)
- What are the responsibilities of the students in the activity?
- Can you design the activity in a way that strengthens its “inclusiveness” by ensuring the participation of all students?
- Can you enhance the “mathematical depth” of the activity?
- Can you enhance the “complexity” of the activity?
- Can you make the “mathematical dimension” of the activity more explicit?
- If you identify any missing or improvable aspects of the activity, can you explain them?
- What are your suggestions for improving the missing or improvable aspects?
- Can you propose an engaging title for the activity?
- Is there anything else you would like to add or mention regarding the activity?
- Can you provide the revised final version of the activity as a single piece?

Appendix 3. Command (C) Type Prompts Used in the Study

1. Using the provided activity template, write an activity aligned with the following learning outcome:
MAT.6.3.2. Ability to make inferences about the properties of shapes formed by two parallel lines intersected by two transversals using mathematical tools and technology
 - a) Make assumptions about the properties of shapes formed by two parallel lines intersected by two transversals.
 - b) List the resulting shapes according to their various properties.
 - c) Compare the shapes with your assumptions by considering side and angle properties.
 - ç) Provide propositions regarding the sum of the interior angles of the shapes formed and the common properties of trapezoids, parallelograms, rhombi, rectangles, and squares.
 - d) Evaluate the contribution of the propositions to the classification of quadrilaterals.
2. Write the suggested duration for carrying out this activity.
3. Clearly state the objective of the activity.
4. Write the teaching–learning methods and techniques to be used in the activity.
5. Write the tools, materials, and references used in the activity.
6. Elaborate on the preparation phase (attention-getter) of the activity. (This may include a real-life example, visual, discussion question, video, etc., related to the targeted mathematical concept.)
7. Elaborate on the implementation phase of the activity. Add the instructions and questions to this section. Then, describe the implementation process of the activity. The implementation phase should include the expected student responses, potential misconceptions, and feedback in case of difficulties.
8. Elaborate on the conclusion phase of the activity. In this section, write how you plan to assess the learning related to the targeted mathematical concept/idea. (This may be a question, problem, survey, etc.)
9. Write in detail the responsibilities of the students in the activity.
10. Design the activity in a way that strengthens its “inclusiveness” by ensuring the participation of all students.
11. Develop the “mathematical depth” of the activity.
12. Develop the “complexity” of the activity.
13. Make the “mathematical dimension” of the activity explicit.
14. If you identify any missing or improvable aspects of the activity, explain them.
15. Write suggestions to improve the missing or improvable aspects.
16. Give an engaging title to the activity.
17. Add any additional remarks if necessary.
18. Provide the revised final version of the activity as a single piece.

Appendix 4. Chat Texts with Gemini 2.5 Pro and Request Prompts

Prompt Types	Chat Text with AI*
Gemini (R)	<p>Prompt: What is the duration you recommend for the implementation of this activity?</p> <p>An excellent question!</p> <p>The duration I recommend for this activity, as I also indicated in the lesson plan, is two class periods, totaling 80 minutes (40 + 40 minutes).</p> <p>The reason I recommend this duration is my desire to provide students with sufficient time, as required by the nature of the constructivist approach:</p> <p>The first 40 minutes; is generally ideal for the stage of arousing curiosity, making assumptions, and initial exploration (playing with materials or GeoGebra).</p> <p>The second 40 minutes; is allocated for steps that require more in-depth thinking, such as analyzing the collected data, inter-group discussion, reaching generalizations, and classifying quadrilaterals.</p> <p>In this way, students find the opportunity to learn without rushing the topic, by internalizing it, and most importantly, by discovering it themselves.</p>
	<p>Prompt: Could you clearly state the aim of the activity?</p> <p>Certainly, let me state the aim of the activity more clearly.</p> <p>The primary aim of this activity is to enable students to discover these properties themselves, like explorers, rather than having them memorize the properties of quadrilaterals as a list.</p> <p>To explain in more detail, we can gather the aims under these three headings:</p> <p>Discovering Through Concrete Experience: Instead of reading an abstract definition, students create two parallel lines and two transversals with their own hands using tools like GeoGebra or straws. When they change the positions of these transversals, they directly observe how the sides and angles of the resulting quadrilateral change. In this way, they “discover” by seeing for themselves, for example, why the opposite angles of a parallelogram are equal.</p> <p>Fostering Generalization Skills: They do not just create a single quadrilateral and stop. They produce many different quadrilaterals (trapezoid, parallelogram, rhombus, rectangle) by using transversals with different slopes. By comparing the data (angle and side measurements) they collect from these different examples, they arrive at general rules and propositions such as, “Apparently, the sum of the interior angles of all the quadrilaterals we created is 360 degrees,” or “When we make the opposite sides parallel, the opposite angles always equal too.”</p> <p>Making Meaningful Classification: At the end of the activity, students make inferences such as, “To create a rectangle, we needed to make the transversals parallel to each other and perpendicular to the parallel lines. This means a rectangle is a special shape whose opposite sides are parallel.” In this way, they establish the hierarchical relationship among quadrilaterals (such as every square being a rectangle, every rectangle being a parallelogram) on a logical foundation, without memorizing.</p> <p>In short, the aim is not to give them the rule “The opposite angles of a parallelogram are equal”; it is to let them experience the excitement and permanence of finding this information themselves.</p>
	<p>Prompt: What are the teaching-learning methods and techniques you will use in the activity?</p> <p>Certainly, let me explain the teaching-learning methods and techniques I will use in this activity and why I prefer them. As I also indicated in my lesson plan, this activity relies on several fundamental, intertwined techniques:</p> <p>Discovery Learning: This is the main philosophy of the activity. Instead of directly telling students the properties of quadrilaterals, I present them with a problem situation (parallel rails) and exploration tools (GeoGebra, ruler, etc.). My aim is to enable them to find the rules and relationships themselves by generating hypotheses, conducting experiments, and analysing results, just like a mathematician. Knowledge becomes much more permanent and meaningful this way.</p> <p>Collaborative Learning: The reason I have students work in groups is my belief that learning is a social process. Within the group, students listen to each other's ideas, defend their own thoughts, encounter different perspectives, and reach the solution together. This process both develops their mathematical communication skills and allows them to learn from one another.</p> <p>Question-Answer Technique: The question-answer here is not a traditional structure where I ask a question and the student gives the single correct answer. I rather ask open-ended questions that encourage thinking, such as “Why do you think so?”, “Did anyone find a different result?”, “What do we need to change to make this shape a rectangle?”. My aim is to reveal their thought processes and ensure they conduct a more in-depth analysis.</p> <p>Brainstorming: I use this technique at the very beginning of the activity when I ask “What do you see?” based on the train track photograph and ask for their hypotheses about possible shapes. The purpose is to activate students' prior knowledge, unleash their creativity, and maximize their curiosity for the lesson.</p> <p>Technology-Assisted Instruction: The reason I use dynamic geometry software like GeoGebra is that it offers students the opportunity to concretize abstract concepts and receive instant feedback. When a student changes the angle of a transversal, they see how the other angles of the quadrilateral change instantly. This dynamic interaction allows them to test their hypotheses immediately and incredibly accelerates and enriches the discovery process.</p>

Appendix 5. Chat Texts with Gemini 2.5 Pro and Command Prompts

Prompt: Write the duration you recommend for the implementation of this activity.

The duration I recommend for this activity, as I also indicated in the plan, is a total of 80 Minutes (2 Class Periods). The reason I recommend this duration is that we need to provide students with sufficient time in a constructivist activity. This duration is ideal for them to explore without rushing, discuss among themselves, test their hypotheses, and make sense of their findings. Generally, the first class period aims for exploration and application, while the second class period aims for compiling the findings and arriving at generalizations.

Prompt: Clearly state the aim of the activity.

Certainly, in the lesson plan I prepared, I stated the aim of the activity as follows: The aim of the activity is for students to discover the quadrilaterals they create with two parallel lines and two transversals using dynamic geometry software or concrete materials, to formulate and test their own hypotheses regarding the side and angle properties of these quadrilaterals, and to grasp the logic of classifying quadrilaterals based on these properties. In short, my aim is not to have them memorize rules, but to enable them to discover these rules themselves, like geometry explorers, and to understand the “why.”

Gemini
(C)

Prompt: Write the teaching-learning methods and techniques you will use in the activity.

Certainly. In this activity, in line with my constructivist philosophy, I will use a combination of various student-centered methods and techniques. As I also indicated in the lesson plan, these are:

- **Discovery Learning:** This is the core strategy of the activity. Instead of directly telling students the properties of quadrilaterals, I provide them with an exploratory environment to find these properties themselves.
 - **Collaborative Learning:** Having students work in small groups encourages them to learn from one another, exchange ideas, and co-construct knowledge.
 - **Question-Answer:** Throughout the lesson, I use critical questions such as “Why?”, “How?”, “What else could it be?” which are not based on rote answers and guide students to think and inquire.
 - **Discussion:** I deepen learning by creating an environment, both within groups and in the whole class, where students present and defend their findings and propositions, and share their ideas.
 - **Brainstorming:** At the very beginning of the activity, I use this technique to activate students' prior knowledge on the subject and to form their initial hypotheses.
-

Appendix 6. Chat Texts with ChatGPT-5 and Request Prompts

Prompt: What is the duration you recommend for the implementation of this activity?

For this type of exploration and discussion-based activity, I can recommend a total duration of approximately 2 class periods (80 minutes):
 Engaging Attention (5–10 min): Introduction to the topic and arousing curiosity with visuals from daily life.
 Application (50–55 min):
 Creating parallel lines and transversals with GeoGebra or paper-drawing (10 min)
 Making hypotheses and taking measurements (15 min)
 Listing results and group discussion (15 min)
 Writing and sharing propositions (10 min)
 Conclusion/Assessment (15 min): Groups presenting their posters/work, teacher's reinforcement.
 If the use of technology (interactive whiteboard, tablets, GeoGebra) is not common in the classroom, the same activity can be done on paper with a ruler and set square; in this case, the duration could be kept slightly longer (approximately 90 minutes).
 If you wish, I can also adapt a shorter version for you that can be implemented within a single class period (40 min). Would you like this?

Prompt: Could you clearly state the aim of the activity?

Of course, we can state the aim of the activity clearly as follows:
 Aim of the Activity
 The aim of this activity is for students to examine the quadrilaterals (trapezoid, parallelogram, rectangle, rhombus, square) formed when they intersect two parallel lines with two transversals, utilizing mathematical tools and technologies,
 to make hypotheses regarding the angle and side properties of these quadrilaterals,
 to verify or falsify their hypotheses by taking measurements,
 to develop propositions from the results they obtain, and
 to evaluate how these propositions, contribute to the classification of quadrilaterals.
 In other words, the activity aims to guide students toward discovering, experimenting, and making their own mathematical inferences, rather than providing them with ready-made knowledge.
 If you wish, I can also write this for you converted into student learning outcomes (e.g., “the student will be able to...”). Would you like that?

ChatGPT
(R)

Prompt: What are the teaching-learning methods and techniques you will use in the activity?

The teaching-learning methods and techniques we can use for the activity are as follows:
 Teaching-Learning Methods and Techniques
 Constructivist approach: To enable students to make their own discoveries and to connect their prior knowledge with new information.
 Question-answer: To encourage students to think, express their hypotheses, and discuss through open-ended questions directed at them.
 Collaborative learning: To ensure they work in small groups to make drawings, measurements, and inferences.
 Discovery learning: To allow them to create the shapes themselves and find their properties using GeoGebra or drawing tools.
 Discussion and peer learning: For groups to share their findings with the class and for different interpretations to be evaluated together.
 Technology-assisted learning: To concretize their experiences using dynamic geometry software (GeoGebra, interactive whiteboard applications, etc.).
 This combination enriches the activity in terms of both active participation and connection to daily life.
 If you wish, I can place these into the template you provided in a lesson plan format as bullet points.
 Would you like me to do that?

Appendix 7. Chat Texts with ChatGPT-5 and Command Prompts

Prompt: Write the duration you recommend for the implementation of this activity.

It is appropriate for all stages of this activity (engaging attention, group work, hypothesis–experiment–comparison, class sharing, and assessment) to be completed within 40 minutes.

Thus, the recommended duration is: 1 class period (40 minutes).

If you wish, I can also create a more detailed timeline, expanding this to 2 class periods (80 minutes).

Would you like this?

Prompt: Clearly state the aim of the activity.

Certainly. Here is the aim of the activity:

Aim: For students, utilizing mathematical tools and technologies, to discover the quadrilaterals formed by two parallel lines and two transversals, to develop hypotheses about the angle and side properties of these shapes, to reach generalizations by comparing the results they obtain, and to evaluate how these generalizations contribute to the classification of quadrilaterals.

ChatGPT
(R)

If you wish, I can also write this aim for you in a more student-centered and concise statement. Would you like that?

Prompt: Write the teaching-learning methods and techniques you will use in the activity.

Of course. The teaching-learning methods and techniques to be used in the activity are as follows:

Constructivist approach (the student actively discovering and constructing knowledge themselves)

Hypothesize–test–verify method

Group work and collaborative learning

Question–answer

Technology-assisted learning (GeoGebra, interactive whiteboard, etc.)

☞ If you wish, I can also organize this list in a concise and formal “bullet-point” manner suitable for the table in the lesson plan. Would you like me to do this?

Appendix 8. Activity Evaluation and Feedback Tool (Bozkurt et al., 2022)

Components	Rating Criteria			
	Very Low (0 points)	Low (1 point)	Medium (2 points)	High (3 points)
Intended Outcome or Product of the Activity	The intended outcome or product of the activity is not understandable.	The activity has multiple unrelated foci, leading to ambiguity about the intended outcome or product.	Although there are uncertainties about the intended outcome or product, following the instructions or completing the steps gives a partial understanding.	The intended outcome or product of the activity is clearly understandable.
Material	The material is neither functional nor useful.	The material is not functional.	The material is not useful.	The material is functional and useful.
Instructions	The instructions are not coherent or executable, making it impossible to evaluate the intended outcome of the activity.	The instructions are coherent, but problems and/or uncertainties exist in reaching the intended outcome.	The instructions are coherent and allow access to the intended outcome, but they contain extra or irrelevant steps.	The instructions are coherent, coherent, and allow access to the intended outcome.
Responsibility	Student responsibilities are not specified.	Student responsibilities are vaguely expressed with uncertainties.	Student responsibilities are clearly expressed but limit students' opportunities for active participation.	Student responsibilities are clearly expressed and allow active participation in the process.
Inclusiveness	The activity contains elements that prevent almost all students from participating.	The activity allows only certain types of students to participate.	The activity allows most students to participate, but some students remain excluded.	The activity allows all students the opportunity to participate.
Mathematical Potential of the Activity Text				
Intended Outcome/Product of the Activity	The intended outcome or product of the activity is not understandable.	The activity has multiple unrelated foci, leading to ambiguity about the intended outcome.	Although there are uncertainties about the intended outcome or product, following the instructions or completing the steps gives a partial understanding.	The intended outcome or product of the activity is clearly understandable.
Material	The material is neither functional nor useful.	The material is not functional.	The material is not useful.	The material is functional and useful.
Instructions	The instructions are not coherent or executable, making it impossible to evaluate the intended outcome of the activity.	The instructions are coherent, but problems and/or uncertainties exist in reaching the intended outcome.	The instructions are coherent and allow access to the intended outcome, but they contain extra or irrelevant steps.	The instructions are coherent, coherent, and allow access to the intended outcome.



This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0). For further information, you can refer to <https://creativecommons.org/licenses/by-nc-sa/4.0/>