Original Research Article

SAGE Open
Vol. 16(3): 1–21
© The Author(s) 2026
DOI: 10.1177/21582440261486168
journals.sagepub.com/home/sgo
Sage

Evaluation of Artificial Intelligence Performance in the Mathematical Problem-Solving Process

Seda Akti Aslan[1], Meltem Miser[2], and Jordan Allison[3]

Abstract

This study examines whether widely used chatbots can support middle-school mathematics learning by analyzing their responses to 8th-grade algebra problems in Turkish. Using a qualitative descriptive case study, we posed ten open-ended items (no visuals or symbols) to two AI models (ChatGPT-4 and Gemini) with a standardized prompt and evaluated their solutions on four expert-defined steps (problem comprehension, strategy selection, execution, and verification). An analytic rubric was developed and content-validated by five mathematics education experts; human scoring reliability was high (Cohen's $\kappa = 0.81$). To incorporate a learner perspective, ten 8th-grade students judged the clarity and usefulness of the explanations. All problem statements, prompts, and analyses were conducted in Turkish. Findings show that ChatGPT achieved 83/100 versus Gemini's 71.5/100, and that both models struggled on cognitively demanding, non-routine items. We discuss implications for instructional use and outline limitations (small sample, single grade level, minimal prompt tuning) and avenues for broader, multi-model, multi-grade evaluations.

Plain Language Summary

This study looked at whether popular artificial intelligence (AI) tools, ChatGPT and Gemini, can help middle school students with algebra problems. We gave each tool ten math problems written in Turkish and then evaluated their answers step by step. We checked if the tools understood the problem, chose the right method, carried out the solution correctly, and checked their final answer. To make the evaluation fair, we used a scoring guide developed and reviewed by mathematics education experts. In addition, ten 8th-grade students were asked if the explanations were clear and helpful. The results showed that ChatGPT performed better overall, with a score of 83 out of 100, compared to Gemini's 71.5. Both tools did well on straightforward problems but had trouble with harder questions that required deeper reasoning or multiple steps. Students reported that the AI explanations were often useful and helped them see different ways to solve a problem. However, some found the solutions too complicated or lacking enough explanation. This means that while AI tools can support learning, teachers still play an important role in guiding students and checking the accuracy of the responses. Overall, the findings suggest that AI chatbots can be a helpful resource in the classroom, especially for practicing math and exploring alternative solution methods. However, they cannot fully replace teacher support, particularly for more complex or challenging problems. Future research should look at other grade levels, more types of math, and different AI tools to better understand how these technologies can be used effectively in education.

[1]Vocational School of Health Services, Fırat University, Elazığ, Türkiye
[2]Republic of Turkey Ministry of National Education, Adıyaman, Türkiye
[3]School of Business, Computing and Social Sciences, University of Gloucestershire, Cheltenham, UK

Corresponding author:
Seda Akti Aslan, Vocational School of Health Services, Fırat University, Rectorate Campus, 23119, Elazığ, Türkiye.
Email: saaslan@firat.edu.tr

Keywords
artificial intelligence, ChatGPT, gemini, mathematics education, problem-solving

Received: 6 November 2025; Revised: 9 June 2026; Accepted: 21 August 2026

1. Introduction

In the context of 21st-century education, changing societal needs and rapid advancements in information and communication technologies significantly influence both learning and teaching processes (Dumont & Istance, 2010; Istance & Kools, 2013; Kerimbayev et al., 2025; Msambwa et al., 2024). These rapid transformations have led to the more effective use of technology in numerous fields. One such field is education, where Artificial Intelligence (AI) has emerged as a rapidly evolving component of educational technology (Engelbrecht & Borba, 2024). Recent research emphasizes that a broad range of stakeholders increasingly recognize AI as a transformative force capable of enhancing learning processes, supporting knowledge creation, and contributing to human development, while also raising critical pedagogical and ethical considerations (Ali et al., 2024). Considering the impact of AI applications across various domains on individuals' quality of life, it is expected to have a similar influence in education. Given AI's increasing presence in everyday life, it is essential to introduce and utilize it as a supportive tool from an early age. In this regard, integrating AI into instructional processes is of great importance (Lampou, 2023).

The literature contains numerous studies on AI's application across different educational fields. These studies include research in mathematics (Engelbrecht & Borba, 2024), chemistry (Atmosukarto et al., 2021), gastronomy and nursing (Chang et al., 2022). One significant finding emerging from these studies is the positive influence of artificial intelligence on learners' problem-solving and problem-analysis skills. This enhancement is primarily attributed to AI-driven educational tools that provide personalized feedback, employ simulation-based learning environments, and utilize adaptive systems capable of dynamically adjusting instructional content based on learners' individual progress and performance patterns (Zawacki-Richter et al., 2019). Problem-solving is widely regarded as one of the most essential competencies students should develop throughout the educational process, as it plays a fundamental role in cognitive development, academic achievement, and the ability to navigate real-world challenges. Within the framework of Bloom's taxonomy, problem-solving is situated at the higher-order thinking levels—namely, analysis, evaluation, and creation—while from a constructivist perspective, it reflects learners' active engagement in constructing knowledge through authentic and meaningful tasks. It serves as a teaching method, a way of thinking, and a subject that requires careful consideration. Furthermore, problem solving, which requires higher-order reasoning, holds an indispensable position in mathematics education (Engelbrecht & Borba, 2024; Yılmaz, 2023). Additionally, it contributes to students' metacognitive thinking abilities (Desoete & De Craene, 2019; Lestari et al., 2019). One of the subjects most closely related to this skill is mathematics. Emerging knowledge, evolving pedagogical approaches, and technological innovations such as dynamic geometry software (e.g., GeoGebra), computer algebra systems, artificial intelligence applications, and online collaborative platforms are continuously reshaping our perspectives on mathematics by enabling more interactive, visual, and personalized learning experiences, as well as our approaches to learning and teaching it (Ministry of National Education [MoNE], 2024). In the 21st century, technology has become an essential tool for learning mathematics (National Council of Teachers of Mathematics [NCTM], 2008). Today, technological advancements occur at a much faster pace, particularly in developed countries. However, researchers in developing countries also demonstrate an increasing interest and enthusiasm for integrating technology into mathematics education (Engelbrecht & Borba, 2024). Overall, the literature contains numerous studies on the integration of technology into mathematics instruction (Artigue, 2002; Baki, 1996; Engelbrecht & Borba, 2024; Son, 2024).

As with all technological developments, AI's integration into instructional processes has drawn researchers' attention. This is primarily because AI-powered chatbots provide real-time feedback and enhance student engagement, making them valuable tools in education. Due to these advantages, chatbots are widely used in instructional processes, particularly in Science, Technology, Engineering, and Mathematics (STEM) disciplines (Zawacki-Richter et al., 2019). Chatbots assist learning (Huang et al., 2022), support exam preparation, and respond swiftly to frequently asked questions (Antony & Ramnath, 2023). For these reasons, chatbots are increasingly recognized as popular and functional educational resources (Holmes & Tuomi, 2022).

The literature includes studies that separately address AI and problem-solving as significant topics in mathematics education (Engelbrecht & Borba, 2024). While existing studies have predominantly examined ChatGPT and Gemini in domains such as language learning, writing assistance, and general education, comparatively limited attention has been given to middle-school-level algebra problems. Whether AI can assist students in overcoming difficulties in the mathematical problem-solving process, which forms the foundation of mathematics, remains a critical issue in mathematics education

(Engelbrecht & Borba, 2024). In this context, analyzing AI performance in solving mathematical problems is expected to provide valuable insights into how AI can be more effectively integrated into instructional processes.

The literature includes studies that address AI and problem-solving as significant topics in mathematics education (Engelbrecht & Borba, 2024). In addition, an emerging body of research has begun to examine the use of ChatGPT and similar large language models in mathematical problem solving, particularly in terms of solution accuracy and reasoning quality. However, existing studies have predominantly focused on general performance outcomes or higher education contexts, and comparatively limited attention has been given to middle-school–level algebra problems. Moreover, few studies have examined AI-generated mathematical solutions using a step-based analytical framework that considers problem comprehension, strategy selection, execution, and verification, especially within non-English instructional contexts. In this respect, the present study aims to contribute to the literature by examining ChatGPT and Gemini's performance on Turkish 8th-grade algebra problems and by incorporating both expert evaluations and student perspectives on the clarity and instructional usefulness of the solutions.

This study aims to examine chatbot responses (ChatGPT and Gemini) to algebraic problems at the eighth-grade level in terms of accuracy, diversity, problem-solving strategies, and comprehensibility. To achieve this aim, the following research questions were addressed:

RQ1. How do ChatGPT-4 and Gemini perform on Turkish 8th-grade algebra problems when evaluated with an expert-validated, step-based rubric?
RQ1a. What is their accuracy at each of four steps (comprehension, strategy, execution, verification)?
RQ1b. To what extent do they propose diverse, alternative solution methods?
RQ1c. How do 8th-grade students judge the clarity/usefulness of the explanations?
RQ2. How consistent are the models' responses across user accounts under the same prompting protocol?

This paper contributes: (i) a rubric aligned with Posamentier & Krulik's problem-solving strategies, content-validated by five experts ($\kappa = 0.81$ for human scoring); (ii) a controlled comparison of ChatGPT-4 and Gemini on ten Turkish, visual-free algebra items with step-level scoring; (iii) a learner-centered clarity assessment from ten 8th-grade students; and (iv) empirical evidence identifying where current LLMs succeed and where they struggle on non-routine items, informing classroom use and future evaluations.

2. Literature Review

2.1. Problem-Solving Process

Although various definitions of the term "problem" can be found in different sources, in the most general sense, a problem is a complex or uncertain question. A problem is a subject of research, discussion, or contemplation (Van De Walle, 1989). It can also be considered a situation in which an individual feels discomfort and, as a result, feels the need to seek a solution using their knowledge and experiences (Baki, 2008). From a mathematical perspective, a problem is a question whose solution or proof is not immediately apparent, even though it is required to be found or demonstrated, making it an unclear or ambiguous issue (Grouws, 2006).

Problem-solving is defined as a process in which difficulties are addressed to eliminate uncertainties (Gelbal, 1991). It is the process by which an individual understands and resolves the gap between the obstacles encountered and the goal they aim to achieve. The construction of meaningful connections between different pieces of mathematical knowledge occurs during the problem-solving process, enabling learners to integrate concepts and procedures coherently (Engelbrecht & Borba, 2024). Problem-solving is a crucial process in education that fosters students' sense of responsibility, encourages them to conduct research, increases their interest in learning, ensures long-term retention, and enhances motivation (Fisher, 1990). Problem-solving requires both analytical and creative thinking, helping individuals understand root causes and generate effective solutions. This dual approach builds confidence in decision-making and overcoming challenges (Aliu & Aigbavboa, 2020). Additionally problem-solving skills help reduce depression and aggression while promoting prosocial behaviors and societal participation (Adrian, 2010).

The ability to solve mathematical problems extends beyond the mere recall of basic operations or the application of memorized procedures; it requires higher-order thinking skills such as analysis, evaluation, and creation, as outlined in Bloom's taxonomy. It is a long and gradually developing process. In this process, mathematical concepts and operations must be integrated and employed together (Engelbrecht & Borba, 2024). Simply reaching the final answer of a mathematical problem does not constitute problem-solving. For it to be considered problem-solving, an individual must encounter novel situations and develop flexible, effective, and elegant solutions. This, in turn, deepens their understanding of mathematical content (Engelbrecht & Borba, 2024).

Posamentier and Krulik (2008) identify a wide range of problem-solving strategies that have been widely cited in mathematics education. To enhance their pedagogical relevance and reduce redundancy, these strategies can be meaningfully grouped into thematic categories. This structure highlights the cognitive demands and educational functions of each strategy.

2.1.1. Representation and Visualization

These strategies focus on structuring and externalizing information to make problems more accessible. They are particularly effective for visual and analytical learners and support the development of conceptual understanding.

Making a Drawing (Visual Representation)
Organizing Data
Accounting for All Possibilities

2.1.2. Heuristic and Exploratory Thinking

These approaches emphasize flexible, trial-and-error methods that promote active engagement with the problem-solving process. They align with constructivist learning theories, encouraging learners to explore and test multiple pathways.

Working Backwards
Solving a Simpler Analogous Problem
Intelligent Guessing and Testing (Including Approximation)

2.1.3. Analytical and Logical Reasoning

Strategies in this category require systematic thinking, deduction, and the evaluation of various scenarios. They foster higher-order cognitive skills, such as those outlined in Bloom's taxonomy, and support deep mathematical reasoning.

Logical Reasoning
Considering Extreme Cases
Accounting for All Possibilities (also relevant here due to its analytical nature)

2.1.4. Perspective-Taking and Strategic Reframing

These strategies involve approaching a problem from alternative viewpoints or reorganizing it strategically. They enhance cognitive flexibility and often reveal more efficient or insightful paths to a solution.

Adopting a Different Point of View
Working Backwards (also fits this category due to its retrospective framing)

2.1.5. Pattern Recognition and Generalization

This category emphasizes the ability to identify recurring structures and extrapolate general principles. It is particularly relevant in fostering algebraic thinking and abstraction.

Finding a Pattern

2.2. ChatGPT and Gemini in Instructional Processes

Recent advancements in natural language processing and machine learning algorithms have led to the rapid adoption of Large Language Model (LLM)-based chatbots capable of analyzing and interpreting human language with increasing accuracy. These systems generate human-like responses not through genuine human reasoning, but through statistical pattern recognition in large-scale datasets (Mikolov et al., 2011). Two prominent representatives of this technology are ChatGPT, developed by OpenAI, and Gemini, developed by Google AI (Mikolov et al., 2011). While both models are integrated into a wide range of instructional contexts, ChatGPT is frequently associated with fluent extended explanation and scaffolded feedback, whereas Gemini places stronger emphasis on multimodal integration, flexible language understanding, and real-time information retrieval (Bessas et al., 2025).

Enriched learning environments that include visual, interactive, and technology-supported elements have been shown to improve student attitudes toward problem-solving by fostering curiosity, enhancing engagement, and supporting active learning processes (Desoete & De Craene, 2019). Dynamic applications such as GeoGebra, real-world contexts, collaborative tasks, and immediate feedback mechanisms contribute to making mathematics instruction more meaningful and motivating. In parallel, the use of AI in education has grown considerably (Arslan, 2017; Lampou, 2023). This expansion continues largely because AI-supported feedback can guide students and support learning in more adaptive ways (Lampou, 2023). Research indicates that teacher candidates have already begun integrating ChatGPT into their instructional practices naturally as part of their questioning behavior. Similar studies confirm that AI-based tools, including ChatGPT, have become part of mainstream educational processes (Lampou, 2023).

ChatGPT contributes to instructional processes through multifaceted support to students, educators, and researchers. It enhances learning experiences, supports personalized learning, assists in writing, clarifies complex concepts, and provides individualized materials and continuous 24/7 access (Shuhaiber et al., 2025; Silva et al., 2024). For educators, ChatGPT facilitates the creation of instructional materials, automates repetitive tasks, and supports professional development activities through summarization, translation, and content generation (Holmes & Tuomi, 2022). For researchers, ChatGPT supports academic workflows by providing assistance in literature reviews, content drafting, and research organization (Baig & Yadegaridehkordi, 2024; Bhullar et al., 2024) Its adaptable structure also contributes to more inclusive learning contexts (Silva et al., 2024), and it has been found to increase interest and motivation especially in language learning and programming education (Heung et al., 2025; Shuhaiber et al., 2025).

Similarly, studies examining Gemini demonstrate that this model can improve vocabulary, grammar, pronunciation, coherence, and writing performance, especially for lower-proficiency learners (Alnasib & Alharbi, 2024). Students have expressed higher satisfaction and reduced anxiety when using Gemini, reporting positive emotional outcomes and increased academic self-efficacy (Bessas et al., 2025). In some studies, the use of Gemini in Arabic language education is supported through collaboration between developers and educators to address accuracy and privacy issues. Other research suggests that Gemini supports personalized and context-aware data analysis by integrating multiple information sources and adapting responses to user input and task requirements (Bessas et al., 2025). Additionally, Gemini has been applied in advanced scientific domains such as quantum computing education, highlighting both its potential and the challenges associated with its instructional integration (Bessas et al., 2025). However, some students and instructors report reliability limitations, methodological constraints, and context-specific effects, indicating the need for more comprehensive research and clearer pedagogical guidelines (Alnasib & Alharbi, 2024).

Despite the substantial advantages of AI-based tools, concerns remain regarding academic integrity, ethical use, data privacy, and potential over-reliance on AI (Silva et al., 2024). One key criticism is that ChatGPT sometimes generates inaccurate or biased content, which may prevent users from deeply understanding underlying reasoning processes (Shuhaiber et al., 2025; Silva et al., 2024). Gemini users have similarly reported issues including repetitive phrasing, restricted vocabulary range, excessively long responses, and uncertainty regarding accuracy and clarity (Alnasib & Alharbi, 2024). This is particularly critical in mathematics education contexts where transparency, strategic reasoning, and clarity of solution steps hold central importance (Zawacki-Richter et al., 2019). Therefore, while ChatGPT and Gemini offer promising contributions to instructional design, their effective use in mathematics requires structured integration, careful teacher monitoring, and clearly defined criteria for evaluating the quality and educational utility of AI-generated responses.

3. Method

3.1. Research Design

This study employed a qualitative research design to investigate the performance of AI-based chatbots in solving algebra problems. Specifically, a descriptive case study approach—an established method within qualitative research—was selected to enable an in-depth examination of the responses generated by two AI tools, ChatGPT and Gemini. A case study is particularly well-suited for exploring complex phenomena within real-life contexts and bounded systems, as it allows for the detailed analysis of processes, behaviors, or events.

In this research, the bounded system consists of a defined set of algebra problems and the corresponding responses provided by the selected AI tools. This framework allows for a focused and contextual analysis of each tool's problem-solving strategies, accuracy, and explanatory depth. A descriptive case study was preferred over other qualitative approaches because it offers a structured yet flexible means of capturing nuanced variations in AI-generated solutions while maintaining fidelity to the specific educational context in which the tools are evaluated. The data were systematically collected and presented through descriptive analysis, highlighting patterns and strategies observed during the problem-solving processes of the AI models.

This study aimed to examine AI-based tools' responses to algebra problems in terms of accuracy, diversity, solution strategies, and clarity. The problems used in the study were adapted from those presented by Posamentier and Krulik (2008).

3.2. Chatbots Used in the Study

- ➢ **Models and Versions.** We used ChatGPT-4 (via the ChatGPT web application) and Gemini (via the standard web-based Gemini application, free version) accessed in September 2025. These models were selected because they are (i) widely used by students and teachers, (ii) easily accessible without specialized infrastructure, and (iii) among the most frequently benchmarked large language models (LLMs) in education research. Both tools advertise Turkish support, enabling an ecologically valid evaluation in the study's target language.
- ➢ **Limitation.** Because model quality can vary across languages and releases, we report versions and access dates to aid reproducibility and we interpret cross-model differences cautiously.

The AI-based chatbots examined in this study were ChatGPT (OpenAI) and Gemini (Google DeepMind). These two tools were selected due to their widespread availability, frequent use by educators and students, and strong support for the Turkish language, making them suitable for evaluating algebra problem-solving performance in an authentic instructional context.

Both models are built on advanced NLP architectures capable of producing structured, step-by-step responses, allowing for the examination of solution correctness as well as clarity and logical coherence. Additionally, ChatGPT and Gemini are among the most frequently compared LLMs in current educational research. Previous studies (e.g., Baig & Yadegaridehkordi, 2024) highlight their pedagogical affordances and limitations, supporting the relevance of selecting these tools for comparative analysis in mathematics education.

In this study, the latest version of ChatGPT (ChatGPT-4) was used due to its improved reasoning and problem-solving abilities. Gemini was included as a comparative model representing an alternative AI infrastructure. The performances of both tools were examined through qualitative analysis.

- ➢ **Language and Setting.** All problem statements, prompts, and model interactions were conducted in Turkish. We selected Turkish to (i) match the target learner population (8th-grade students in Türkiye), (ii) avoid translation artifacts that could distort mathematical meaning, and (iii) reflect real classroom use, where both teachers and students interact in Turkish.

3.3. Data Collection Instrument

For this study, 10 open-ended algebra questions originally written in Turkish were used (Appendix 1). These questions were selected based on expert opinions and a review of the national mathematics curriculum, and were determined to be appropriate for the cognitive level and learning outcomes of 8th-grade students. To avoid potential misinterpretation in chatbot processing, no visual elements, figures, or symbolic representations were included. English translations of the original Turkish questions are provided in Appendix 2 for reference.

The ten algebra problems were selected according to three explicit criteria: content coverage, difficulty level, and cognitive demand. In terms of content, the items were aligned with algebra-related learning outcomes in the Turkish 8th-grade mathematics curriculum. With respect to difficulty, the set included both routine procedural problems and non-routine tasks requiring multi-step reasoning. Cognitive demand was addressed by including items corresponding primarily to the application, analysis, and evaluation levels of Bloom's taxonomy. Each problem was analytically associated with one dominant Posamentier and Krulik (2008) problem-solving strategy for classification purposes, while allowing alternative strategies to emerge in solution processes. The adequacy of the item set was confirmed by five mathematics education experts.

The expert group consisted of five university faculty members specializing in mathematics education. All experts held doctoral degrees in mathematics education or closely related fields and had a minimum of eight years of teaching and research experience. Their academic backgrounds included curriculum studies, mathematical problem solving, and assessment in middle-school mathematics. None of the experts were involved in the data analysis phase beyond validation procedures.

Each question was developed within Posamentier and Krulik's (2008) problem-solving strategy framework, and each item reflected one of the ten identified strategies. A categorisation matrix was created to map each problem to its corresponding strategy (e.g., working backwards, finding a pattern, considering extreme cases). Five mathematics education experts reviewed both the questions and the matrix to ensure content validity. The AI-generated responses were evaluated using an analytic rubric developed by the researchers, which was also reviewed and validated by the same expert group. The rubric assessed correctness, completeness of the solution process, and coherence of reasoning. While each item was associated with a single dominant problem-solving strategy for the purpose of analytical classification, this association did not constitute a one-to-one constraint on solution processes. During the analysis phase, AI-generated responses were examined without restricting them to the predefined strategy, allowing the identification of multiple strategies within a single response where applicable. To improve methodological transparency, one illustrative sample item from the problem set is presented below. This example is included to demonstrate the structure of the algebra problems used in the study and to clarify how AI-generated responses were evaluated using the analytic rubric, with particular attention to problem comprehension, strategy selection, execution, and verification.

Sample Item:

> "There are 8 blue socks, 6 green socks and 12 black socks in a drawer. Without looking at the socks in the drawer, what is the minimum number of socks that must be taken from the drawer to be certain that 2 socks of the same colour will be taken?"

Additionally, the clarity and usefulness of the explanations produced by ChatGPT and Gemini were evaluated by ten 8th-grade students selected through purposive sampling. The student group consisted of six female and four male students enrolled in a public middle school, with mixed levels of prior achievement in algebra as identified by their mathematics teacher. All participating students had prior experience with basic algebra topics covered in the national curriculum.

Expert review was employed to ensure alignment between the problem set, curricular objectives, and the analytical framework adopted in the study. The use of an analytic rubric enabled the systematic evaluation of both solution accuracy and reasoning processes, consistent with the focus of the study on problem-solving performance.

3.4. Data Collection Process

The data collection process was carried out in the following stages:

1. Preparation of Questions:
   ➢ Ten open-ended algebra problems were prepared (Appendix 1).
   ➢ To prevent potential misinterpretations by the chatbots, no figures, visuals, or scientific symbols were included.
2. Obtaining Responses from Chatbots:

The prepared problems were individually posed to ChatGPT and Gemini using a single standardized prompt to ensure comparability across responses. The exact prompt was: "Solve the following problem and explain your reasoning step by step." No additional scaffolding, follow-up questions, or prompt-tuning strategies were employed during data collection.

Each problem was submitted only once to avoid bias from repeated interactions, but the chatbots were instructed to generate multiple drafts in order to explore the range of possible solution strategies. This approach allowed the researchers to evaluate the diversity and consistency of the responses. All outputs were copied and compiled into structured answer sheets for further analysis.

3. Evaluation of Responses:

The responses generated by ChatGPT and Gemini were evaluated using a problem-specific analytical rubric developed by the researchers. This rubric was structured to assess the AI-generated solutions in relation to the problem-solving strategies proposed by Posamentier and Krulik (2008). For each question, the rubric included the following components:

Identification of the problem-solving strategy: *"Which of the problem-solving strategies outlined by* Posamentier and Krulik (2008) *does this problem correspond to?"*

Evaluation of solution steps: *"Are the solution steps presented above appropriate and aligned with the identified strategy?"*

Evaluation of the scoring accuracy: *"Is the scoring based on the solution steps accurate and justifiable?"*

Open-ended feedback: *"Please share any additional comments regarding the solution process and the assigned score, if applicable."*

Limitation: we did not conduct prompt tuning; future work should compare multiple pedagogically enriched prompts.

Additionally, the quality, usefulness, and clarity of the solutions generated by ChatGPT and Gemini were evaluated by ten 8th-grade students selected through purposive sampling, based on their willingness to participate and their basic proficiency in algebra. Prior to data collection, written informed consent was obtained from parents/guardians, and verbal assent was obtained from the students. The ethical protocol of this study was reviewed and approved by a Social and Human Sciences Ethics Committee. The name of the committee, meeting date, and decision number have been anonymized at this stage in accordance with the double-blind review process.

3.5. Expert Validation

Five experts who were university faculty members in mathematics education reviewed the rubric and item–strategy alignment prior to data collection. Each expert had at least eight years of experience, and they confirmed the content validity, verified the alignment between items and problem-solving strategies, and provided feedback to improve clarity in scoring criteria. The rubric was refined based on their suggestions before being used to evaluate chatbot responses.

To examine internal consistency of the AI tools, all responses were re-submitted to ChatGPT and Gemini using a different user account, and the newly generated solutions were evaluated using the same rubric. This allowed comparison of multiple outputs under the same conditions. In addition, inter-rater reliability for human scoring was assessed by having two independent raters evaluate the original responses. Cohen's kappa coefficient was calculated and yielded $\varkappa = 0.81$, indicating a high level of agreement and confirming the reliability of the rubric-based scoring.

The chatbots were informed that the answers belonged to a student and were asked to evaluate them using the same assessment rubric.

4. Scoring Process:
   - ➢ Each chatbot could earn a maximum score of 10 per problem.
   - ➢ Since the study involved 10 problems, the highest possible total score for each chatbot was 100.
   - ➢ The same scoring system was applied to the evaluations conducted by the researchers.

In addition, each problem was broken down into four pedagogically meaningful steps, as determined through expert consultation. These steps reflect the key components of the problem-solving process specific to each problem—such as interpreting the problem, selecting an appropriate strategy, executing the solution, and validating the result. Scoring was conducted step-by-step, with partial points assigned to each stage. This step-based evaluation allowed for a more detailed and diagnostic analysis of the AI-generated solutions. These components are reflected in Table 1 as Step 1 to Step 4.

Since all tasks, prompts, and student judgments were carried out in Turkish, the study reflects the linguistic and curricular context of Turkish secondary education.

Table 1. *The accuracy of the solutions Provided by ChatGPT and Gemini to the problems*

Question No		Step 1	Step 2	Step 3	Step 4
1	ChatGPT	4 point	0 point	0 point	0 point
2		4 point	6 point	-	-
3		3 point	5 point	2 point	-
4		4 point	2 point	-	-
5		3 point	6 point	1 point	-
6		5 point	3 point	2 point	-
7		5 point	5 point	-	-
8		5 point	5 point	-	-
9		5 point	2 point	2 point	1 point
10		1,5 point	1,5 point	0 point	-
1	Gemini	0 point	0 point	0 point	0 point
2		4 point	6 point	-	-
3		3 point	5 point	2 point	-
4		0 point	0 point	-	-
5		3 point	6 point	1 point	-
6		5 point	3 point	2 point	-
7		5 point	5 point	-	-
8		5 point	5 point	-	-
9		5 point	2 point	2 point	1 point
10		0,75 point	0,75 point	0 point	-

Note. Step 1–4 refer to the four key stages in the solution process of each problem, determined based on expert input.

3.6. Data Analysis

Independent coding and consensus-based resolution were used to enhance analytical transparency and consistency in the identification of problem-solving strategies.

Analytical Framework. To evaluate chatbot performance, we adopted a four-step rubric adapted from Posamentier and Krulik's (2008) problem-solving framework:

1. **Comprehension:** Identifying and correctly interpreting the problem statement.
2. **Strategy:** Selecting an appropriate problem-solving strategy or approach.
3. **Execution:** Accurately applying the chosen strategy to perform calculations or transformations.
4. **Verification:** Checking and validating the correctness of the solution.

This framework allowed for a step-based evaluation of chatbot responses, enabling a more fine-grained analysis than binary correct/incorrect scoring. The rubric was content-validated by five mathematics education experts, with an inter-rater reliability coefficient of $\kappa = 0.81$, ensuring consistency in scoring. Strategy coding was performed independently by two researchers based on predefined categories derived from Posamentier and Krulik's framework, and inter-rater agreement for this coding process was assessed using Cohen's kappa, as reported above.

To answer the research question and sub-questions, qualitative analysis methods were employed. The data obtained from AI responses and expert evaluations were analyzed using content analysis. This process was conducted manually by the researchers, guided by predefined codes derived from the problem-solving strategies of Posamentier and Krulik (2008). To ensure systematic coding and reduce potential bias, the analysis was independently conducted by two researchers and subsequently cross-checked. Although no qualitative data analysis software was used, inter-rater agreement was assessed to enhance the reliability of the findings. Specifically, Cohen's Kappa coefficient was calculated to determine the consistency between coders, yielding a value of 0.81, which indicates a strong level of agreement.

To determine the performance of the chatbots, their responses to each problem were examined by both the researchers and ChatGPT/Gemini.

Three key aspects were considered during the analysis to provide a comprehensive understanding of the chatbots' performance:

3.6.1. Performance Evaluation

The primary metric of success was based on the scores assigned to each response by expert raters, using the rubric developed by the researchers. This evaluation aimed to assess the correctness, logical coherence, and completeness of each solution. By aggregating these scores across all problems, the overall performance of ChatGPT and Gemini could be compared both quantitatively and qualitatively.

3.6.2. Solution Diversity

To explore the capacity of the chatbots for flexible and creative problem-solving, their ability to generate multiple, distinct solution strategies was examined. This dimension is particularly important given that one of the research aims was to evaluate the extent to which AI tools align with the heuristic and strategic aspects of mathematical thinking. Instances where the same problem elicited alternative valid approaches were identified, coded, and analyzed to gain insight into the internal variability and problem-solving depth of each chatbot.

3.6.3. Clarity Analysis

The clarity and comprehensibility of the solutions were evaluated by ten 8th-grade students, who represent the target audience for the problem set. These students were selected through purposive sampling to ensure variation in mathematical achievement levels. Informed consent was secured from both the students and their guardians. During this phase, students were asked to reflect on whether they found the chatbot-generated responses meaningful and understandable. Their feedback provided insight into how effectively the explanations communicated mathematical reasoning at an age-appropriate level. For the purposes of analysis, student responses were anonymized and labeled from S1 to S10. This component of the study complemented expert evaluations by incorporating the perspectives of actual learners, thus strengthening the assessment of clarity from a user-centered viewpoint.

As a result of these analyses, data were obtained regarding the success of chatbots in mathematical problem-solving, their ability to propose different strategies, and the clarity of their explanations.

We provide a study flowchart (Figure 1) summarizing design, data collection, and analyses.

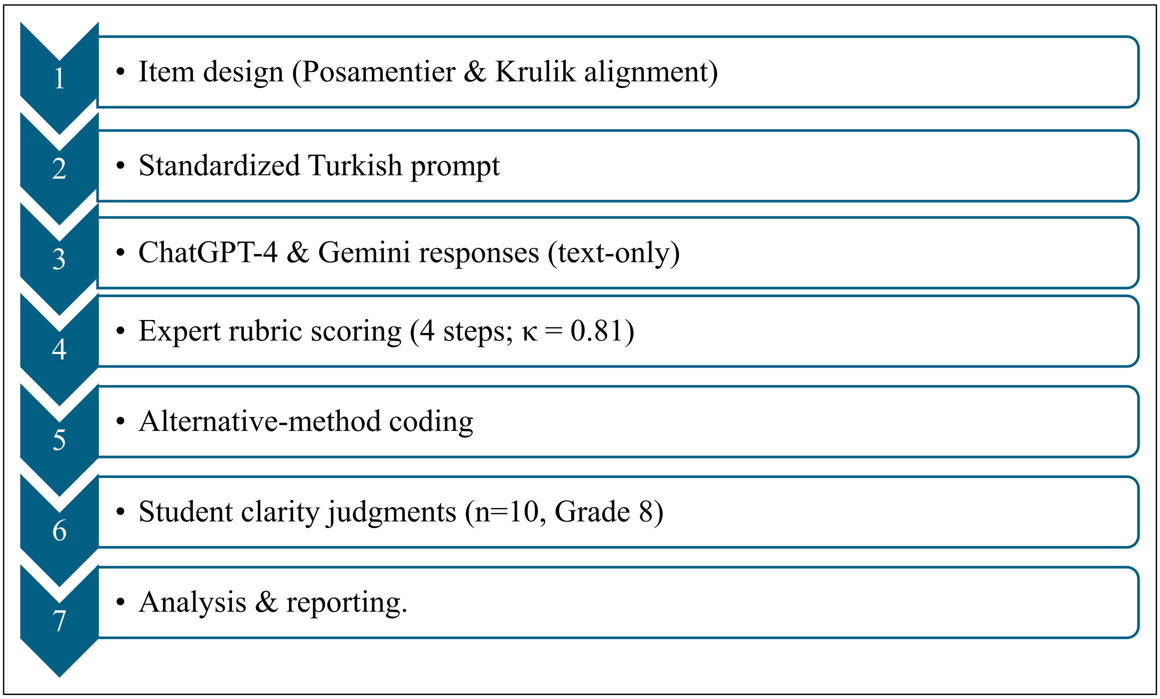

Figure 1. *Study flow*

4. Findings

4.1. Findings on ChatGPT and Gemini's Problem-Solving Performance

To address the first sub-question of the study, ChatGPT and Gemini were presented with math problems and their responses were scored using a predefined rubric. As a result of the scoring, ChatGPT scored 83 points and Gemini scored 71.5 points. This indicates that ChatGPT outperformed Gemini in mathematical problem-solving, as evidenced by higher overall scores assigned through the rubric-based evaluation. Success in this context refers to the correctness, completeness, and logical coherence of the solutions, as well as the chatbot's ability to generate accurate and appropriately structured responses aligned with the intended problem-solving strategies. In addition to these findings, beyond the rubric prepared by the researchers, the chatbots were also asked to evaluate their own responses. As a result of the scoring, ChatGPT assigned a score of 78 to its own responses, while Gemini assigned a score of 73 to its own responses. In line with this sub-problem of the study, the accuracy of the answers provided by ChatGPT and Gemini to the problems was also examined. This evaluation was based on problem-solving steps previously determined by experts.

Table 1 shows the rubric with four steps:

Step 1 – Comprehension (understanding the problem),
Step 2 – Strategy (planning a solution method),
Step 3 – Execution (carrying out the plan),
Step 4 – Verification (checking and interpreting the solution).

Table 1 shows that ChatGPT and Gemini differed in the accuracy of their solutions. Each problem was scored out of 10 points using a four-step expert-developed rubric (understanding the problem, selecting a method, executing the strategy, and verifying the result), which allowed for a detailed evaluation of each stage of problem-solving.

ChatGPT generally performed better across most problems, while Gemini either failed to produce a solution or generated lower-quality responses. For instance, in Problem 1, ChatGPT partially identified the correct structure and earned points only from the initial step, whereas Gemini could not provide a meaningful response. A similar pattern reoccurred in Problem 4.

Both chatbots struggled particularly with Problems 1, 4, and 10—items requiring logical reasoning, abstraction, and strategic planning. These problems demanded non-routine thinking and complex pattern recognition, which are well documented as cognitively demanding in mathematics education (e.g., Van de Walle et al., 2013). Their performance suggests that current LLMs face clear challenges with open-ended, multi-step reasoning tasks.

In contrast, both ChatGPT and Gemini successfully solved more procedural algebra problems (e.g., linear equations, evaluating expressions), indicating that LLMs perform more consistently when working within structured and routine problem contexts.

4.2. Findings on ChatGPT and Gemini's Performance in Providing Alternative Solution Methods

To address the second sub-question of the study, an analysis was conducted to determine whether ChatGPT and Gemini proposed different solution methods for the given problems. Table 2 presents information on the number of different solution approaches each chatbot provided for the problems.

Table 2 shows that ChatGPT and Gemini generally performed similarly in proposing alternative solution methods, with clearer differences appearing in Problems 1 and 9. In Problem 1, ChatGPT proposed a creative numerical pairing strategy, while Gemini did not offer an alternative approach. In Problem 9, however, the pattern was reversed: Gemini produced six alternative approaches compared with ChatGPT's four, drawing on computational and number-theoretic methods. These findings indicate that both models generate diverse strategies for standard algebraic problems, while ChatGPT appears more flexible in non-routine contexts where no procedural template is available.

It was also observed that Gemini occasionally preferred more technical, programming-oriented approaches, especially in Questions 4 and 9, where it relied on iterative structures or pseudo-code-like reasoning. This computational perspective may be associated with its design and development background. In contrast, in Problems 6 and 10, both models used similar strategies such as systematic case analysis, combinatorics, and tree diagrams, which reflect conventional mathematical reasoning commonly emphasized within school curricula.

Table 2. *Alternative Solution Methods Provided by ChatGPT and Gemini*

Question No	Alternative solution methods	
	ChatGPT (38)	Gemini (34)
1	Systematic matching and observation, mathematical generalization, matching graph. (3)	There is no solution. (0)
2	Solving using a formula, finding the numbers and calculating, reversing and summing method, approach with symmetric polynomials. (4)	Solving using formulas, solving the system of equations, using Vieta's formulas. (3)
3	Finding a cycle in the units digit, using modular arithmetic, simple multiplication Cycle. (3)	Finding cycles in the units digit, using modular arithmetic, computer programming. (3)
4	Systematic trial and optimization, arranging term groups, combinatorial approach, functional approach. (4)	Trying all possibilities (systematic search), using inequalities, solving with programming languages. (3)
5	Solving the equation directly, direct substitution, proportion and inversion method, general method. (4)	Solving the equation directly, solving by setting up a proportion, variable substitution, graphical solution. (4)
6	Method based on fundamental definitions, summing the reciprocals of all divisors, approach based on factors and reciprocal sum, systematic approach. (4)	Method based on fundamental definitions, prime factorization method, functional approach, computer programming. (4)
7	Considering the worst-case scenario (Pigeonhole Principle), general logical approach, using formulas, experimental approach. (4)	Considering the worst-case scenario (Pigeonhole Principle), probabilistic approach, visualization, logical reasoning. (4)
8	Basic equation solving, proportional method, fractional method, direct number definition. (4)	Probabilistic approach, visualization, logical reasoning, solving through visualization. (3)
9	Common denominator method, direct coefficient solving, modular arithmetic usage, graphical solution. (4)	Testing by assigning values to x, table method, GCD-LCM relationship, computer programming, modular arithmetic usage, graphical solution. (6)
10	Using tree diagrams, combination method, separating consecutive winning and total winning conditions, showing probability flow. (4)	Using tree diagrams, combination method, Permutation method, sample space method. (4)

4.3. Findings on the Problem-Solving Strategies Used by ChatGPT and Gemini

In line with the third sub-research question, the problem-solving strategies employed by ChatGPT and Gemini were examined. When identifying these strategies, the framework proposed by Posamentier and Krulik (2008) was taken into account. This framework outlines a set of heuristic problem-solving strategies designed to enhance mathematical thinking. These include techniques such as working backwards, finding a pattern, considering extreme cases, logical reasoning, making a drawing, and solving a simpler analogous problem, among others. The goal of these strategies is to encourage students to approach problems from multiple perspectives and to develop flexible, analytical thinking skills. Table 3 presents the problem-solving strategies observed in the responses of the conversational AI models alongside the strategies identified by experts, allowing for a comparative evaluation. Table 3 presents the problem-solving strategies used by the conversational AI models alongside the strategies recommended by experts.

Table 3. *Findings Based on the Thematic Categorization of Problem-Solving Approaches Observed in ChatGPT and Gemini*

Question No	Thematic categorization of problem-solving approaches		
	ChatGPT	Gemini	Strategies recommended by experts
1	Representation and Visualization, Analytical and Logical Reasoning	Representation and Visualization, Analytical and Logical Reasoning	Representation and Visualization, Analytical and Logical Reasoning, Heuristic and Exploratory Thinking
2	Heuristic and Exploratory Thinking, Perspective-Taking and Strategic Reframing, Representation and Visualization, Analytical and Logical Reasoning	Heuristic and Exploratory Thinking, Perspective-Taking and Strategic Reframing, Representation and Visualization, Analytical and Logical Reasoning	Heuristic and Exploratory Thinking, Perspective-Taking and Strategic Reframing, Representation and Visualization, Analytical and Logical Reasoning
3	Pattern Recognition and Generalization, Analytical and Logical Reasoning	Pattern Recognition and Generalization, Analytical and Logical Reasoning	Pattern Recognition and Generalization, Analytical and Logical Reasoning, Heuristic and Exploratory Thinking
4	Representation and Visualization, Representation and Visualization, Analytical and Logical Reasoning	Representation and Visualization, Representation and Visualization, Analytical and Logical Reasoning	Representation and Visualization, Representation and Visualization, Analytical and Logical Reasoning
5	Representation and Visualization, Perspective-Taking and Strategic Reframing, Heuristic and Exploratory Thinking, Perspective-Taking and Strategic Reframing, Analytical and Logical Reasoning	Representation and Visualization, Perspective-Taking and Strategic Reframing, Heuristic and Exploratory Thinking, Perspective-Taking and Strategic Reframing, Analytical and Logical Reasoning	Representation and Visualization, Perspective-Taking and Strategic Reframing, Analytical and Logical Reasoning
6	Heuristic and Exploratory Thinking, Heuristic and Exploratory Thinking, Analytical and Logical Reasoning	Heuristic and Exploratory Thinking, Representation and Visualization, Analytical and Logical Reasoning	Heuristic and Exploratory Thinking, Pattern Recognition and Generalization, Analytical and Logical Reasoning
7	Heuristic and Exploratory Thinking, Analytical and Logical Reasoning	Heuristic and Exploratory Thinking, Analytical and Logical Reasoning	Heuristic and Exploratory Thinking, Analytical and Logical Reasoning, Representation and Visualization
8	Heuristic and Exploratory Thinking, Perspective-Taking and Strategic Reframing, Analytical and Logical Reasoning	Heuristic and Exploratory Thinking, Analytical and Logical Reasoning	Heuristic and Exploratory Thinking, Perspective-Taking and Strategic Reframing, Analytical and Logical Reasoning, Representation and Visualization
9	Heuristic and Exploratory Thinking, Perspective-Taking and Strategic Reframing, Analytical and Logical Reasoning	Heuristic and Exploratory Thinking, Perspective-Taking and Strategic Reframing, Analytical and Logical Reasoning	Analytical and Logical Reasoning, Representation and Visualization
10	Representation and Visualization, Analytical and Logical Reasoning	Representation and Visualization, Analytical and Logical Reasoning	Representation and Visualization, Analytical and Logical Reasoning

Table 3 shows that although ChatGPT and Gemini frequently apply similar thematic strategies—such as Analytical and Logical Reasoning or Heuristic and Exploratory Thinking—the depth and variety of strategies they employ are more limited than those identified by experts. In several items, experts associated the same problem with multiple thematic categories (e.g., three or more in Questions 1 and 7), whereas both chatbots tended to rely on a narrower set. This indicates that while AI tools demonstrate partial alignment with mathematical problem-solving processes, they lack the flexibility and cognitive range typically observed in expert reasoning.

These results also have instructional implications. Although LLMs can reproduce standard reasoning pathways, they often struggle to shift between different reasoning modes when tasks require abstraction or representational flexibility. In contrast, when the problem structure was more procedural and well-defined (e.g., Questions 5 and 10), both chatbots aligned more closely with expert categorizations. This suggests that current LLMs perform more effectively in procedural contexts than in complex, open-ended tasks. Therefore, instructional designs that incorporate AI support should include scaffolding and teacher oversight to ensure that students engage with diverse cognitive strategies rather than relying solely on surface-level procedural outputs.

4.4. Findings on Student Perspectives Regarding ChatGPT and Gemini's Responses

In line with the fourth sub-research question, student perspectives on the quality, usefulness, and clarity of the responses provided by ChatGPT and Gemini were examined. The findings related to the "quality and usefulness of responses" were presented first.

To gather student perspectives, they were asked the following questions:

Are you confident in the accuracy of the given response? Why?
Did the response help you solve a similar problem?
Was it beneficial for you to see multiple solution approaches in the response? Why?
Did you learn something new from the response?

Based on the student responses, the themes and codes that emerged are presented in Table 4.

Table 4. *Student Perspectives on the Quality and Usefulness of ChatGPT and Gemini's Responses: Thematic Analysis by Sentiment*

Theme	Sub-theme (code)	Frequency (f)	Sentiment
Perceived Usefulness	Helped solve similar problems	4	Positive
	Encouraged different perspectives	3	Positive
	Enabled faster problem-solving	2	Positive
	Improved understanding via alternatives	2	Positive
	Information more memorable	1	Positive
	Increased awareness of AI use	1	Positive
	Learned something new	2	Positive
	Provided sufficient explanation	5	Positive
	Included different solution methods	2	Positive
	Clarified problem structure	1	Positive
	Responses clear and understandable	1	Positive
	Explained in detail	2	Positive
	Well-structured and comprehensive	1	Positive
	No lingering doubts remained	1	Positive
	Alternative responses were helpful	1	Positive
	Responses answered the question	2	Positive
	Helped with retention	1	Positive
Perceived Limitations	Solution too result-oriented	1	Negative
	Solution lacked sufficient explanation	4	Negative
	Response unclear	3	Negative
	Solution too complex	4	Negative
	Too many solution paths caused confusion	1	Negative
	Not explained in a simple manner	1	Negative
	Too much numerical computation, little explanation	1	Negative
	Confusing on first read	1	Negative

The participants consisted of ten 8th-grade students who were selected through purposive sampling based on their willingness to participate and their teachers' judgments of their basic proficiency in algebra. This ensured that the students were capable of interpreting and evaluating the AI-generated responses. Although the sample size was small, data saturation was considered to have been reached, consistent with qualitative research practices where rich, in-depth perspectives are prioritized over broad numerical generalizability.

Table 4 shows that most students expressed positive views on the usefulness and quality of ChatGPT and Gemini's responses, particularly emphasizing the accuracy of solution steps and clarity of language. For example, S2 stated, *"Yes, I am confident in the accuracy of the response because I believe it follows the correct steps."* Students also noted that the AI responses helped them solve problems more quickly and offered alternative methods. As S6 indicated, *"Actually, yes. Because it can help me solve problems faster using different methods."* Similarly, S7 reported, *"Yes, it was helpful. Because even if I didn't understand one method, the other one helped me."*

Multiple solution pathways were also perceived to support better retention and deeper understanding. S8 noted, *"Yes, it was actually beneficial because some responses were more concise and easier to remember."* Likewise, S9 added, *"Yes, it was helpful. Seeing multiple approaches made the process more understandable and reliable. Also, I didn't have any lingering doubts about whether there was another possible solution."*

However, some students expressed negative perspectives related to clarity and cognitive load. S3 reported, *"No, it wasn't helpful. Everything was too complicated."* Similarly, S10 stated, *"It was confusing. Maybe one or two solutions would have been enough."*

During the continuation of the interview, students were asked additional questions to assess their views on the clarity of the responses:

Do you think the given response fully answers the question?
Did you have difficulty understanding any part of the response?
Was the response clear enough for you? Why?
Were complex concepts in the response explained sufficiently?

Based on the student responses, the emerging themes and codes are presented in Table 5.

In Table 5, most students reported positive views regarding the clarity of the responses provided by ChatGPT and Gemini. Students emphasized the value of multiple possibilities, detailed explanations, and sufficient elaboration in making the responses understandable. For example, S9 stated, *"Yes, it answers the question because it provides multiple possibilities and determines how many ways it can conclude."* Similarly, S5 remarked, *"The responses were clear because they were explained in detail."* S9 further added, *"Yes, all the methods were well-structured and understandable. Additionally, alternative solutions eliminated any uncertainties."* Likewise, S7 commented, "It was clear because everything was explained in detail."

Table 5. *Student Perspectives on the Clarity of ChatGPT and Gemini's Responses*

Question No	Theme	Code	Frequency (f)
1	Yes (f=8)	Presenting different possibilities	3
	No (f=2)	Solution being too result-oriented	1
		Lack of sufficient explanation	1
		Solution being unclear	1
2	Yes (f=4)	Lack of sufficient explanation	2
		Solution being complex	2
		Solution being unclear	1
		Solution not being explained simply	1
	No (f=6)	Solution being clear and understandable	1
3	Yes (f=7)	Including details in solutions	2
		Offering different solution methods	1
	No (f=3)	Solution being complex	1
		Lack of sufficient explanation	1
4	Yes (f=5)	Providing sufficient explanation	5
	No (f=5)	Solution being complex	1
		Lack of sufficient explanation	1

In contrast, some students noted difficulties related to lack of clarity or cognitive overload. S5 mentioned, *"At first, I found it confusing while reading, but after rereading, I understood it."* Likewise, S7 said, *"No, it was too complicated."* S10 also stated, *"No. There were too many numerical calculations and too little explanation, which made it difficult for me to understand."*

These findings suggest that although AI-generated responses may be correct, they are not always pedagogically suitable for every learner. Dense numerical content and limited scaffolding may overwhelm middle school students, especially those who rely on step-by-step guidance. This indicates the importance of integrating instructional design principles such as simplification, progressive explanation, and clearer conceptual framing when using AI-generated solutions in mathematics education.

5. Discussion and Conclusion

5.1. Discussion of Findings Related to RQ1: Performance, Strategies, and Clarity

Results indicate that ChatGPT outperformed Gemini on rubric-based scoring (83 vs. 71.5) yet both models struggled on non-routine items requiring multi-step reasoning. Furthermore, correctness alone did not guarantee clarity for 8th-grade learners; several student comments pointed to either insufficient scaffolding or overly dense numerical exposition. These findings align with the step-based analysis presented in the Results section.

Importantly, we refrain from attributing stylistic tendencies (e.g., iterative or pseudo-code-like reasoning) to any particular developer culture; instead, we report them descriptively as observed response patterns under our Turkish, text-only setup.

The research findings indicate that ChatGPT outperforms Gemini in mathematical problem-solving tasks. ChatGPT scored 83 points, offering more accurate and diverse solutions, whereas Gemini scored 71.5 points and produced a higher number of incorrect or partially correct responses. This discrepancy may reflect underlying differences in their natural language processing architectures and training objectives. ChatGPT, particularly in its latest version (GPT-4), appears to possess a more robust internal reasoning structure and greater linguistic flexibility, which may contribute to its superior performance in educational contexts—a pattern also observed in prior comparative studies (Baig & Yadegaridehkordi, 2024).

However, it is important to note that both models exhibited limitations when dealing with cognitively demanding problems. Their inadequate performance on questions 4 and 10—problems that required multi-step reasoning and abstraction—suggests that large language models (LLMs) may still struggle with non-routine mathematical tasks. These performance differences constitute empirical findings of the study and are further discussed in relation to their pedagogical implications in the following paragraphs. Constraints related to generalizability and scope are addressed separately in the Limitations section.

Moreover, while ChatGPT demonstrated a broader range of solution methods, Gemini tended to offer more technical and programming-oriented responses. For instance, in Questions 4 and 9, Gemini applied algorithmic logic and referred to computer science principles, such as pseudocode and iterative search—suggesting a distinct orientation shaped by its development background. Although these approaches may be suitable for advanced learners, they can be cognitively inaccessible for middle school students, potentially hindering rather than enhancing mathematical understanding (Desoete & De Craene, 2019).

These findings emphasize that the performance of AI chatbots is not merely a function of correctness but also of pedagogical appropriateness. A tool that generates correct answers may still fall short if the reasoning is opaque or misaligned with learners' developmental levels. As such, educational practitioners and developers should consider not only algorithmic sophistication but also instructional clarity and user adaptation when integrating AI into teaching and learning processes.

Consistent with studies in the literature (Zawacki-Richter et al., 2019), both models were observed to employ effective problem-solving strategies in most cases. However, in some instances, the models did not fully implement expert-recommended strategies, emphasizing the limitations of chatbots in problem-solving. The research findings align with previous studies suggesting that AI enhances problem-solving abilities (Zawacki-Richter et al., 2019). Nevertheless, the limited number of studies examining the specific performance of ChatGPT and Gemini in algebraic mathematical problems (Engelbrecht & Borba, 2024) increases the contribution of this research to the existing literature. In particular, ChatGPT's ability to provide diverse solutions highlights the potential of AI-based educational technologies to support flexible thinking and individualized learning (Adıgüzel et al., 2023).

However, this potential must be considered alongside certain limitations. While AI-generated responses can offer variety and speed, they do not always align with national curriculum goals, grade-level expectations, or established pedagogical principles. For instance, some solutions produced by AI may prioritize efficiency over clarity or introduce concepts that are beyond students' cognitive readiness. In this sense, the integration of AI into mathematics education should be guided by instructional frameworks that prioritize developmental appropriateness and curricular coherence. Without such alignment, AI tools risk offering technically correct yet educationally unsuitable responses—particularly for middle school learners still building foundational algebraic reasoning. These observations reflect pedagogical implications of the findings rather than methodological constraints; the latter are addressed in the Limitations section.

The research findings, consistent with several studies in the literature, indicate that AI-based chatbots can serve as supportive tools in education (Antony & Ramnath, 2023; Holmes & Tuomi, 2022). These studies generally emphasize the benefits of chatbots in terms of increased engagement, quick access to information, and support for learning autonomy. For instance, research suggests that educational chatbots support efficient conceptual understanding in STEM-related learning contexts through timely feedback and scaffolded explanations, while studies have reported increased academic self-efficacy among learners using chatbot-based systems (Chang et al., 2022).However, many of these studies focus on domains such as language learning or general information retrieval, with limited emphasis on mathematical reasoning and problem-solving, particularly in algebra. This study expands on the existing literature by focusing specifically on how ChatGPT and Gemini perform in solving algebraic problems and how their outputs are perceived by students in terms of clarity, usefulness, and strategic diversity.

Moreover, while previous research often highlights the positive aspects of AI integration, the present findings also reveal significant pedagogical limitations, such as the misalignment between AI-generated responses and students' cognitive readiness or curricular expectations. For instance, some students in this study found the responses too complex or insufficiently explained—issues not fully addressed in prior works. This discrepancy underscores the importance of contextual and discipline-specific evaluations when assessing the educational value of AI tools.

Chatbots such as ChatGPT and Gemini have the potential to enhance students' problem-solving skills and make mathematics lessons more effective (Desoete & De Craene, 2019). However, it is crucial to provide appropriate guidance to students in using these tools and to verify the accuracy of the responses they generate, in order to support meaningful and reliable learning processes (Fischer et al., 2013). In this context, the performance of ChatGPT and Gemini in mathematical problem-solving highlights the potential of AI in education and offers a new perspective for advancements in this field.

The coexistence of student feedback indicating that some responses were "too complex" while others were "not sufficiently explained" raises important questions about the source of confusion. This divergence suggests that the challenge lies not solely in the AI's content but in its alignment with learners' cognitive readiness. For some students, the explanations may have lacked sufficient scaffolding or conceptual elaboration, making the solutions feel abrupt or incomplete. For others, the abundance of solution steps, dense numerical content, or advanced terminology may have overwhelmed their processing capacity. Therefore, the confusion reported by students appears to stem from a combination of factors: the inherent complexity of certain algebra problems, the AI's tendency to present answers in a concise or overly technical manner, and the absence of pedagogical adjustments based on learner variability. This highlights the need for educational AI tools to adapt not only to content demands but also to students' developmental levels—ensuring that explanations are both clear and cognitively accessible. Teachers' guidance remains crucial in bridging this gap by contextualizing AI-generated content, simplifying complex ideas, and encouraging reflective comparison between different solution methods (Desoete & De Craene, 2019).

5.2. Discussion of Findings Related to RQ2: Consistency Across Accounts

With the advancement of AI technologies, the potential for transformation in education systems is high. AI-supported learning environments have been shown to provide numerous benefits, from offering personalized learning experiences to reducing teachers' workloads (Luckin & Holmes, 2016). The use of AI in educational settings can help identify students' individual needs and offer personalized learning paths (Zawacki-Richter et al., 2019). Additionally, it can create an inclusive learning environment for students with special needs (Holmes, 2019). AI can assist teachers with lesson planning and tracking student performance (Selwyn, 2019). AI-supported assessment systems can provide instant feedback to students, accelerating the learning process, and enable automated grading of exams, helping teachers use their time more efficiently (Ferrara & Qunbar, 2022). It is essential to remember that integrating AI into educational environments presents significant opportunities for both students and teachers. However, developing these technologies in alignment with pedagogical and ethical principles is crucial (Holmes, 2019). Teachers play a vital role in integrating AI tools into educational environments. They should act as facilitators, promoting critical thinking, raising ethical awareness, and managing personalized learning processes. Since AI tools can sometimes provide incomplete or incorrect information, teachers must help students cultivate a critical perspective to distinguish misinformation (Williamson et al., 2020). In such cases, encouraging students to compare AI-generated responses with other sources can foster habits of information verification and responsible use (Bain, 2004).

Beyond the role of individual educators, educational institutions must take a proactive stance by developing clear guidelines, policies, and ethical frameworks to govern the use of AI tools in classrooms. These frameworks should address not only data privacy, academic integrity, and bias prevention but also offer practical strategies for ensuring developmental appropriateness and pedagogical alignment. Institutions should also provide professional development programs to enhance AI literacy among educators and establish protocols for evaluating the quality and suitability of AI-generated educational content. Without such institutional structures, the responsibility of ethical AI use falls disproportionately on teachers, potentially leading to inconsistent practices and unregulated use. Therefore, a systemic approach that combines teacher training with institutional oversight is essential for embedding AI into education in a way that is equitable, transparent, and pedagogically sound (Arvin et al., 2023; Williamson et al., 2020).

The use of AI tools in education also raises ethical concerns. Teachers must ensure that students use AI not for academic dishonesty but as a means to support their learning process (Bhullar et al., 2024). Additionally, they should guide students in protecting their personal information and becoming responsible digital citizens (Williamson et al., 2020).

Of course, for teachers to fulfill these responsibilities effectively, they must be well-equipped and continuously improve their AI literacy (Arvin et al., 2023). To successfully integrate AI tools into instructional designs, teachers need to be provided with the necessary digital competencies (Kohnke et al., 2025).

The study results indicate that chatbot AI systems employ different solution strategies, sometimes fail to find solutions, or provide incorrect answers to mathematical problems. The errors AI makes while solving math problems can significantly impact students' learning experiences and perceptions of mathematics. When AI systems make mistakes, it may lead to confusion and frustration among students. This finding is consistent with prior research indicating that students generally perceive AI-based tools as reliable; however, encountering errors or inaccuracies may negatively affect their trust in such systems (Shuhaiber et al., 2025).

The primary purpose of artificial intelligence in educational settings is to facilitate learning by providing instant feedback and continuous instructional support (Lampou, 2023). However, when AI tools fail to provide correct solutions, the flow of instruction may be disrupted, leading to misconceptions of mathematical concepts. In such cases, teachers' guidance becomes even more crucial (Fischer et al., 2013).

The key takeaway is that AI errors present both risks and opportunities. If students develop critical thinking and verification habits, these mistakes can enhance the learning process. However, if AI errors go unnoticed, students may acquire incorrect information, leading to long-term conceptual misunderstandings. For instance, an AI model might incorrectly solve a system of linear equations—misapplying substitution or elimination—causing students to adopt flawed strategies. Similarly, AI-generated explanations might oversimplify fraction operations (e.g., suggesting that fractions should be added by directly adding numerators and denominators) or misinterpret variables by treating them as constants. Such errors, if internalized, could hinder students' ability to reason logically, solve problems accurately, and build foundational mathematical understanding. Thus, close teacher supervision and instructional scaffolding are essential when integrating AI tools into learning environments.

6. Recommendations

This study was limited to a specific mathematical domain (algebra) and a particular grade level (8th grade in primary education). As such, the generalizability of the findings is constrained. Future research should include a broader range of mathematical topics, such as geometry, statistics, and calculus, as well as varying educational levels. Expanding the participant pool to include students from diverse socio-economic and educational backgrounds would also enhance the external validity of the results.

Future studies should also incorporate a wider array of AI-based tools, not only ChatGPT and Gemini but emerging models specialized in education. Comparative analyses could help identify which platforms are best suited for specific mathematical problem types or learner profiles. Longitudinal studies could further reveal long-term effects of AI on mathematical learning and conceptual development.

Additionally, richer prompting strategies, including explicit scaffolding, should be explored to determine how prompts influence model performance and student comprehension. Incorporating multimodal items such as diagrams or graphs could provide a more authentic assessment of mathematical reasoning. Furthermore, reporting detailed model and version metadata will be critical for reproducibility and for understanding differences in performance across AI systems.

Teacher training programs should be strengthened to enhance AI literacy, including prompt engineering, error detection, ethics, and pedagogical adaptation. Curriculum developers should collaborate with AI specialists to design AI-supported instructional modules aligned with national standards, including scaffolding strategies and adaptive features.

School administrators and policymakers should establish institutional policies and ethical guidelines for AI usage, addressing academic integrity, data privacy, bias mitigation, and equitable access. Additionally, student-facing platforms using AI must be evaluated for instructional clarity, age appropriateness, and cognitive accessibility, with continuous feedback loops embedded to ensure ongoing improvement.

Finally, pedagogical strategies should be developed to mitigate common AI-generated errors and misconceptions in algebraic problem solving. As AI systems can produce procedural inaccuracies, misuse symbolic notation, or provide mathematically fluent but conceptually incorrect solutions, teachers should explicitly train students to critically evaluate AI-generated outputs by verifying intermediate steps, checking solutions through substitution, and comparing AI responses with formal algebraic principles taught in class. For instance, structured error-analysis activities, in which students identify, classify, and correct AI mistakes, may strengthen metacognitive awareness and conceptual understanding. Furthermore, instructional scaffolds such as guided prompts, constraints on solution formats, and teacher-curated exemplars can reduce overreliance on AI while promoting disciplined mathematical reasoning. Embedding these practices into algebra instruction would support students in using AI as a cognitive aid rather than an authoritative source.

7. Limitations

Despite the contributions of the present study, several limitations should be acknowledged. One limitation concerns the exclusion of visual- or table-based mathematical tasks. To avoid potential misinterpretation by the chatbots and to ensure comparability across responses, all problems were presented in a text-only format. However, many algebra tasks in authentic classroom settings involve tables, diagrams, graphs, or other visual representations that support conceptual understanding and strategy selection. The absence of such representations may limit the ecological validity of the findings and restrict the extent to which the results reflect real classroom problem-solving conditions. Future studies should therefore incorporate multimodal problem formats, including visual and tabular elements, to more comprehensively examine AI-supported mathematical problem solving in realistic instructional contexts. Future studies may examine whether AI model performance varies systematically across different problem-solving strategy types, an analysis that was beyond the scope of the present study.

ORCID iDs

Seda Akti Aslan https://orcid.org/0000-0001-9345-6194
Jordan Allison https://orcid.org/0000-0001-8513-4646

Ethics Considerations

This study was reviewed and approved by the Social and Human Sciences Ethics Committee of Fırat University (Meeting Date: 13 June 2025, Meeting Number: 2025/13, Decision No: 29).

Consent to Participate

Written informed consent was obtained from the parents/guardians of all student participants, and verbal assent was obtained from the students themselves. Participation was voluntary, and confidentiality was assured.

Funding

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

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Declaration of Generative AI Use in the Writing Process

During the preparation of this work, the authors used Claude (Anthropic) solely to improve the language and readability of the manuscript (grammar and English-language editing). After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

Adıgüzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, *15*(3), ep429. https://doi.org/10.30935/cedtech/13152

Ali, O., Murray, P. A., Momin, M., Dwivedi, Y. K., & Malik, T. (2024). The effects of artificial intelligence applications in educational settings: Challenges and strategies. *Technological forecasting and social change*, *199*, 123076. https://doi.org/10.1016/j.techfore.2023.123076

Aliu, J., & Aigbavboa, C. (2020). Reviewing Problem-Solving as a Key Employability Skill for Built Environment Graduates. In *Collaboration and Integration in Construction, Engineering, Management and Technology* (pp. 399–403). https://doi.org/10.1007/978-3-030-48465-1_67 (Accessed: March 2025.

Alnasib, B., & Alharbi, N. (2024). Challenges and Motivation: Assessing Gemini's Impact on Undergraduate EFL Students in Classroom Settings. *World Journal of English Language*, *14*(5), 501–514. https://doi.org/10.5430/wjel.v14n5p501

Antony, S., & Ramnath, R. (2023). A Phenomenological Exploration of Students' Perceptions of AI Chatbots in Higher Education. *IAFOR Journal of Education*, *11*(2), 7–38. https://doi.org/10.22492/ije.11.2.01

Arslan, K. (2017). Artificial Intelligence in Education and Its Applications. *Western Anatolian Journal of Educational Sciences*, *11*(1), 71–88. https://dergipark.org.tr/tr/download/article-file/1174773

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*(3), 245–274. https://doi.org/10.1023/A:1022103903080

Arvin, N., Hoseinabady, M., Bayat, B., & Zahmatkesh, E. (2023). Teacher Experiences with AI-based Educational Tools. *AI and Tech in Behavioral and Social Sciences*, *1*(2), 26–32. https://doi.org/10.61838/kman.aitech.1.2.5

Atmosukarto, I., Sin, C. W., Iyer, P., Tong, N. H., & Yu, K. W. P. (2021). Enhancing adaptive online chemistry course with AI-chatbot. In: 2021 IEEE International Conference on Engineering, Technology & Education (TALE). (pp. 838–843). IEEE.

Baig, M., & Yadegaridehkordi, E. (2024). ChatGPT in the higher education: A systematic literature review and research challenges. *International Journal of Educational Research*, *127*, 102411. https://doi.org/10.1016/j.ijer.2024.102411

Bain, K. (2004). *What the best college teachers do*. Harvard University Press.

Baki, A. (1996). Is the computer everything in mathematics teaching? *Hacettepe University Journal of Education*, *12*(12), 135–143. http://www.efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/1276-published.pdf

Baki, A. (2008). *Mathematics education from theory to practice*: Harf Education Publishing.

Bessas, N., Tzanaki, E., Vavougios, D., & Plagianakos, V. P. (2025). Comparative analysis of ChatGPT and Gemini; Implications for junior high school physics education: Opportunities and ethical challenges. *International Journal of Advanced Multidisciplinary Research and Studies*, *5*(1), 7–18. https://doi.org/10.62225/2583049X.2025.5.1.3610

Bhullar, P., Joshi, A., & Chugh, R. (2024). ChatGPT in higher education - a synthesis of the literature and a future research agenda. *Education and Information Technologies*, *29*(16), 21501–21522. https://doi.org/10.1007/s10639-024-12723-x

Chang, C. Y., Hwang, G. J., & Gau, M. L. (2022). Promoting students' learning achievement and self-efficacy: A mobile chatbot approach for nursing training. *British Journal of Educational Technology*, *53*(1), 171–188. https://doi.org/10.1111/bjet.13158

Desoete, A., & De Craene, B. (2019). Metacognition and mathematics education: An overview. *ZDM*, *51*(4), 565–575. https://doi.org/10.1007/s11858-019-01060-w

Dumont, H., & Istance, D. (2010). *The nature of learning: Using research to inspire practice*. OECD Publishing.

Engelbrecht, J., & Borba, M. (2024). Recent developments in using digital technology in mathematics education. *ZDM Mathematics Education*, *56*(2), 281–292. https://doi.org/10.1007/s11858-023-01530-2

Ferrara, S., & Qunbar, S. (2022). Validity arguments for AI-based automated scores: Essay scoring as an illustration. *Journal of Educational Measurement*, *59*(3), 288–313. https://doi.org/10.1111/jedm.12333

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, *48*(1), 56–66. https://doi.org/10.1080/00461520.2012.748005

Fisher, R. (1990). *Teaching children to think*. Stanley Thornes Ltd.

Grouws, D. A. (2006). *Handbook of research on mathematics teaching and learning*. Macmillan.

Heung, Y. M. E., & Chiu, T. K. (2025). How ChatGPT impacts student engagement from a systematic review and meta-analysis study. *Computers and Education: Artificial Intelligence*, *8*, 100361. https://doi.org/10.1016/j.caeai.2025.100361

Holmes, W. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Holmes, W., & Tuomi, I. (2022). State of the art in AI in education. *European Journal of Education*, *57*(4), 542–570. https://doi.org/10.1111/ejed.12533

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of computer assisted learning*, *38*(1), 237–257. https://doi.org/10.1111/jcal.12610

Istance, D., & Kools, M. (2013). OECD work on technology and education: Innovative learning environments as an integrating framework. *European journal of education*, *48*(1), 43–57. https://doi.org/10.1111/ejed.12017

Kerimbayev, N., Adamova, K., Shadiev, R., & Altinay, Z. (2025). Intelligent educational technologies in individual learning: a systematic literature review. *Smart Learning Environments*, *12*(1), 1. https://doi.org/10.1186/s40561-024-00360-3

Kohnke, L., Zou, D., Ou, A. W., & Gu, M. M. (2025). Preparing future educators for AI-enhanced classrooms: Insights into AI literacy and integration. *Computers and Education: Artificial Intelligence*, *8*, 100398. https://doi.org/10.1016/j.caeai.2025.100398

Lampou, R. (2023). The integration of artificial intelligence in education: Opportunities and challenges. *Review of Artificial Intelligence in Education*, *4*, e15. https://doi.org/10.37497/rev.artif.intell.educ.v4i00.15

Lestari, W., Pratama, L. D., & Jailani, J. (2019). Metacognitive skills in mathematics problem solving. *Daya Matematis: Jurnal Inovasi Pendidikan Matematika*, *6*(3), 286–295. https://doi.org/10.26858/jds.v6i3.8537

Luckin, R., & Holmes, W. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education.

Mikolov, T., Deoras, A., Povey, D., Burget, L., & Černocký, J. (2011). Strategies for training large scale neural network language models. In In IEEE Workshop on Automatic Speech Recognition and Understanding (pp. 196–201).

Ministry of National Education [MoNE]. (2024). *Mathematics curriculum*: MoNE Publishing.

Msambwa, M. M., Daniel, K., & Lianyu, C. (2024). Integration of information and communication technology in secondary education for better learning: A systematic literature review. *Social Sciences & Humanities Open*, *10*, 101203. https://doi.org/10.1016/j.ssaho.2024.101203

National Council of Teachers of Mathematics [NCTM]. (2008). *Principles and standards for school mathematics*: NCTM.

Posamentier, A. S., & Krulik, S. (2008). *Problem-solving strategies for efficient and elegant solutions, grades 6-12: a resource for the mathematics teacher*. Corwin press.

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Shuhaiber, A., Kuhail, M. A., & Salman, S. (2025). ChatGPT in higher education-A Student's perspective. *Computers in Human Behavior Reports*, *17*, 100565. https://doi.org/10.1016/j.chbr.2024.100565

Silva, C. A. G. D., Ramos, F. N., De Moraes, R. V., & Santos, E. L. D. (2024). ChatGPT: Challenges and benefits in software programming for higher education. *Sustainability*, *16*(3), 1245. https://doi.org/10.3390/su16031245

Son, T. (2024). Intelligent tutoring systems in mathematics education: A systematic literature review using the substitution, augmentation, modification, redefinition model. *Computers*, *13*(10), 270. https://doi.org/10.3390/computers13100270

Van De Walle, J. A. (1989). *Elementary school mathematics: Teaching developmentally*: Longman.

Van de Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2013). *Elementary and middle school mathematics: Teaching developmentally* (8th ed.). Pearson Higher Ed.

Williamson, B., Eynon, R., & Potter, J. (2020). Pandemic politics, pedagogical possibility: Social impacts of COVID-19. *Learning, Media and Technology*, *45*(2), 107–114. https://doi.org/10.1080/17439884.2020.1761641

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, *16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Appendix

Appendix 1

Matematiksel Problem Çözme Soruları (Turkish).

1. Geçen cuma akşamı Nehir, evinde bir akşam yemeği düzenlemiş ve 17 arkadaşını davet etmiştir. Nehir, misafirlerin her birine 1′den 18′e kadar numaralandırılmış kâğıtlardan birini vermiş, kalan bir kâğıdı ise kendisi almıştır. Yemek masasında misafirleri ikili gruplar hâlinde eşleştirdiğinde, her bir ikilideki sayıların toplamının tam kare bir sayı olduğunu fark etmiştir. Buna göre Nehir'in elindeki kâğıtta yazan sayı kaçtır?
2. Toplamları 12 ve çarpımları 4 olan iki reel sayı veriliyor. Bu iki sayının çarpımsal terslerinin toplamını bulunuz.
3. n sayısının birler basamağını bulunuz.
4. a ve b, $\{0,1,2,3,4,5,6,7,8,9\}$ kümesinden seçilen ve birbirinden farklı iki rakam olsun.
   ab ifadesinin alabileceği en büyük değeri bulunuz.
5. $x = \ldots$ verildiğine göre, y'nin değerini bulunuz.
   (Not: Orijinal soruda verilen matematiksel ifade aynen korunmalıdır.)
6. 360 sayısının pozitif bölenlerinin toplamı 1170′tir. Buna göre 360 sayısının tüm pozitif bölenlerinin terslerinin toplamı kaçtır?

7. Bir çekmecede 8 mavi, 6 yeşil ve 12 siyah çorap bulunmaktadır. Çekmeceden bakmadan çorap çekildiğinde, aynı renkten iki çorabın kesin olarak elde edilebilmesi için en az kaç çorap çekilmelidir?
8. Bir sayının $\frac{3}{5}$'i 12 ise, bu sayının $\frac{2}{3}$'ü kaçtır?
9. x ve y pozitif tam sayılar olmak üzere $xy = \ldots$ verilmiştir. Buna göre x ve y'nin değerlerini bulunuz. *(Not: Orijinal matematiksel koşul aynen yazılmalıdır.)*
10. Songül ve Hasan bir tenis turnuvasında karşı karşıya gelmektedir. Bu maç sisteminde, bir oyuncu art arda iki oyunu kazanırsa ya da toplamda üç oyun kazanırsa maçı kazanmış sayılmaktadır. Buna göre maç kaç farklı şekilde sona erebilir?

Appendix 2

Mathematical Problem-Solving Questions.

1) Last Friday evening, Nehir invited 17 friends to a dinner party at her home. She gave each guest a paper labeled with a number from 1 to 18 and kept the remaining paper for herself. When she paired up all the guests at the dinner table, she noticed that the sum of the numbers in each pair was a perfect square. Based on this observation, what is the number on Nehir's paper?
2) Let x and y be two real numbers such that their sum is 12 and their product is 4. Find the sum of the multiplicative inverses of these two numbers.
3) Determine the units digit of 8^{19}.
4) Let a, b, c and d be distinct digits, each selected from the set $\{1, 2, 3, 4\}$. What is the maximum possible value of the expression $.b + b.c + c.d + a.d$.
5) Given that $\frac{1}{x+4} = 4$, find the value of $\frac{1}{x+6}$.
6) The sum of the divisors of 360 is 1170. What is the sum of the reciprocals of all the divisors of 360?
7) A drawer contains 8 blue socks, 6 green socks, and 12 black socks. What is the minimum number of socks that must be drawn blindly from the drawer to guarantee obtaining a matching pair of the same color?
8) If $\frac{1}{8}$ of a number is $\frac{1}{5}$, what is $\frac{5}{8}$ of that number?
9) Let x and y be positive integers such that $\frac{x}{4} + \frac{y}{5} = \frac{19}{20}$. Determine the values of x and y.
10) Songül and Hasan are opponents in a tennis tournament. In this match format, the first player to either win two consecutive games or a total of three games wins the match. In how many different ways can the match end?