

Examination of Grade 7 Ratio Concept Tasks Designed by ChatGPT Based on Cognitive Demand Levels

Merve Koçyiğit Gürbüz^{1*} 
Kübra Alan² 
Bahadır Yıldız² 

¹ Ankara Provincial Directorate of National Education, Ankara, Türkiye, mervekociyigitgurbuz@gmail.com

² Hacettepe University, İstanbul, Türkiye, kubraalan37@gmail.com, bahadir@bahadiryildiz.net

*Corresponding author

Received: 01.02.2024
Accepted: 01.07.2024
Available Online: 24.07.2024

Abstract: The process that enables students to reach the gains in a mental and physical way while doing the mathematical tasks given under the supervision of the teacher is defined as an activity. The student's state of being active in the implementation process of the activities positively affects the permanence and efficiency of teaching. The process of creating tasks is a long and challenging journey as it involves a series of complex stages. Today, artificial intelligence technologies are utilized to facilitate this process. ChatGPT, one of the artificial intelligence tools, can support the teacher in preparing content, creating a lesson plan, evaluation process, etc., as well as supporting the teacher in creating activity design. It is important to understand how much the created tasks direct students to be cognitively active and to examine their cognitive demand levels in order to manage the teaching process efficiently and to improve the scope of the tasks. The aim of this study is to evaluate the tasks prepared for the 7th grade ratio concept through ChatGPT within the framework of cognitive demand levels. In the study, a case study design was utilized to examine the cognitive demand levels of the tasks created by ChatGPT 3.5 and ChatGPT 4, to determine the functionality of the activities, and to determine the deficiencies, if any, and how they can be eliminated. Content analysis was used to analyze the data obtained. It was concluded that the tasks created can be used by making necessary improvements, if any; thus, teachers can enrich their course content by saving time and energy. In this context, it is recommended to carry out tasks that will encourage the use of artificial intelligence tools by teachers in the world of education.

Keywords: Mathematical task, activity, cognitive demand level, ratio, chatgpt

1. Introduction

Numerous methods, techniques, and strategies are used to improve the quality of learning and teaching processes. Tasks, which are fundamental elements in achieving this goal, enable students to actively achieve the objectives both mentally and physically while performing mathematical tasks and additional activities either individually or in groups under teacher supervision (Nayir & Bulut, 2022). Tasks are designed, selected, or organized according to needs and aligned with learning objectives. The appropriate use of tasks in teaching processes holds significant importance. The tasks may aim to teach, reinforce, or evaluate a concept (Dede, Doğan, & Aslan-Tutak, 2021). Two perspectives on task use in mathematics education are notable. The concept of task involves problem-solving or activity processes related to a concept or subject that has been taught (Van de Walle et al., 2019), whereas the concept of activity focuses on making students active participants in learning a subject or concept (Dede, Doğan, & Tutak, 2020). From this perspective, tasks provide opportunities for students to develop skills such as reasoning and modeling in learning and teaching environments (Stein, Grover, & Henningsen, 1996).

Tasks inherently involve activity (Doyle, 1983); they not only enhance interaction, cooperation, and problem-solving skills (Atara et al., 2000) but also positively affect the retention and effectiveness of teaching (Lugosi & Urobi, 2022). Understanding the extent to which tasks engage students cognitively is crucial for managing teaching processes effectively and refining the scope of the tasks. Hence, determining the cognitive demand levels of the tasks is essential (Engin, 2015). The cognitive demand level was defined by Stein et al. (2000) as “the type and level of thinking required for students to successfully deal with and solve the task.” According to this framework, tasks are classified into four cognitive demand levels: “memorization tasks,” “procedures without connection tasks,” “procedures

with connection tasks,” and “doing mathematics tasks.” These levels are referred to as “cognitive demand levels.” Studies suggest that cognitive demand levels can decrease or increase during the preparation and implementation of tasks in the classroom (Engin, 2015; Henningsen & Stein, 1997; Sarpkaya, 2011; Stein et al. 1996). Appropriate cognitive demand levels need to be maintained during task preparation and implementation (Henningsen & Stein, 1997). Tasks should ideally involve higher-order cognitive thinking strategies to deepen students’ understanding of mathematical relationships and concepts (Stein et al., 2000). Therefore, tasks should be designed to maximize cognitive demands during preparation and implementation (Ecemiş, 2017; Ubuz et al., 2010).

Mathematical tasks typically serve three purposes: “teaching a concept,” “reinforcing a concept,” and “evaluating the learning of a concept” (Pekkan & Esmer, 2021). Designing mathematical tasks is pivotal in terms of process planning and operational paradigm. This involves analyzing and designing the objectives, determining the task design purposes, evaluating student knowledge, selecting relevant examples, drafting instructions, assessing tasks, and planning task implementation (Yeşildere, 2022).

Given the multifaceted nature of task preparation, leveraging tools that streamline this intricate and time-consuming process is essential.

Artificial intelligence technologies are instrumental in various demanding tasks such as coding, design of plans for content marketing, test creation, and data analysis. They are also employed in specialized domains such as disease diagnosis and case file preparation. They are being used in many fields from health to engineering, and from industry to marketing (Korucu & Biçer, 2020). These technologies are increasingly integrated into education, enhancing effectiveness and efficiency through capabilities such as generating exam questions, drafting education texts, providing detailed explanations, analyzing exam results, and employing diverse teaching techniques (Arslan, 2020). New concepts and professional groups have emerged with the development of artificial intelligence. Prompt engineering, a burgeoning discipline within artificial intelligence, focuses on optimizing instructions to ensure that artificial intelligence models provide appropriate responses (Breuss, 2023).

Artificial intelligence technologies, including chatbots such as ChatGPT, play a pivotal role in education by supporting student learning and alleviating teacher burdens in tasks such as question creation, content selection, and lesson planning (Pressey, 1950). The use of artificial intelligence technologies, such as personalized education, chatbots, intelligent agents, and artificial intelligence-supported assessment systems (Arslan, 2020), provides opportunities to improve the teaching processes such as personalized student learning and rapid response to diverse inquiries during lessons. Developed by OpenAI, ChatGPT is a dialogue-specialized artificial intelligence chatbot prototype (Kelly, 2022). The prompts presented as input to the system determine the scope of the output (Breuss, 2023). In other words, providing the input most appropriately for the desired output ensures that the resulting output is aligned with the intended goal.

Initially released as ChatGPT 3 in November 2022 (Edwards, 2023), ChatGPT 3.5 has since evolved with updated settings using supervised and reinforcement learning. The most recent OpenAI iteration is ChatGPT 4 (Terrasi, 2023). ChatGPT 3.5 is a freely available language model that uses deep learning to generate content similar to human-written text that can answer many questions (Scharth, 2022). It can also perform complex operations such as writing lyrics, writing poetry, and coding (Scharth, 2022). In comparison, ChatGPT 4 surpasses its predecessor, ChatGPT 3.5, by handling more complex inputs, including text, images, and videos, while delivering faster, more secure, and content-relevant responses (OpenAI 2023b). ChatGPT provides users with various responses aligned with their queries, enabling students to expand the scope of questions or explore different frameworks for answers. ChatGPT aids teachers in content creation, lesson planning, question formulation, evaluation processes, and task

design. Therefore, teachers can leverage artificial intelligence technologies to save time, enhance energy efficiency, and improve educational outcomes.

Creating tasks for classroom use is often challenging and time-intensive for teachers. To alleviate these challenges, the researchers prepared mathematical tasks using ChatGPT, one of the artificial intelligence tools, and examined their cognitive demand levels.

1.1 Aim of the Study

This study aimed to evaluate the tasks prepared by ChatGPT 3.5 and ChatGPT 4 for the Grade 7 ratio concept within the framework of cognitive demand levels. In line with this purpose, the study sought to answer the following questions:

1. How compatible are the tasks prepared by ChatGPT 3.5 and ChatGPT 4 with the given objective?
2. How suitable are these tasks for the given age group/grade level?
3. Do the tasks differ according to cognitive demand levels?

Additionally, the study aimed to identify the advantages and disadvantages of ChatGPT 3.5 and ChatGPT 4 versions in creating tasks.

2. Method

This was a qualitatively designed case study. A case study is a systematic design including steps such as collecting information, organizing the collected information, interpreting it, and reaching research findings. Its aim is to provide the best understanding of a situation rather than generalization (Denzin & Lincoln, 1994; Merriam, 1988). What distinguishes this method from many other research methods is its preference for understanding various topics in education, especially when questions of what, how, and why are asked (Çepni, 2012; Yin, 2003). This study used a case study design to examine the cognitive demand levels of the tasks created by ChatGPT 3.5 and ChatGPT 4, determine the functionality of the tasks, and identify the deficiencies and ways to address them.

2.1 Data collection

This study created tasks for the following objectives using two different versions of ChatGPT (3.5 and 4).

The relevant objectives from the Grade 7 curriculum related to the concept of ratio were as follows:

M.7.1.4.1 Determine the value the other will take if one of the multiplicities in the ratio is 1.

M.7.1.4.2 Find the other when one of the two multiplicities and the ratio between the multiplicities are provided.

Three different types of tasks were created: introduction, reinforcement, and evaluation. The prompts directed to ChatGPT during the task creation process were as follows:

1. What does the statement “Determine the value that one of the multiplicities in the ratio will take if one of them is 1” mean? Can you explain it with examples and an introduction and reminder task?
2. “Given one of two multiplicities in ratio to each other, find the other.” This objective belongs to the Grade 7 mathematics course. Can you think like a math teacher and prepare a deepening task related to this objective?
3. “Find the other when given one of two multiplicities their ratio to each other.” This objective belongs to the Grade 7 mathematics lesson. Now, can you think like a math teacher and prepare an assessment task related to this objective that we can use at the end of the lesson?

Each question was directed to both ChatGPT 3.5 and ChatGPT 4, resulting in six different tasks.

2.2 Data analysis

The content analysis method was used in this study. This involved bringing together similar data within the framework of certain concepts and themes and organizing and interpreting them in a way that is understandable to the reader (Creswell, 2012).

For the coding during the content analysis, the levels in the Task Analysis Guide (Stein et al., 2000) were used, as the Guide was deemed appropriate for the research purpose. The Guide was developed in line with a research project called QUASAR, conducted between 1990 and 1995. It was prepared to determine the cognitive demand level of hundreds of mathematical tasks found in the mathematics textbooks of the schools involved in the project and implemented in classrooms. The framework for coding the tasks according to cognitive demand levels, based on the Guide (Stein et al., 2000, p. 16, trans., Reçber, 2012), is presented in Table 1. The Guide contains a list of cognitive demand characteristics at each level and is used as a rating tool applicable to any mathematical task (Smith & Stein, 1998).

Table 1

Cognitive Demand Levels of Tasks Based on the Task Analysis Guide

LOW-LEVEL DEMANDS (LLD)	
Memorization Tasks (LLD-M)	The use of previously learned or given information, rules, or formulas involves remembering definitions. These tasks cannot be solved using a method because either no method exists to apply or the time interval is too short to apply a method. No ambiguity is involved. Such tasks are straightforward repetitions of previously learned material. The information to be repeated is clear and unambiguous. No association of the information, rules, formulas, or definitions with their underlying concepts and meanings is observed.
Procedures Without Connection (LLD-PC)	These tasks are algorithmic. The method to be used is explicitly requested or is evident from prior instruction or the context of the task. The cognitive demand for successful completion is limited. A little uncertainty exists regarding what to do and how to do it. No association with the underlying concept and meaning of the method used is observed. The focus is on producing correct answers rather than developing mathematical understanding. These tasks do not require an explanation or only require an explanation of the procedure.
HIGH-LEVEL DEMANDS (HLD)	
Procedures with Connection (HLD-PWC)	These tasks focus students' attention on the method to foster a deep understanding of mathematical concepts and ideas. They implicitly or explicitly suggest the application of general methods closely related to the underlying conceptual ideas. Often, these tasks present multiple representations, such as visual diagrams, dexterous situations, symbols, and problem situations. Making connections between multiple representations helps improve understanding. These tasks require a certain level of cognitive effort. Although general methods are followed, these tasks cannot be completed without a thinking process. Students need to engage with the conceptual ideas underlying the methods to be able to complete and understand the task.

Doing Mathematics Tasks
(HLD-DM)

These tasks require complex and non-algorithmic thinking, implying no clearly stated, repeated, predictable, or suggested approach in the task, its instructions, or examples. Students must explore and understand the nature of mathematical concepts, operations, and relationships. They need to observe and organize their own cognitive processes, access relevant knowledge and experiences, and use them appropriately in their work. Students must analyze the task, evaluate possible solution strategies, and consider the limitations of those solutions. Completing these tasks requires significant cognitive effort, and students may experience some anxiety due to the unpredictable nature of the solution process.

Source: Stein et al. (2000, p. 16).

Four levels of cognitive demands are defined in the Efficiency Analysis Guide, and their corresponding codes are presented in Table 2.

Table 2

Codes Used for Cognitive Demand Levels of Tasks

Level of cognitive demands	Abbreviation
LOW-LEVEL DEMANDS – Memorization Tasks	LLD-M
LOW-LEVEL DEMANDS – Procedures Without Connection	LLD-PC
LOW-LEVEL DEMANDS – Procedures with Connection	HLD-PWC
LOW-LEVEL DEMANDS – Doing Mathematics Tasks	HLD-DM

To enhance coding reliability, the first researcher coded the data initially, followed by an independent coding and comparison by the second researcher. Then, the coders discussed any discrepancies in the levels and reached a consensus on the common levels.

3. Findings

In this section, the findings from the analyzed data are examined within the framework of the questions posed to ChatGPT and then interpreted in relation to the research problems.

3.1 Question 1

QUESTION 1: *What does the statement “Determine the value that the other will take if one of the multiplicities in the ratio is 1” mean? Can you explain it with examples and an introductory and recall task?*

3.1.1 ChatGPT 3.5

The statement "Determine the value of one of the multiplicities in the ratio when one of them is 1" indicates that when one component of a ratio is set to 1, the corresponding value of the other component can be determined proportionally. This objective aims to gain a deeper understanding of the concept of ratio.

Introductory task: To introduce the lesson, the teacher can assign students a task where they calculate the ratios of different objects or scenarios. For example, students may be asked to determine the ratio of red balls to blue balls when 6 red and 12 blue balls are present in a box. This task helps students grasp how to determine the value of one component of a ratio when the other is set to 1.

Recall task: Following the lesson, the teacher should present students with various ratios and prompt them to explain which component's value is determined when one part is set to 1. Encouraging classroom discussion allows students to revisit the fundamental principles of proportion. Providing a

platform for students to exchange ideas helps them develop a deeper understanding of the concept of ratio.

3.1.2 ChatGPT 4

This objective involves understanding the characteristics of ratios and predicting the value of one part when the other is 1 in ratios.

3.1.2.1 Introductory Task: "Let's Make Pizza Together"

1. Before starting the lesson, students engage in a pizza-making task. Each student receives materials, such as cardboard pizza bases, glue, scissors, and colored paper.
2. Students are instructed to determine the amount of ingredients needed to make a pizza slice. For example, they can decide to use two sheets of green paper, one sheet of red paper, and one sheet of ...
- ... brown paper for each pizza slice. 3. Students calculate and record these ingredient proportions in a table.

3.1.2.2 Recall Task: "Pizza Slices on Sale!"

1. In the next class, students learn that each pizza slice sells for 2 dollars.
2. Students are asked to determine the cost of one slice of pizza based on the ratio and predict the cost if it were 1. They record their findings in a table.

Table 3

Cognitive Demand Levels of Introductory/Recall Tasks

Task Type	Introductory/Recall	
ChatGPT version	GPT 3.5	GPT 4
Cognitive demand level	LLD-M	HLD-PWC

When the introductory task prepared by ChatGPT 3.5 was analyzed within the framework of cognitive demand levels, it was determined to be at the level of LLD-M. This approach was chosen because, after presenting the initial situation for students to apply, a new but similar situation was given. Students were expected to apply the same approach again, essentially repeating the initial task. This repetition aligned with the characteristics of a memorization task.

The recall task from ChatGPT 3.5 was found to be at the level of HLD-PWC. This task required students to use their prior knowledge to explain and discuss the situation through examples, encouraging them to establish relationships between events and reason through different scenarios. This aligned with the level of connected methods.

Both tasks created by ChatGPT 3.5 for the first question were compatible with the content of the learning objective and appropriate for the age group/grade level. They effectively prepared content for student comparisons between units and included concrete elements.

For the tasks designed by Chat GPT 4, both the introductory and recall tasks were analyzed according to their cognitive demand levels and found to be at the level of HLD-PWC. In both tasks, students were expected to recognize the relationships between concepts using their basic knowledge. Creating a table with the new information required students to use multiple representations, demanding a certain level of cognitive effort.

The scope of both tasks created by ChatGPT 4 for Question 1 was also compatible with the content of the learning objective and the age group/grade level. These tasks provided students with the opportunity to analyze the relationships between the specified ratios and included examples from daily life, enhancing their understanding of the unit.

3.2 Question 2

QUESTION 2: “Find the other when given one of two multiplicities given their ratio to each other.” This objective belongs to the Grade 7 mathematics lesson. Now, can you think like a math teacher and prepare a deepening task related to this objective?

3.2.1 ChatGPT 3.5

Of course! Here is a math task that you can prepare to deepen the understanding of the statement “Given one of two multiplicities in proportion to each other, find the other.”

Task Name: Ratio-Based Puzzle

Target Objective: To develop the ability to determine one in proportion to another given their ratio.

Task Description:

1. Divide the class into four equal groups and provide each group with a piece of paper and a pencil.
2. Assign each group a question that involves determining one of two multiplicities based on a given ratio. For example: “There are 24 students in a class. Of these students, $\frac{3}{4}$ are girls. How many girls are there in the class?” In this question, $\frac{3}{4}$ represents the ratio of girls to total students. Students must calculate the number of girls based on this ratio. Note: You can also give different questions to other groups. In this way, each student will work on a different question, and at the end, they will share their answers.
3. Allow time for each group to solve their respective question using correct proportional reasoning on their paper.
4. Each group presents their solution to the class, explaining how they have arrived at their answer.
5. One student from each group will write their solution on the board or a large piece of paper and explain the process.
6. Students from other groups can ask questions and discuss their approaches to understanding the solution.
7. Review the correct answers together, addressing any misunderstandings or mistakes.
8. Conclude by offering additional examples or questions about ratios for further practice.

This task aims to enhance students’ critical thinking and problem-solving skills about ratios, while also fostering collaboration and discussion among peers.

3.2.2 ChatGPT 4

Of course! Here is a deepening task to help students master “Find one of the two multiplicities given their ratios to each other.”

Task Name: Ratio Puzzles

Objective: To enable students to understand the concept of ratios and develop their ability to find one of two multiplicities given their ratio to each other.

Materials:

- Paper and pencil

Event Duration: Approximately 15 min

Task Stages:**Step 1: Reminding the Concept of Proportion (5 min)**

In class, have a brief discussion with the students to review the concept of proportion. Review the definition of a ratio and demonstrate its use with examples. Ensure students understand the concept of ratios.

Step 2: Ratio Puzzles Questions (10 min) Ask students to take their paper and pencils and solve the ratio puzzles. They need to find the other multiple based on the given ratio in each question. You can adjust the difficulty level of the questions according to the students' needs. Here are some sample questions:

Question 1: In a restaurant, there are nine chairs at a table for six people. If another table has 12 seats, how many chairs should be at this table?

Question 2: Three tablespoons of coffee are used for one cup of coffee. How many cups of coffee can be made using nine tablespoons of coffee?

Table 4

Cognitive Demand Levels of Deepening Tasks

Task type	Deepening	
ChatGPT version	GPT 3.5	GPT 4
Cognitive demand level	LLD-M	LLD-PC

When the deepening task prepared by ChatGPT 3.5 was examined within the framework of the cognitive demand level, the task was found to be at the level of LLD-M. The main purpose of this task was to solve the question given in Step 2; in this step, students are expected to reach the answer using only basic knowledge rather than establishing a relationship between old knowledge and concepts. In the following steps of the task, students are suggested to discuss their answers. However, this discussion does not raise the level of the task, as it is more about checking the results rather than a new situation.

For Question 2, it was determined that the scope of the deepening task created with ChatGPT 3.5 was compatible with the content of the learning target and the age group/grade level. The content was designed to emphasize that a proportion occurred when two ratios were equal.

When this task prepared by ChatGPT 4 for the deepening task was analyzed according to the cognitive demand levels, it was evaluated as LLD-PC. The information requested by the question was present in prior knowledge and did not have any ambiguity. It focused on producing correct answers instead of developing a deeper mathematical understanding. However, the process design of the task was quite detailed, and it was thought to facilitate the teacher in shaping the lesson plan.

For Question 2, it was determined that the scope of the deepening task created with ChatGPT 4 was compatible with the content of the learning objectives and the age group/grade level, as it allowed the students to establish the relationship between two multiplicities and was supported by concrete examples.

3.3 Question 3

QUESTION 3: *“Find the other when given one of the two multiplicities and the ratio between them.” This objective belongs to the Grade 7 mathematics lesson. Now, can you think like a math teacher and prepare an evaluation task related to this objective for use at the end of the lesson?*

3.3.1 ChatGPT 3.5

Of course! Here is a math task that you can use to assess the understanding of “Given one of two multiplicities and the ratio of the two, find the other.”

Task Name: Ratio-Finding Questions

Objective: To evaluate the ability to find one of two multiplicities given the ratio of one to the other.

Task Description:

1. Arrange the class for individual work. Provide each student with a piece of paper and a pencil.

2. Hand out a test or worksheet with ratio-based questions. Sample questions include:

(a) A store has 120 apples and 80 oranges. What is the ratio of apples to oranges? (b) A basket contains 36 red, 48 green, and 72 blue balls. What is the ratio of red balls to green balls? (c) A music player contains 250 songs. If $\frac{3}{5}$ of these songs are rock songs, how many are pop songs? (d) A country has 500,000 men and 600,000 women. What is the ratio of men to women?

3. Encourage students to create their own real-life scenarios involving ratios and share these with the class. For example, they can discuss ratios related to product prices in a supermarket.

3.3.2 ChatGPT 4

Of course! Here is an assessment task for the skill of “Finding one of two multiplicities given their ratios.”

Task Name: Ratio Investigation

Objective: To evaluate students’ ability to use ratios to find one of the two multiplicities when the other is provided.

Materials:

-Paper and pencil

Event Duration: Approximately 15 min

Task Stages:

Step 1: Creating a Series of Questions (10 min)

Develop a set of questions that test students’ ability to find one of two multiplicities given its ratio to another. Sample questions:

Question 1: A team can clear an area of 9 m in 12 h. Given this ratio, how many meters can the same team clear in 15 h?

Question 2: There are 120 trees in a park. One-third of these trees are fruit trees. How many fruit trees are there?

Step 2: Solving the Questions (5 min)

Allow students time to solve the questions you have prepared. Ask them to find the unknown number using the given ratio. Have students write their solutions on their papers.

Table 5

Cognitive Demand Levels of Evaluation Tasks

Task Type	Evaluation	
ChatGPT version	GPT 3.5	GPT 4
Cognitive demand level	LLD-PC	LLD-PC

When the evaluation task created by ChatGPT 3.5 was analyzed according to the cognitive demand levels, it was found to be at the level of LLD-PC. The question content addressed the information available from prior learning, and the method to be applied was clear, causing the task level to remain at this level. In addition, the questions in the task were found to be answer-oriented rather than directing students toward higher-level thinking skills.

For Question 3, it was determined that the scope of the evaluation task created with ChatGPT 3.5 was compatible with the content of the learning target and the age group/grade level, as it established proportional relationships and included relevant instructions.

When the task prepared by ChatGPT 4 for the evaluation task was analyzed according to the cognitive demand levels, it was determined to be at the level of LLD-PC. The task focused on reaching the correct answer using basic knowledge rather than establishing relationships with different concepts. In addition, little uncertainty existed about what needed to be done in the task content.

For Question 2, it was determined that the scope of the evaluation task created with ChatGPT 4 was compatible with the content of the learning target and the age group/grade level. This task reinforced proportional thinking by having students first find the ratios by establishing a relationship between the given multiplicities and then equalizing them.

The cognitive demand levels of the tasks prepared by ChatGPT are summarized in Table 6.

Table 6

Cognitive Demand Levels of All Tasks

Task Type	Introductory/Recall		Deepen		Evaluation	
	GPT 3.5	GPT 4	GPT 3.5	GPT 4	GPT 3.5	GPT 4
Cognitive demand level	LLD-M	HLD-PWC	LLD-M	LLD-PC	LLD-PC	LLD-PC

As shown in Table 6, only one of the six tasks prepared by ChatGPT was at a high cognitive demand level. This meant that approximately 85% of the tasks remained at the LLD level. Two of the five tasks at the LLD level were at the memorization task level, and both tasks were created by ChatGPT 3.5.

When all the findings were evaluated, it was determined that ChatGPT could prepare mathematical tasks suitable for the targeted objective and grade level. However, the cognitive demand levels of the tasks it prepared remained at low levels.

3.4 Comparison of ChatGPT 3.5 and ChatGPT 4

When the task examples were analyzed, it was evident that the presentation style of the tasks differed according to the version. ChatGPT 4 offered a more systematic design, whereas the explanations made by ChatGPT 3.5 regarding the task level remained at a low level compared with ChatGPT 4. Conversely, the explanations provided by ChatGPT 3.5 were more comprehensive compared with that by ChatGPT 4. These differences between the explanations were seen when comparing the answers to the second question in both ChatGPT 3.5 and ChatGPT 4. This result aligned with the findings of Stojanov (2023). Although ChatGPT 4 explained the implementation process of the task briefly and clearly, ChatGPT 3.5 explained it more indirectly.

When the quality and content of the questions were evaluated, it was noticed that both versions failed to generate tasks related to high-level cognitive demand in mathematics. However, teachers were encouraged to facilitate classroom discussions based on student answers. In addition, both versions included explanations about the skills students would acquire after completing the tasks.

When evaluated within the framework of cognitive demand levels, the tasks prepared by ChatGPT 4 had higher cognitive demand levels. For example, in preparing an introductory task, the task designed by ChatGPT 3.5 was geared toward memorization (low demand levels), whereas the task prepared by ChatGPT 4 was designed for a connected method (high demand levels).

However, despite ChatGPT 4 being based on higher demand levels than ChatGPT 3.5, most of the tasks it prepared remained at low levels. For instance, when examining ChatGPT 4's deepening task design, it was found to be created according to the disconnected method tasks (low demand levels). Similarly, when analyzing the evaluation tasks, it is clear that both versions produced designs that included disconnected method tasks from low demand levels. Therefore, it was interpreted that ChatGPT 3.5 did not create a task design for high demand levels. However, ChatGPT 4 aimed for higher demand levels and did not fully achieve the task of doing mathematics.

In addition, both versions could not create task designs for high-level demand mathematics tasks, also known as doing mathematics tasks.

4. Conclusions and Discussion

In this study, ChatGPT, one of the artificial intelligence tools, was given the objectives that would constitute the course content, and instructions were given to prepare various types of tasks related to these objectives. The tasks created were then evaluated according to their cognitive demand levels. The analysis of the results showed that ChatGPT could design tasks appropriate to the age, grade, objective, and knowledge level of the students. Additionally, it provided various suggestions for the implementation of these tasks. In this context, it was determined that ChatGPT, as an artificial intelligence chatbot, could help teachers based on the content of the questions asked. Thus, while saving time and energy, teachers could enrich their course content. Artificial intelligence tools that can be considered teacher assistants (Brakina et al., 2021) can also be used for various functions, such as interactive games and creating custom questions, translating, and using them as language learning tools, which can actively engage students in the lesson. This situation brings advantages such as increased student motivation and a liberal approach to presenting their own ideas. Some studies reported similar results (Almelweth, 2022; Jones, 2021).

When the tasks prepared with ChatGPT 3.5 and ChatGPT 4 versions, which constituted the first research question, were examined, it was observed that the tasks were created in a way that would guide the teacher and allow for flexible modifications. In addition, it was possible to design the teaching process in different dimensions thanks to features such as changing the level of the tasks and adjusting them to be student-centered or teacher-centered. Thus, the task content could be developed according to the environment in which the task would be implemented; tasks that could adapt to the changes in the

teaching process could be implemented in the classroom. Considering the aforementioned situations, integrating artificial intelligence tools into lessons or using these tools to design course content can add different dimensions to the educational process (Arslan, 2020). Benefiting from artificial intelligence technologies also requires teachers to adapt to information technologies and encourage their students to learn in this direction (Yuan, 2022).

When evaluating the findings related to the second research question, which involved assessing the cognitive demand levels of the tasks prepared with both versions of ChatGPT, it was observed that the cognitive demands of the tasks remained at low levels. This situation suggested that teachers should expand the scope of the tasks or provide more comprehensive prompts to ChatGPT when preparing tasks. In addition, the fact that no tasks including the demand level of "Doing Mathematics Tasks" were created indicated that the chatbot had deficiencies in creating such tasks. In summary, ChatGPT 4 was found to generate tasks at higher demand levels compared with ChatGPT 3.5, but these tasks were not at sufficient cognitive demand levels.

As seen in the tasks created by ChatGPT, it was advantageous that the designed tasks could be directed, organized in a short time, flexibly adapted to the teaching strategy, and promote cooperation between students and teachers (Dwivedi & Joshi, 2021). The purpose of creating tasks was to provide conceptual understanding and foster a discussion process rather than merely assigning homework or solving questions (Özgen, 2017). Tasks that served a single purpose would lead to a uniform transfer of knowledge, causing students' interpretation and solution strategies to be similar and resulting in memorized knowledge (Özgen, 2017). Therefore, when designing tasks, it would be useful to include a guiding scheme for the type of tasks and their implementation in the process. Both ChatGPT versions could prepare content for the type of task asked. However, the tasks prepared by ChatGPT4 were found to be more systematic. The applicability of the prepared content can be considered as an indicator of its use as a helpful tool for teachers. In this way, teachers can save time and energy and enrich their course content. However, it is thought that it would be more beneficial for teachers to revise and use these tasks.

5. Recommendations

Today, it is clear that technology has ushered in a new era. The rapid advancement and the increasing number of artificial intelligence tools direct us to learn how to use these tools effectively. Therefore, it is suggested that efforts should be made to encourage teachers in the field of education to use artificial intelligence tools. These efforts can include in-service seminars or artificial intelligence courses. Moreover, introducing students to these tools and sparking their interest in coding may benefit their future technological developments. Researchers are recommended to conduct further studies on the role of artificial intelligence as a teacher assistant.

Although ChatGPT, which can be an assistant for the teacher, offers flexibility in the content it provides, some factors may be overlooked when applied in real-life situations. These factors include student psychology, classroom environment, and teacher psychology. As ChatGPT can design content independently of these factors, teachers need to consider that the designed content may include elements that do not directly address these factors. Therefore, teachers should revise and adapt the content as needed.

In addition, teachers and researchers who aim to maintain high cognitive demand levels in tasks are recommended to carefully consider those when creating prompts for ChatGPT. By doing so, they can ensure that the tasks generated align with their educational objectives and meet the desired cognitive demand levels.

References

- Almelweth, H. (2022). The effectiveness of a proposed strategy for teaching geography through artificial intelligence applications in developing secondary school students' higher-order thinking skills and achievement. *Pegem Journal of Education and Instruction*, 12(3), 169-176. <https://doi.org/10.47750/pegegog.12.03.18>
- Arslan, K. (2020). Artificial intelligence and applications in education. *The Western Anatolian Journal of Educational Sciences*, 11(1), 71-88. Retrieved from <https://dergipark.org.tr/en/pub/baebd/issue/55426/690058>
- Atara, S., Roberto, W. L., Woon, C., & Kamber, D. (2000). An implementation of active learning and its effect on the quality of student learning. *Innovations in Education and Training International*, 37(4), 381-389. <https://doi.org/10.1080/135580000750052991>
- Barakina, E. (2021). Digital technologies and artificial technologies in education. *European Journal of Contemporary Education*, 10(2), 285- 296. <https://doi.org/10.13187/ejced.2021.2.285>
- Breuss, M. (2023). *Prompt engineering: A practical example*. Retrieved from <https://realpython.com/practical-promptengineering/>
- Çepni, S. (2012). *Araştırma ve proje çalışmalarına giriş [Introduction to research and project work]*. Trabzon, Celepler Yayıncılık.
- Dede, Y., Doğan, M.F., & Aslan-Tutak, F. (Ed). (2020). *Matematik eğitiminde etkinlikler ve uygulamaları [Activities and applications in mathematics education]* (2nd Edition). Pegem Akademi, Ankara.
- Denzin, N. K., & Lincoln, Y. S. (Ed) (1994). *Handbook of Qualitative Research*. Thousand Oas, CA: Sage Publications.
- Doyle, W. (1983). Academic work. *Review Of Educational Research*, 53(2), 159-199. <https://doi.org/10.3102/00346543053002159>
- Dwivedi, V., & Joshi, Y. (2021). ICT perspectives of productive and sustainable development for 21st century higher education institutions. *International Journal of Education and Development using Information and Communication Technology*, 17(3), 81-96. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1335689.pdf>
- Ecemiş, Ü. (2017). *An international comparison of the cognitive demand levels of tasks in Turkey's 5th grade mathematics books (Thesis Number: 481774)* [Master Thesis, Ankara University]. Turkish Council of Higher Education Thesis Center.
- Edwards, B. (2023). OpenAI's GPT-4 exhibits "human-level performance" on professional benchmarks. Retrieved from <https://arstechnica.com/information-technology/2023/03/openai-announces-gpt-4-its-next-generation-ai-language-model/>
- Engin, Ö. (2015). *A comparison of the cognitive demand levels of tasks in 7th grade mathematics textbook in Turkey with those in the national curriculum and the textbooks of other countries (Thesis Number: 396167)* [Master Thesis, Ankara University]. Turkish Council of Higher Education Thesis Center.
- Henningsen, M., & Stein, M. (1997). Mathematical tasks and student cognition: classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, 28(5), 542-549. <https://doi.org/10.5951/jresmetheduc.28.5.0524>

- Jones, J. (2021). Integrating machine learning in mathematics classroom. *Mathematics Teacher: Learning and Teaching PK-12*, 114(8), 624-628. <https://doi.org/10.5951/MTLT.2021.0001>
- Kelly, S. M. (2022, December 5). *This AI chatbot is dominating social media with its frighteningly good essays*. Retrieved from <https://web.archive.org/web/20221210112243/https://www.cnn.com/2022/12/05/tech/chatgpt-trnd/index.html>
- Korucu, A., & Biçer, H. (2022). *Eğitimde yapay zeka uygulamaları [Artificial intelligence applications in education]* (3rd ed.) Ankara: Pegem Publishing House.
- Lugosi, E., & Uribe, G. (2022). Active learning strategies with positive effects on students' achievements in undergraduate mathematics education. *International Journal of Mathematical Education in Science and Technology*, 53(2), 403-424. <https://doi.org/10.1080/0020739X.2020.1773555>
- Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass.
- Nayir, Ö., & Bulut, S. (2022). Etkinlik kavramının matematik öğretimindeki tarihsel gelişimi [Historical development of the concept of activity in mathematics teaching]. *Eğitimde yapay zeka uygulamaları* (3rd ed.). Ankara: Pegem Yayınevi.
- OpenAI (2023b). *GPT 4 Technical Report*. Retrieved from <https://cdn.openai.com/papers/gpt-4.pdf>
- Özgen, K. (2017). Matematiksel öğrenme etkinliği türlerine yönelik kuramsal bir çalışma: Fonksiyon kavramı örnekleme [A theoretical study on types of mathematical learning activities: Function concept sampling]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 17(3), 1437-1464. <https://doi.org/10.17240/aibuefd.2017.17.31178-338839>
- Pressey, S. L. (1950). Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. *The Journal of Psychology*, 29(2), 417-447. <https://doi.org/10.1080/00223980.1950.9916043>
- Reçber, H. (2012). *An international comparison of level of cognitive demands of the tasks in Turkey's 8th grade mathematics curriculum and textbook (Thesis Number: 342456)* [Master Thesis, Ankara University]. Turkish Council of Higher Education Thesis Center.
- Sarpkaya, G. (2011). *The investigation of mathematical tasks related to primary school secondary level algebra learning field in terms of cognitive demands: Mathematics course books and classroom implementations. (Thesis Number: 298508)* [Master Thesis, Gazi University]. Turkish Council of Higher Education Thesis Center.
- Scharth, M. (2022). *The ChatGPT chatbot is blowing people away with its writing skills*. The University of Sydney. Retrieved from <https://www.sydney.edu.au/news-opinion/news/2022/12/08/the-chatgpt-chatbot-is-blowing-people-away-with-its-writing-skill.html>
- Silver, E. A. ve Herbst, P. G. (2008). Theory in mathematics education scholarship. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 39-68). Charlotte, NC: Information Age Publishing.
- Smith, M. S., & Stein, M. K. (1998). Reflections on practice: Selecting and creating mathematical tasks: From research to practice. *Mathematics teaching in the middle school*, 3(5), 344-350. <https://doi.org/10.5951/MTMS.3.5.0344>
- Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, 2(1), 50-80. <https://doi.org/10.1080/1380361960020103>

- Stein, M. K., Smith, M. S., Henningsen, M., & Silver, E. A. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development*. New York: Teacher College.
- Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study. *International Journal of Educational Technology in Higher Education*, 20(35), 1-17. <https://doi.org/10.1186/s41239-023-00404-7>
- Terrasi, V. (2023). *GPT 4: How is it different from GPT 3.5?*. Search Engine Journal. Retrieved from <https://www.searchenginejournal.com/gpt-4-vs-gpt-3-5/482463/>
- Ubuz, B., Erbaş, A., Çetinkaya, B., & Özgeldi, M. (2010). Exploring the quality of the mathematical tasks in the new Turkish elementary school mathematics curriculum guidebook: the case of algebra. *ZDM Mathematics Education*, 5(42), 483-491. <https://doi.org/10.1007/s11858-010-0258-5>
- Wakhata, R., Mutarutinya, V., & Balimuttajjo, S. (2023). Exploring the impact of Stein et al.'s levels of cognitive demand in supporting students' mathematics heuristic problem-solving abilities. *Frontiers in Education*, 8, 1-13. <https://doi.org/10.3389/educ.2023.949988>
- Yeşildere, S. (2022). Matematiksel etkinliklerin tasarım ilkeleri. Eğitimde yapay zekâ uygulamaları [Design principles of mathematical activities. Artificial intelligence applications in education] (3rd ed) Ankara: Pegem Yayınevi.
- Yıldırım, A., & Şimşek, H. (2013). *Sosyal bilimlerde nitel araştırma yöntemleri [Qualitative research methods in the social sciences]* (9th ed.). Ankara: Seçkin Publishing House.
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, California: Sage.
- Yuan, Y. (2022). Quantitative analysis of Chinese classroom teaching activity under the background of artificial intelligence. *Education and Information Technologies*, 27(8), 11161-11177. <https://doi.org/10.1007/s10639-022-11080-x>

Article Information Form

Authors Notes: The authors would like to express their sincere thanks to the editor and the anonymous reviewers for their helpful comments and suggestions.

Authors Contributions: All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Conflict of Interest Disclosure: No potential conflict of interest was declared by the author.

Copyright Statement: Authors own the copyright of their work published in the journal and their work is published under the CC BY-NC 4.0 license.

Supporting/Supporting Organizations: No grants were received from any public, private or non-profit organizations for this research.

Ethical Approval and Participant Consent: It is declared that during the preparation process of this study, scientific and ethical principles were followed and all the studies benefited from are stated in the bibliography.

Plagiarism Statement: This article has been scanned by iThenticate.