UDC: 37.091:004.738:53.08

Mehmet Akif Arduç
Dr., Science Teacher
Ministry of National Education, Science Education, Adıyaman, Türkiye
ORCID ID 0000-0002-1026-2843
arducakif@gmail.com

Gülşah Gürkan
Ph.D., Assistant Professor
Malatya Turgut Ozal University, Science Education, Malatya, Türkiye
ORCID ID 0000-0003-0297-3060
gulsah.gurkan@ozal.edu.tr

BRIDGING THE GAP: HOW CHATGPT ENHANCES ASSESSMENT TO CLOSE LEARNING GAPS IN MIDDLE SCHOOL

Abstract. This study investigates the impact of ChatGPT-supported assessments on addressing students' learning gaps. With the increasing integration of artificial intelligence (AI) in education, understanding its potential benefits in enhancing students' academic achievement is of great importance. For this purpose, a quasi-experimental design was implemented within a quantitative methodology framework. The study involved seventh-grade students from a public middle school in Turkey during the 2024-2025 academic year. To assess students' learning levels, seventh-grade assessment questions from the "The Solar System and Beyond" unit, published by the Ministry of National Education (MEB), were used. These multiple-choice questions consisted of four options, with one correct answer, and underwent validity and reliability testing. Prior to publication, the Ministry of National Education (MoNE) sought expert opinions and conducted pilot implementations to analyze item performance. All questions were designed to be appropriate for students' developmental levels and were explicitly aligned with the "Solar System and Beyond" unit.The experimental group participated in a ChatGPT-supported question-and-answer process during the "The Solar System and Beyond" unit. Instead of providing direct answers, students received step-by-step guidance toward the correct solution and immediate feedback. The results showed that ChatGPT-supported assessments effectively addressed learning gaps, and the experimental group demonstrated statistically significant improvement in learning outcomes. This study provides valuable insights into the potential of ChatGPT-supported chatbots to reduce learning gaps in educational settings and suggests that they can contribute to improving students' academic performance. Additionally, the study offers new empirical evidence that AI-supported chatbots can actively facilitate learning by supporting students' reasoning processes rather than merely providing correct answers, thus presenting a novel approach to AI-supported formative assessment in education.

Keywords: ChatGPT (GPT-4); assessment and evaluation; middle school students; learning gaps; solar system and beyond.

1. INTRODUCTION

Artificial intelligence refers to systems that perform certain tasks algorithmically by simulating the components of human intelligence [1]. In recent years, artificial intelligence, which has become part of countries' educational visions, has frequently progressed with various applications for developing educational content and intelligent teaching methods, assessing students and improving teacher-student interaction, providing course activities and content presentation, providing feedback to students, and monitoring progress. In this context, the inclusion of artificial intelligence technology in education offers new opportunities that facilitate student learning [2]. Artificial intelligence, one of the advanced learning technologies, appears as chatbots, expert systems and intelligent teaching systems. ChatGPT, one of the large language models developed by OpenAI, and Gemini, developed by Google, are among the

© Mehmet Akif Arduç, Gülşah Gürkan, 2025.

widely used artificial intelligence chatbots today. Unlike other artificial intelligence-based chatbots, ChatGPT, one of the large language models developed by OpenAI, is more successful in activities that require creativity and is specifically designed to process and understand human language [3].

The use of generative AI tools such as ChatGPT in a purposeful manner can help with different aspects of academic research and publication, such as brainstorming, language correction, and formatting. AI-based chatbots such as ChatGPT can be used to interact with students and support their argumentation skills. Chatbots, equipped with reasoning and knowledge, can be used as a tool for learning approaches, performing sampling much faster than humans [4]. ChatGPT, which can process various text-based requests, has the ability to respond to requests in a conversational style by processing basic queries and answers. These advantages of ChatGPT can be used to eliminate learning deficiencies and ensure complete learning in education and training activities. As artificial intelligence tools develop and the use of these tools by students increases, educators must ensure that learning outcomes are compatible with these evolving dynamics. Theoretical frameworks such as Bloom's Complete Learning Model, active learning, and constructivist learning theories can serve as a guide by adapting learning objectives to take into account the presence of artificial intelligence [5].

Bloom's Mastery Learning Model argues that if an appropriate and planned approach is taken, almost all students can learn all the new behaviours that are intended to be taught, and academic success will increase [8]. In the Mastery Learning Model, where the process can be adjusted according to the student level, the learning products are evaluated throughout the process and the student's learning difficulties and deficiencies are resolved in a timely manner. Since the next unit is not passed on without resolving the learning deficiencies of the students, usually all students reach the desired level of teaching objectives. This situation contributes to the reduction of individual differences in terms of academic success and the creation of a positive effect on the student's success level [9]. One of the most important features of the Complete Learning Model is that each student can progress at their own pace. Generative artificial intelligence language tools such as ChatGPT can provide personalized content according to each student's learning needs. In this context, students can get more information on the subjects they are deficient in through ChatGPT and their questions can be answered instantly. The instant feedback given to student questions with ChatGPT can be used as a tool to provide complete learning in subjects in which students are deficient and to eliminate these deficiencies quickly. In this context, with the instant feedback given to the student during the assessment process, progress is easier and learning deficiencies can be gradually eliminated, and students' cognitive load can be reduced [6].

In addition to the benefits listed above, there are also concerns about some negative effects of ChatGPT. If students and teachers rely too much on AI tools for content generation, the possibility of skipping the actual exchange of information poses a real ethical and practical problem. Recently, Elsevier, one of the leading research publishers, has adopted a publication policy that encourages disclosure of the use of AI technologies to promote research transparency and trust. In particular, discussions have begun on whether texts produced by ChatGPT constitute plagiarism [7]. However, with students having quick and easy access to ready-made information, students' skills such as creative thinking, problem solving, and critical thinking may be negatively affected [8]. It is debated that ChatGPT, a model designed to understand and generate natural language, should be regulated to prevent cheating. Therefore, the need to consider the benefits and concerns created by artificial intelligence tools such as ChatGPT arises, as the emergence and use of these tools is inevitable. At this point, it is the responsibility of teachers to guide students on the appropriate and effective use of such artificial intelligence tools, when deemed necessary [9]. When looking at the benefits and concerns of

ChatGPT, it is seen that the effects of these artificial intelligence tools for the education sector and society as a whole should be critically examined.

The problem statement. Learning deficiencies can be addressed by providing students with AI-supported feedback similar to the feedback teachers provide when taking tests at home or in class. In this context, students receiving direct feedback on the accuracy of their answers during the assessment process allows them to develop their skills to address questions and problems more quickly. In this way, AI-supported feedback activities can allow teachers and students to save time. Teachers try to focus on more complex teaching aspects by saving time they previously spent on checking and correcting repeated errors [10]. The process feedback of chatbots can become more similar to the feedback provided by teachers. The current study will provide valuable insights into the potential of ChatGPT-powered chatbots to address learning gaps, increase academic achievement, and reduce educational inequalities. In this context, the current study is structured under the headings of "Related Studies", which provides a brief discussion of existing academic studies on ChatGPT in order to understand the direction of the literature; "Methodology", which explains the applied methodology; "Conclusion", which includes the results of the effect of instant feedback given to students by ChatGPT during the assessment process in eliminating learning deficiencies; "Discussion", which provides a discussion of the findings of ChatGPT-supported feedback on the experimental and control groups; "Recommendations and Implications", which offers some suggestions for the application of ChatGPT in the assessment process; and finally, "Limitations", which presents the conclusion, implications and limitations of the study.

Analysis of recent studies and publications. Recent advancements in AI have enabled ChatGPT to serve as an automated writing assessment tool, allowing students to receive more frequent feedback while alleviating the workload for teachers who are managing large classes [11]. Instructional feedback is an effective approach to improving student learning. However, instructional feedback can be difficult, laborious, and costly to implement in practice. This has led to the use of pre-trained language models as inspiration. Automatic feedback that teachers and students receive from online learning environments can positively affect students' engagement in the learning process and improve their performance [12]. Studies on the effectiveness of AI-supported assessments and evaluations are still quite limited; however, some studies on the effectiveness of ChatGPT in the assessment-evaluation process are given below. In the study conducted by [13], students were helped to learn better by providing ChatGPT-supported feedback. The study found that ChatGPT was capable of producing more detailed feedback and summarizing students' performance more fluently and consistently than human instructors. However, it was concluded that ChatGPT was highly compatible with the teacher when evaluating the subject matter of students' assignments and could benefit students in improving their learning skills.

In a study conducted by [11], the quality of formative feedback provided by ChatGPT was compared to that of human feedback. The study evaluated 200 instances of human-generated feedback and 200 instances of AI-generated feedback, assessing how well the feedback adhered to criteria, offered clear improvement instructions, maintained accuracy, prioritized key aspects, and employed a supportive tone. As a result, it was concluded that while well-trained raters provided higher quality feedback than ChatGPT, generative AI could be particularly useful in situations where a well-trained rater is not available.

[14] carried out a study to determine the accuracy of ChatGPT's argumentation feedback and the factors affecting the evaluation of arguments, and found that ChatGPT showed an impressive precision rate (91.8%) and recall rate (63.2%) when providing feedback on arguments. This study, which qualitatively compared ChatGPT and teacher feedback, revealed that both types of feedback have their advantages and disadvantages.

[15], aimed to explore students' experiences with ChatGPT when seeking help in generating internal feedback and to determine the factors affecting the quality of feedback provided by ChatGPT. As a result of the qualitative study, it was concluded that including ChatGPT as a feedback tool in process-oriented writing education helped students develop their skills and be successful in writing.

[16], aimed to examine the effect of ChatGPT on the academic performance and learning perception of 11th grade students in the field of electromagnetism. One group of participants accessed ChatGPT (experimental group) while the other group did not access ChatGPT (control group). During the study, the experimental group was instructed to use ChatGPT to get help with course content, and both quantitative and qualitative data were collected to evaluate the effect of ChatGPT on student learning and perception. As a result of the study, it was seen that ChatGPT had a positive effect on students' academic success and learning perception.

Although studies on technology-focused redesign of AI-supported assessment and evaluation processes have increased over time, empirical evidence on the effectiveness of using these tools in the assessment process is quite limited. In particular, in terms of learning deficiencies and misconceptions in the assessment process, there is a need for studies on redesigning chatbots such as ChatGPT to provide feedback to students, as well as provide hints, reinforcement and motivation. In order to fill this gap, the current study aimed to investigate the effect of ChatGPT on eliminating learning deficiencies by using the feature of guiding students to the correct answer with hints while administering an achievement test to them.

The research goal. Artificial intelligence language model tools are changing education in terms of teaching, learning, assessment and management. With its advanced natural language processing capability, ChatGPT enables a transition from traditional, linear learning assessment models to more dynamic and personalized approaches by imitating human-like text production [36]. With this transition, learning outcomes can be improved by adapting educational experiences to individual needs and providing instant feedback. The most powerful element of the quality of the teaching service is feedback and correction [17]. Students who expect feedback are generally more successful than others. However, crowded classes cause feedback-correction activities to be carried out only occasionally. Providing timely feedback to students is considered a characteristic of good feedback. Therefore, it is necessary to provide quality and sufficient feedback in the learning process. Teachers often strive to deliver timely feedback during lessons but face numerous contextual constraints, whereas ChatGPT can offer feedback instantly and iteratively. Therefore, it is assumed that ChatGPT's feedback is always more immediate. The adoption of ChatGPT and similar artificial intelligence technologies in student grading and assessment is a harbinger of important and broader implications for the field of education [18]. In this context, the aim of the study was to investigate the effect of the assessment process carried out with instant feedback, given according to the cognitive development levels of students, and supported by artificial intelligence using ChatGPT, in eliminating the learning deficiencies of middle school students. Within the scope of this purpose, the following research questions were sought:

1. Is there a statistically significant difference between the pre-tests of the experimental and control groups?
2. Is there a statistically significant difference between the post-tests of the experimental and control groups?
3. Is there a statistically significant difference between the pre-tests and post-tests of the experimental group?
4. Is there a statistically significant difference between the pre-tests and post-tests of the control group?

2. RESEARCH METHODS

2.1. Research Model

Since the aim of the study was to examine the effect of the ChatGPT-supported assessment process on completing students' learning deficiencies and the study group could not be selected by random sampling, a quasi-experimental design was preferred among quantitative method designs. A quasi-experimental design is used in cases where it is not possible to form groups by random assignment and where a specific intervention is applied to one of the groups differently from the other group [19]. There are experimental and control groups in the study. A subject-based achievement test was applied to the groups after they completed the "Solar System and Beyond" unit. While ChatGPT support was used during the achievement test application phase in the experimental group, an intervention-free application was applied in the control group. This achievement test was also accepted as a pre-test in order to compare the pre-intervention levels of the groups. Once the assessment process was completed, a parallel test was administered as a post-test to assess the impact of the intervention. The research model is shown in Table 1.

Table 1

Research design

Group	Pre-Processing	Experimental Processing	Post-Processing
Experimental Group	Processing of the Solar System and Beyond Unit (16 hours)	Implementation of the Achievement Test supported by ChatGPT (also accepted as a pre-test)	Implementation of the final test
Control Group	Processing of the Solar System and Beyond Unit (16 hours)	Implementation of the Achievement Test (no intervention, also considered as a pre-test)	Implementation of the final test

2.2. Study Group

The study was conducted in a province located in the Southeastern Anatolia Region of Turkey at the beginning of the 2024-2025 academic year. The study was conducted by researchers in a state middle school. In this direction, the easily accessible convenience sampling method, one of the purposeful sampling methods, was used when determining the study group. This sampling method saves researchers money, time and effort in line with the purpose of the study in cases where the use of the random sampling method is not possible. Since there were two seventh grades in the school where the study was conducted, the seventh grades were determined as the study group and were randomly assigned as the experimental and control groups. Since the assignment of students to the classes was made randomly by the school administration, the academic and demographic characteristics of the two class groups were similar. Both groups consist of 20 students. This sample size is considered sufficient in the literature for a large effect size (e.g., Cohen's d>0.8), [20]. The experimental group consists of 13 female and 7 male students, while the control group includes 12 female and 8 male students. The students are between the ages of 12 and 13. The "Solar System and Beyond Unit" was used in the study because it contains information that can be memorized and includes topics that students have difficulty learning. During the research process, 5 students who did not participate in any lesson or assessment process were not included in the study. All the remaining students were included in the process because they volunteered.

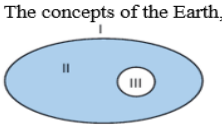
2.3. Data Collection Tools

To align with the study objective, seventh-grade-level tests published by the Ministry of National Education (MoNE) were utilized to assess students' learning levels after completing the Solar System and Beyond Unit. The questions consist of multiple-choice items with four options and one correct answer, and validity and reliability studies have been conducted. MoNE obtains expert opinions before publishing these questions and analyzes the data by conducting pilot applications. The questions were prepared by MoNE in accordance with the student level and published for the Solar System and Beyond Unit. Although there are achievement tests related to the subject in the literature, they were not used because they did not fully comply with the purpose of the research or because the number of questions was too high. The tests used as pre-test and post-test were different parallel tests. The two tests were examined by science education experts and four questions that were found to have a level difference were changed. The pre-test was applied as an achievement test after the unit was taught for 16 hours and the intervention process was carried out at this stage. The achievement test was applied to both groups after the intervention.

2.4. Implementation Process

First, a pilot study was conducted with three students. In order to create the assessment process to be carried out, ChatGPT was introduced by giving the relevant unit of the textbook, curriculum and questions (*I will give you the content of the Solar System and Beyond Unit from the Science textbook and science curriculum and share questions about the unit, I want you to give sentences that will guide you without directly giving the correct answer for each wrong answer in these questions*). The 20-question Solar System and Beyond Unit questions published by the MoNE were shared with ChatGPT in order. An example of a chat made with ChatGPT is given below:

9. The concepts of the Earth, universe and space are modeled according to their sizes as follows.



Accordingly, which of the following correctly lists the concepts that are appropriate to be written in the spaces numbered I, II and III?

	I	II	III
A)	Universe	Earth	Space
B)	Earth	Space	Universe
C)	Space	Universe	Earth
D)	Universe	Space	Earth

— ***Share for the student who answered A***
— ***Think about it:*** *The universe is the largest structure and encompasses everything, this is true. However, is the location given for the size of the Earth, correct? Is the Earth part of the universe or space? What can we think of the Earth in?*
— ***Share for the student who answered B***
— ***Let's think again:*** *The Earth should be inside as the smallest structure. However, reconsider whether the Earth is part of the universe or space. How should we position the universe as the most comprehensive structure? Try again by thinking about the relationship between the Earth and space.*
— ***Share for the student who answered C***
— ***Let's think a little more:*** *The universe, as the largest structure, encompasses everything. But here you should remember that the universe should be the outermost. Is the Earth inside the universe or is the universe inside the Earth? You can reconsider your answer by thinking about the size of space relative to the universe.*
— ***Share for student who answered D***
— *Congratulations, you found the correct answer! The universe is the largest structure that encompasses everything, and space is a part of the universe. The Earth is located in space. In other words, the largest structure is the universe, space is located in the universe, and the Earth is a smaller area that is part of space. You got this order right; you did a great job!*

Figure 1. Sample question process used in the pilot group

After the pilot process was completed, various applications (Quizlet, Kahoot, Socrative, ClassMarker, Google Forms, etc.) were used to place the obtained explanations appropriately. With the help of the applications used, the explanation produced by ChatGPT was added to the option the student answered for each question. The process continued with three people, and after this pilot application process, it was decided to carry out the entire process via ChatGPT after seeing some deficiencies of each program. In line with the answers obtained from this process, the achievement test process to be used in the experimental group was structured.

In the experimental group, when the process started, students were asked to write the texts given via ChatGPT. After the process was introduced to ChatGPT, the question answering process was started. A sample correspondence for a sample question is presented in Figure 2.

— I will give you a test with 20 questions. The questions will appear in front of me in order. I will write the answer options for the questions you give. Don't give me the correct answer, just guide me to reach the correct answer!

— I understand, then I expect you to share the questions with me. When you send me the questions, I will guide you according to your answers.

— The pdf file with 20 questions has been shared.

*— **Question 2:** Viking 1, sent on August 20, 1975 to examine the surface structure of Mars and to investigate whether there is life there, is the first spacecraft to successfully land on Mars and conduct its examinations. Accordingly, Viking 1 can be given as an example of which of the following technological tools produced for space research?*

a) Space shuttle
b) Space station
c) Space probe
d) Space telescope

Write your answer and I will guide you to the correct answer.

— D

— The answer you gave is space telescope, but since Viking 1's mission is to land on the surface of Mars and conduct studies, it is not used to observe space like space telescopes. Think about it a little more. Considering that Viking 1 landed on the surface of Mars and roamed around there collecting scientific data, which would be a more suitable option?

— C

— Great, you're on the right track! Viking 1 is, in fact, a space probe. Space probes are used to land on planets and conduct research there, just like Viking 1 did on Mars. Congratulations!

— Shall we continue?

Figure 2. Sample question process used in the experimental group

The application was carried out by one of the researchers of this study. The researcher is a person who is aware of and has used artificial intelligence tools such as ChatGPT and has a PhD in science education. Students were asked to answer questions via mobile devices as shown in the example process in Figure 2. Students in the experimental group received instant feedback while answering each question and were presented with directions created by ChatGPT according to the options they indicated as answers. Correct answers were not given directly to the students, but they were guided until they found the correct answer according to the options they indicated. However, in order to compare the achievement test results (as a pre-test) between the experimental and control groups, students in the experimental group were asked to write their first answers on an optical form before writing them on ChatGPT. Considering the difficulty of this situation, three additional observers were present in addition to the researcher in the class with three groups of desks. Students were told that they would be given two minutes for each question and that the application would be completed within one class hour, and it was observed that no student had finished the process by the 35th minute. At the end of the process, the optics and ChatGPT texts were compared to determine which option the first answers were. Since a difference was found in two questions in one of the five forms examined, the results of all students were compared and their first answers to the achievement

test were recorded correctly. In addition, support was received from a science education expert to examine the information of four randomly selected students from the group, and no deficiencies were detected.

In the control group, mobile devices were used again and only the correct answer option was presented by indicating whether the answers given by the students were “correct” or “incorrect”. The application was carried out with the Kahoot Web 2.0 tool. It was recorded that all students completed the test in the 32nd minute of the study. Since this application provided the total correct answers, an optical form was not needed again. A test that was completely parallel to the achievement test was used as the final test. The questions were presented to the students in written form and they were asked to answer the questions. The process was continued and completed as specified in the experimental and control groups. The groups were gathered in the same classroom environment with three observers. It was observed that all students completed the test in the 30th minute of the process. It was confirmed that 20 students from each group participated in the study and their data were organized.

2.5. Data Analysis

The study included experimental and control groups, with two different achievement tests administered to students as pre-tests and post-tests. The study aimed to identify differences both between the groups and within each group. Based on the obtained data, it was examined whether there was a statistically significant difference between the pre-tests and post-tests of the experimental group and the control group. It was also examined whether there was a statistically significant difference between the pre-tests and post-tests within the experimental and control groups. In order to decide which statistical tests to use for this purpose, the data obtained for both groups were entered into the SPSS program and the tests were decided upon by performing analyses.

In order to determine the difference between the experimental and control groups (separately for the pretest and posttest), when the data of both groups were analyzed for the pretest and posttest; as seen in Table 2, the skewness and kurtosis values were between -1.5 and +1.5, and the Shapiro-Wilks (p) value was found to be greater than .05 These values indicate that the data followed a normal distribution [21]. Considering the Levene test result, it was determined that the homogeneity of the variances of the groups (p> .05) was met, and considering the fact that the test used was an equal-interval test with a normal distribution, the data **were** analyzed with the unrelated samples t-test. However, in the analysis of the posttest, the ANCOVA test was used instead of the unrelated samples t-test in order to minimize the margin of error and eliminate the effect of the pretest.

To determine which test to use for assessing a statistically significant difference between the pre-test and post-test within the experimental and control groups, the difference scores were first calculated, and their distribution was analyzed for normality. It can be seen in Table 1 that the skewness and kurtosis values for both groups were between -1.5 and +1.5, and the Shapiro-Wilks (p) value was determined to be greater than .05. These values indicate that the data followed a normal distribution [21]. In the light of the analyses obtained, it was seen that a related sample t-test should be performed. The skewness and kurtosis, Shapiro-Wilks Test and Levene Test values are presented in Table 2.

Table 2

Skewness-kurtosis, Shapiro-Wilks test and Levene Test values

Variables	Groups	Skewness	Kurtosis	Shapiro-Wilks (p)	Levene Test (F)	Levene Test (p)
Pre-test	Experimental Group	.505	.324	.609	1.054	.311

	Control Group	-.658	-.275	.188		
Post-test	Experimental Group	-.089	-.516	.388	.384	.539
	Control Group	.365	-.296	.837		
Difference Scores	Experimental Group	.247	.280	.299		
	Control Group	-.213	.637	.983		

3. THE RESULTS AND DISCUSSION

The independent sample t-test was used to examine whether there was a statistically significant difference between the pre-test scores of the experimental group and the control group, as it met the assumptions. The results are presented in Table 3 and Table 4.

3.1. Results regarding the first sub-problem

Table 3

ANCOVA results for group × covariate interaction

Source	Sum of Squares	df	Mean of Squares	F	p
Group × Covariant	2.559	1	2.559	.113	.739
Reg.	12817.840	1	12817.840	563.587	.000
Group	.548	1	.548	.024	.878
Error	818.760	36	22.743		
Total	247850.000	40			

Table 4

Pre-Tests independent samples t-test results

Variable	Groups	*N*	$\bar{X}$	S	sd	*t*	p
Pre-test	Experimental Group	20	73.25	17.11	38	.046	.964
	Control Group	20	73.00	17.50			

Table 4 compares the pre-test scores of the experimental and control group students. It was seen that the mean scores of the groups are quite close to each other ($\bar{X}$ experimental group = 73.25, $\bar{X}$ control group = 73.00). It was seen that there is no statistically significant difference between the mean scores of the experimental and control groups [t(38) = .046, p = .964]. p> .05 indicates that the difference is not statistically significant.

3.2. Results regarding the second sub-problem

ANCOVA test was used to examine whether there was a statistically significant difference between the post-test scores of the experimental group and the control group, as it met the assumptions. The results are presented in Table 5.

Table 5

ANCOVA results of post-test scores by group

Source of Variance	Sum of Squares	sd	Mean of Squares	F	p	η^2
Reg.	12816.181	1	12816.181	577.362	.000*	
Group	104.629	1	104.629	4.714	.036*	.01
Error	822.15	37	22.21			
Total	13810.38	39				

**p<.05*

When the ANCOVA test results in Table 5 were examined, it was seen that the adjusted achievement test scores are statistically significantly different from each other. F(1,37)=4.714, p< .05. When the descriptive statistics in Tables 6 and 7 are examined, it was seen that the experimental group post-test mean scores ($\bar{X}$=78.25) were higher than the control group post-test mean scores ($\bar{X}$ =74.75). The η^2 value calculated in Table 5 as .01 shows that the effect is small.

3.3. Results regarding the third sub-problem

Whether there was a statistically significant difference between the pre-test and post-test scores of the experimental group was examined using the paired sample t-test, as it met the assumptions. The results are presented in Table 6.

Table 6

Experimental group paired sample t-test results

Group	Variable	*N*	$\bar{X}$	S	sd	*t*	p	Cohen's d
Experimental Group	Pre-test	20	73.25	17.11	19	-4.359	.000*	.275
	Post-test	20	78.25	19.07				

**p<.01*

Table 6 compares the pre-test and post-test scores of the experimental group students. It is seen that the experimental group post-test mean scores ($\bar{X}$=78.25) are higher than the achievement test mean scores ($\bar{X}$=73.25). When the difference is examined as significant, it is seen that t(19) = -4.359, p = .000, p<.05 shows that the difference is statistically significant. The calculated Cohen's d value is .275, which shows that the effect is small.

3.4. Results regarding the fourtf sub-problem

The dependent sample t-test was applied to determine if there was a statistically significant difference between the pre-test and post-test scores of the control group, as the assumptions were met. The results are shown in Table 7.

Table 7

Control group paired sample t-test results

Group	Variable	*N*	$\bar{X}$	S	sd	*t*	p
Control Group	Pre-test	20	73.00	17.50	19	-1.789	.090
	Post-test	20	74.75	18.81			

**p<.05*

Table 7 compares the pre-test and post-test scores of the control group students. It is seen that the experimental group post-test mean scores ($\bar{X}$=74.75) are higher than the pre-test mean scores ($\bar{X}$ =73.00). When we look at whether the difference is significant, it is seen that t(19) = -1.789, p = .090, p > .05 shows that the difference is not statistically significant.

In this study, the effect of ChatGPT supported assessment process in eliminating the learning deficiencies of students for the science course "Solar System and Beyond" unit was investigated. In this section of the study, the discussion of the obtained findings with the studies in the literature, and the results and recommendations of the research are included. The results regarding the first sub-problem of the study showed that there was no statistically significant difference between the achievement tests applied as pre-tests to the experimental and control groups. This indicates that the experimental and control groups were at similar levels in terms of learning outcomes. These results allow for a more reliable evaluation of the effect of the application and show that the final differences resulted from the application. In addition, this increases the internal validity of the study [22].

When the results regarding the second sub-problem of the study were examined, a statistically significant difference was determined in favor of the experimental group among the students' post-tests. This suggests that the ChatGPT-supported assessment process is effective in eliminating students' learning deficiencies. The fact that students did not directly get the correct answer while answering the questions and instead encountered feedback for each wrong answer creates the need to spend more time on the question and makes it difficult for students to forget, thus ensuring permanent learning. [16], concluded that the academic success and learning perceptions of students in the group using ChatGPT were higher than those in the group not using ChatGPT, as a result of their study aimed at examining the effect of ChatGPT on the academic performance and learning perceptions of 11th-grade students in the field of electromagnetism. [17] concluded in their study on feedback that feedback has an effect on the permanence of learning and success. The feedback and guidance provided by ChatGPT enable students to learn the reasons for their wrong answers and give them the opportunity to analyse them.

[23], stated that providing artificial intelligence-supported feedback increases students' participation in the course and strengthens their learning outcomes in their meta-analysis study. Similarly, [24] stated that providing artificial intelligence-supported feedback increases student motivation and success. Their study shows that it improves students' problem-solving skills. Based on this, it can be said that ChatGPT-supported assessment process provides students with an opportunity for in-depth learning and allows them to analyse their thinking process.

In the light of results obtained, it is seen that the average scores of the control group students increased less and this increase was not significant. In the control group, students were only given the option of correct answer by being told whether their answers were correct or wrong in the assessment process. This did not allow students to understand why they were wrong and the total time spent with the question was less. Therefore, this result is an expected outcome. [25] found that feedback has a limited effect on the learning process when used only to provide correct answers. This result suggests that providing guidance while providing feedback will help with permanent learning.

Current literature shows that the use of artificial intelligence support in teaching is effective. [26] stated that the use of artificial intelligence support makes an important contribution because it can respond to the individual needs of students. When this situation is considered specifically for this study, it is important to provide guidance to each student according to the answer they give in a student profile that gives four different answers for a question with four options in order to eliminate individual deficiencies. While a student can reach the correct answer with one feedback text, some students can reach the correct answer after two or three feedbacks. With the application made in this study, each student was able to

access additional information as much as they needed according to their own knowledge level and provides more personalized learning for the learning to be permanent. In addition, it should be known that technology alone cannot positively affect learning processes and the pedagogical infrastructures of the innovations encountered should also be ensured. It should not be forgotten that the use of artificial intelligence tools such as ChatGPT in teaching has various disadvantages [27].

When the relevant literature is examined, it is evident that lesson plans and programs have been prepared using artificial intelligence tools such as ChatGPT, support has been provided through these tools during the lesson process, in student evaluations, and in language learning processes, and they have been used effectively. [28], [29], [30]. When we look at the published studies, it is seen that there is no study in which the answer was tried to be found by guiding the students without giving the answer directly in the assessment process as in this study. When looking at the literature, it is seen that although different techniques are used in assessment studies, artificial intelligence tools such as ChatGPT are not yet used. The fact that this technology is relatively new can be shown as one of the reasons for this situation. The results obtained with the research conducted in this study will contribute to the use of alternative methods to be used in the assessment process.

5. CONCLUSIONS AND PROSPECTS FOR FURTHER RESEARCH

This study has shown that ChatGPT-supported assessment process is effective in eliminating students' learning deficiencies for the science course "Solar System and Beyond" unit. As a result, the use of artificial intelligence tools such as ChatGPT, which provide a personalized learning opportunity, in the assessment process allows students to make their learning permanent. The results of this study, which are also consistent with the studies in the literature, show that the use of artificial intelligence tools such as ChatGPT in the assessment process yields effective results. Upon examining the findings related to the third and fourth sub-problems of the study, a statistically significant difference was found between the pre-test and post-test scores of the experimental group, whereas no such difference was observed in the control group. Although the students solved similar questions to the pre-test in the post-test, the fact that no significant difference was observed in the control group shows how important the assessment process is. The fact that ChatGPT provided guidance until the students in the experimental group got the correct answer for the questions they answered incorrectly, offered individual feedback, and exposed the students to the question more are among the reasons for this result.

This study was built on providing feedback and guidance to students without directly giving them the correct answer, and effective results were achieved. In this context, it is recommended that teachers use sufficient time for the assessment process and make the assessment process a more effective component of learning. In order to be able to carry out similar studies professionally, it is recommended that teachers participate in activities that will improve themselves in the area. It is recommended that program makers provide sample applications in line with current technologies to teachers and that they provide guidance in a structure that will consider which information is missing for students who mark each option, as used in this study, during the creation of published resources. It is recommended that the infrastructure of mobile devices and internet access, which can be accessed by schools and classes in case of need, be established in all schools. It is recommended that the effect of this assessment process should be investigated on a large scale by using it in different subjects and branches.

The study was limited to a specific group of students at a specific school with groups limited to 20 people, mobile devices and internet access being needed in the classroom

environment to be able to do the application, difficulties in student coordination, and problems related to teaching the use of the application. The study was limited to the "Solar System and Beyond Unit" and to a specific period of time. The study was limited to students' different technology use proficiencies and ChatGPT use proficiencies. The study also limited the use of a multiple-choice test as an achievement test.

FUNDING

No funding was received for conducting this study.

ACKNOWLEDGEMENT

This work was presented in the form of an oral abstract by G. Gülşah and M. A. Arduç at the 2024 Symposium on Measurement, Selection, and Placement, Ankara, Türkiye.

REFERENCES (TRANSLATED AND TRANSLITERATED)

[1] N. P. Melville, L. Robert, and X. Xiao, "Putting humans back in the loop: An affordance conceptualization of the 4th industrial revolution," *Information Systems Journal*, vol. 33, no. 4, pp. 733–757, 2023. [Online]. Available: https://doi.org/10.1111/isj.12422 (in English).

[2] K. D. Caratiquit and L. J. C. Caratiquit, "ChatGPT as an academic support tool on the academic performance among students: The mediating role of learning motivation," *Journal of Social, Humanity, and Education*, vol. 4, no. 1, pp. 21–33, 2023. [Online]. Available: https://doi.org/10.35912/jshe.v4i1.1558 (in English).

[3] S. Wolfram, *What is ChatGPT doing ... and why does it work?*, Wolfram Media Inc., 2023. [Online]. Available: https://doi.org/10.31855/bc47ee6b-75c (in English).

[4] Y. Su, Y. Lin, and C. Lai, "Collaborating with ChatGPT in argumentative writing classrooms," *Assessing Writing*, vol. 57, p. 100752, 2023. [Online]. Available: https://doi.org/10.1016/j.asw.2023.100752 (in English).

[5] R. H. Mogavi et al., "Exploring user perspectives on ChatGPT: Applications, perceptions, and implications for AI-integrated education," 2023. [Online]. Available: https://doi.org/10.48550/ARXIV.2305.13114 (in English).

[6] J. A. Flad, *The effects on increasing cognitive load on self-report and dual task measures of mental effort during problem solving*, Doctoral dissertation, University of Southern California, 2002. (in English).

[7] U. A. Bukar et al., "Text analysis on early reactions to ChatGPT as a tool for academic progress or exploitation," *SN Computer Science*, vol. 5, p. 366, 2024. [Online]. Available: https://doi.org/10.1007/s42979-024-02714-7 (in English).

[8] M. Sullivan, A. Kelly, and P. McLaughlan, "ChatGPT in higher education: Considerations for academic integrity and student learning," *Journal of Applied Learning & Teaching*, vol. 6, no. 1, pp. 1–10, 2023. [Online]. Available: http://dx.doi.org/10.37074/jalt.2023.6.1.17 (in English).

[9] J. Willems, "ChatGPT at universities - the least of our concerns," 2023. [Online]. Available: https://doi.org/10.2139/ssrn.4334162 (in English).

[10] J. Skalka, M. Drlík, and J. Obonya, "Automated assessment in learning and teaching programming languages using virtual learning environment," in *2019 IEEE Global Engineering Education Conference (EDUCON)*, 2019. [Online]. Available: https://doi.org/10.1109/EDUCON.2019.8725127 (in English).

[11] J. Steiss et al., "Comparing the quality of human and ChatGPT feedback of students' writing," *Learning and Instruction*, vol. 91, 2024. [Online]. Available: https://doi.org/10.1016/j.learninstruc.2024.101790 (in English).

[12] A. P. Cavalcanti et al., "Automatic feedback in online learning environments: A systematic literature review," *Computers and Education: Artificial Intelligence*, vol. 2, p. 100027, 2021. [Online]. Available: https://doi.org/10.1016/j.caeai.2021.100027 (in English).

[13] W. Dai et al., "Can large language models provide feedback to students? A case study on ChatGPT," in *23rd IEEE International Conference on Advanced Learning Technologies (ICALT)*, 2023. [Online]. Available: https://doi.org/10.1109/ICALT58122.2023.00100 (in English).

[14] L. Wang et al., "ChatGPT's capabilities in providing feedback on undergraduate students' argumentation: A case study," *Thinking Skills and Creativity*, vol. 51, p. 101440, 2024. [Online]. Available: https://doi.org/10.1016/j.tsc.2023.101440 (in English).
[15] A. C. F. Tam, "Interacting with ChatGPT for internal feedback and factors affecting feedback quality," *Assessment and Evaluation in Higher Education*, pp. 1–17, 2024. [Online]. Available: https://doi.org/10.1080/02602938.2024.2374485 (in English).
[16] S. Alneyadi and Y. Wardat, "ChatGPT: Revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools," *Contemporary Educational Technology*, vol. 15, no. 4, p. 448, 2023. [Online]. Available: https://doi.org/10.30935/cedtech/13417 (in English).
[17] J. Hattie and H. Timperley, "The power of feedback," *Review of Educational Research*, vol. 77, no. 1, pp. 81–112, 2007. [Online]. Available: https://doi.org/10.3102/003465430298487 (in English).
[18] X. Yang, Q. Wang, and J. Lyu, "Assessing ChatGPT's educational capabilities and application potential," *ECNU Review of Education*, 2023. [Online]. Available: https://doi.org/10.1177/20965311231210006 (in English).
[19] J. W. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*, 4th ed., Sage Publications, 2014. (in English).
[20] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed., Erlbaum, 1988. (in English).
[21] B. G. Tabachnick and L. S. Fidell, *Using multivariate statistics*, 7th ed., Pearson, 2019. (in English).
[22] J. R. Fraenkel, N. E. Wallen, and H. H. Hyun, *How to design and evaluate research in education*, 8th ed., McGraw-Hill, 2012. (in English).
[23] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?" *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, p. 39, 2019. [Online]. Available: https://doi.org/10.1186/s41239-019-0171-0 (in English).
[24] S. Knight, A. F. Wise, and B. Chen, "The impact of using artificial intelligence in classrooms on student learning outcomes," *Journal of Educational Computing Research*, vol. 58, no. 4, pp. 780–804, 2020. (in English).
[25] V. J. Shute, E. G. Hansen, and R. G. Almond, "You can't fatten a hog by weighing it—or can you? Evaluating an assessment for learning system called ACED," *International Journal of Artificial Intelligence in Education*, vol. 18, no. 4, pp. 289–316, 2008. (in English).
[26] K. Arslan, "Artificial intelligence and its applications in education," *Western Anatolia Journal of Educational Sciences*, vol. 11, no. 1, pp. 71–88, 2020. (in English).
[27] C. Kooli, "Chatbots in education and research: A critical examination of ethical implications and solutions," *Sustainability*, vol. 15, no. 7, p. 5614, 2023. [Online]. Available: https://doi.org/10.3390/su15075614 (in English).
[28] P. Bruneau, J. Wang, L. Cao, and H. Trương, "The potential of ChatGPT to enhance physics education in Vietnamese high schools," 2023. [Online]. Available: https://doi.org/10.35542/osf.io/36qw9 (in English).
[29] K. Taneja et al., "Jill Watson: A virtual teaching assistant powered by ChatGPT," in *International Conference on Artificial Intelligence in Education*, pp. 324–337, Springer Nature Switzerland, 2024. [Online]. Available: https://doi.org/10.1007/978-3-031-64302-6_23 (in English).
[30] I. Kostka and R. Toncelli, "Exploring applications of ChatGPT to English language teaching: Opportunities, challenges, and recommendations," *TESL-EJ*, vol. 27, no. 3, 2023. [Online]. Available: https://doi.org/10.55593/ej.27107int (in English).

Text of the article was accepted by Editorial Team 09.07.2025

ЗАПОВНЮЮЧИ ПРОГАЛИНИ: ЯК CHATGPT СПРИЯЄ ОЦІНЮВАННЮ НАВЧАЛЬНОГО ПРОГРЕСУ В СЕРЕДНІЙ ШКОЛІ

Мехмет Акіф Ардуч
доктор наук, викладач природничих наук
Міністерство національної освіти, природнича освіта,
м.Адияман, Туреччина
ORCID ID 0000-0002-1026-2843
arducakif@gmail.com

Гюльшах Гюркан
доктор філософії, доцент

Університет Малатья Тургут Озал,
природнича освіта, м. Малатья, Туреччина
ORCID ID 0000-0003-0297-3060
gulsah.gurkan@ozal.edu.tr

Анотація. Це дослідження вивчає вплив оцінювання за допомогою ChatGPT на подолання прогалин у навчанні учнів. Зі зростанням інтеграції штучного інтелекту (ШІ) в освіту розуміння його потенційних переваг у підвищенні академічної успішності учнів має велике значення. З цією метою було проведено квазіекспериментальне дослідження в рамках кількісної методології. У дослідженні протягом 2024-2025 навчального року взяли участь учні сьомого класу державної середньої школи в Туреччині. Для оцінювання рівня навчальних досягнень учнів використовувались питання з підручника для сьомого класу з розділу "Сонячна система та поза нею", опублікованого Міністерством національної освіти (МНО). Ці питання з вибором однієї правильної відповіді складалися з чотирьох варіантів і пройшли перевірку на валідність і надійність. Перед публікацією Міністерство національної освіти (МНО) звернулось до експертів та провело пілотне впровадження для аналізу ефективності завдань. Усі запитання були розроблені відповідно до рівня розвитку учнів і чітко узгоджені з розділом "Сонячна система та поза нею". Під час вивчення розділу "Сонячна система та поза нею" експериментальна група брала участь у процесі опитування за допомогою ChatGPT. Замість прямих відповідей студенти отримували покрокові підказки до правильного рішення та негайний зворотний зв'язок. Результати показали, що оцінювання з підтримкою ChatGPT ефективно усуває прогалини в навчанні, а експериментальна група продемонструвала статистично значуще покращення результатів навчання. Це дослідження дає цінну інформацію про потенціал чат-ботів з підтримкою ChatGPT для скорочення прогалин у навчанні в освітніх установах і припускає, що вони можуть сприяти поліпшенню академічної успішності студентів. Крім того дослідження пропонує нові емпіричні докази того, що чат-боти з підтримкою ШІ можуть активно сприяти навчанню, керуючи процесами міркувань учнів, а не просто надаючи правильні відповіді, представляючи в такий спосіб новий підхід до формувального оцінювання з підтримкою ШІ в освіті.

Ключові слова: ChatGPT (GPT-4); аналіз та оцінювання; учні середніх класів; прогалини в навчанні; Сонячна система та поза нею.

This work is licensed under Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.