LEVERAGING GENERATIVE AI TO ENHANCE SIXTH GRADERS' SCIENTIFIC ARGUMENTATION QUALITY ON THE GREENHOUSE EFFECT TOPIC

ERENAY ATAY

BOĞAZİÇİ UNIVERSITY

2025

LEVERAGING GENERATIVE AI TO ENHANCE SIXTH GRADERS' SCIENTIFIC ARGUMENTATION QUALITY ON THE GREENHOUSE EFFECT TOPIC

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

In partial fulfillment of the requirements for the degree of

Master of Arts

in

Educational Technology

by

Erenay Atay

Boğaziçi University

2025

Leveraging Generative AI to Enhance Sixth Graders' Scientific Argumentation Quality on the Greenhouse Effect Topic

The thesis of Erenay Atay

has been approved by:

Assist. Prof. Mutlu Şen Akbulut _______________________________
(Thesis Advisor)

Assist. Prof. Gaye Defne Ceyhan _______________________________

Assoc. Prof. Bahadır Namdar _______________________________
(External Member)

May 2025

DECLARATION OF ORIGINALITY

I, Erenay Atay, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature……………………………………………...

Date…………………………………………...............

ABSTRACT

Leveraging Generative AI to Enhance Sixth Graders' Scientific Argumentation Quality on the Greenhouse Effect Topic

This case study aimed to explore how a learning module designed based on the Argument-Driven Inquiry (ADI) instructional model with generative AI integration facilitated sixth-grade students' written scientific argument quality in the context of the greenhouse effect, their interaction with the chatbot, and their perceptions of this experience. Conducted after school hours at a public middle school in Mersin, Turkey, the study involved five high-achieving sixth-grade students over three weekly 120-minute sessions. During each session, participants investigated different aspects of the greenhouse effect results, engaging in various simulations, text materials, and chatbot interactions to form their scientific arguments. Data, including participants' written scientific arguments, chatbot conversation logs, and perception forms, were analyzed using content analysis to assess argument quality and participant experiences. Findings showed that the chatbot's feedback and guidance improved argument quality, particularly in refining claims, evidence, and reasoning. Although participants initially struggled with interpreting complex responses, the chatbot generally produced contextually appropriate responses. Participants reported enhanced topic understanding and new knowledge acquisition. Due to technological limitations at the school, the small sample size led the study to focus on individual written argumentation, offering deep insights into chatbot efficacy while limiting the transferability of findings to larger populations.

ÖZET

Üretken Yapay Zekanın Altıncı Sınıf Öğrencilerinin Sera Etkisi Konusundaki Bilimsel Argümantasyon Kalitesini Artırmadaki Rolünün İncelenmesi

Bu vaka çalışmasının amacı Argüman Tabanlı Sorgulayıcı Araştırma (ATSA) öğretim modeline dayalı olarak tasarlanan ve üretici yapay zeka entegrasyonu içeren bir öğrenme modülünün, sera etkisi bağlamında altıncı sınıf öğrencilerinin yazılı bilimsel argüman kalitesine ve bu deneyime ilişkin algılarına nasıl katkı sağladığını araştırmaktır. Çalışmaya beş başarılı altıncı sınıf öğrencisi katılmış ve öğrenme modülü, üç hafta boyunca art arda uygulanan üç oturumda gerçekleştirilmiştir. Öğrencilerin yazılı bilimsel argümanları, chatbot etkileşimleri ve öğrenme modülüne ilişkin yazılı görüşleri veri olarak toplanmış ve içerik analizi yöntemiyle değerlendirilmiştir. Bulgular, sohbet robotunun verdiği geri bildirim ve rehberliğin, özellikle iddia, kanıt ve gerekçelerin geliştirilmesinde, yazılı argüman kalitesini artırdığını ortaya koymuştur. Katılımcılar başlangıçta karmaşık yanıtları yorumlamakta zorlansalar da, sohbet robotu genel olarak bağlama uygun ve destekleyici yanıtlar üretmiştir. Katılımcılar konuya dair anlayışlarının arttığını ve yeni bilgiler edindiklerini bildirmiştir. Okulun yetersiz teknolojik altyapısından kaynaklanan sınırlamalar nedeniyle küçük bir örneklemle yürütülen bu çalışma, bireysel yazılı argümantasyona odaklanmıştır. Bu durum, sohbet robotunun etkililiğine dair derinlemesine bağlamsal içgörüler sunarken, elde edilen bulguların daha geniş gruplara genellenebilirliğini sınırlamıştır.

ACKNOWLEDGMENTS

I sincerely thank my supervisor, Assist. Prof. Mutlu Sen-Akbulut, for her unwavering support, insightful guidance, patience, and encouragement throughout my journey to complete this thesis. I feel immensely fortunate to have had the privilege of working under her guidance and will remain deeply grateful for this opportunity.

I extend my gratitude to the members of my thesis committee, Assist. Prof. Gaye Defne Ceyhan and Assoc. Prof. Bahadır Namdar, for their insightful comments and valuable suggestions throughout this process.

I am deeply grateful to my family, particularly my mother, Ayşen Atay, for their unwavering support and constant presence throughout this journey. I sincerely thank my colleague, Nuray Gültekin, for providing technological devices during the data collection process, offering detailed insights about her classroom to guide me, and for her continuous support during the thesis writing process. I also sincerely thank my dear friend, İpek Paksoy, for her insightful ideas that nurtured my creativity and for consistently keeping my motivation high throughout this endeavor. Finally, I express my gratitude to my close friend, Cemre Duran, for always listening, understanding, and providing steadfast support during this period.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX FIGURES

CHAPTER 1

INTRODUCTION

The Earth's average temperature has been rising each year, primarily due to the greenhouse effect. This phenomenon exacerbates global warming and accelerates climate change, resulting in widespread consequences for agriculture, public health, and natural ecosystems (IPCC, 2023). The rise in these gases can be attributed to excessive production and consumption (Jeffry et al., 2021; Mitchell, 1989). For instance, carbon dioxide (CO_2), a major greenhouse gas, has significantly increased, particularly due to the extensive use of fossil fuels in transportation, energy generation, and industrial production (Jay et al., 2023). Although it is possible to mitigate global temperature rise by maintaining greenhouse gas emissions at specified levels, reducing emissions is a gradual process. Due to the prolonged presence of greenhouse gases in the atmosphere, their immediate removal is not feasible, and phasing out human activities such as fossil fuel use requires time and systemic transformation. Even if emissions decline, any future emissions exceeding the amount removed from the atmosphere will continue to elevate global temperatures. Achieving a balance between greenhouse gas emissions and removals, known as net-zero emissions, has the potential to stabilize global temperatures (IPCC, 2023). Consequently, the extent of global warming hinges on future emissions influenced by human activities and decisions. One practical pathway to achieving this balance is through the adoption of zero-carbon transportation and heating systems by both governments and citizens, which can significantly contribute to reaching net-zero carbon emissions (Jay et al., 2023).

Therefore, public action plays a crucial role in regulating human activities that contribute to greenhouse gas emissions. Addressing this challenge requires a collective responsibility, beginning with raising awareness at the individual level. One effective approach is to educate the public about the scientific dimensions of climate change and support informed decision-making processes (Gautier, 2012). Through such educational efforts, individuals can gain insight into how scientific knowledge is constructed and recognize that science is also a social endeavor. This awareness, in turn, contributes to the development of scientific literacy among citizens (Driver et al., 2000).

Developing such scientific literacy begins in the classroom, where students are introduced to practices that mirror how science operates in the real world. To promote scientific literacy, students need to practice evaluating evidence and relating it to explanations through reasoning, which is also the practice of scientific argumentation (Lieskovský & Sunyík, 2022). Scientific argumentation is accepted as a way of doing science; therefore, to raise scientifically literate citizens, it is regarded as a fundamental practice in science education (Driver et al., 2000).

Scientific argumentation can be developed through both writing and speaking about science (Jimenez & Erduran, 2007). In order to write scientifically, students need to learn how to construct explanations that combine evidence and reasoning to support their claims (McNeill et al., 2006), a process central to scientific argumentation.

Constructing a scientific argument requires students to integrate evidence with reasoning to support their claims, an intellectually demanding task that involves identifying relevant data and comprehending the scientific principles underlying the phenomenon (Sadler, 2004). Moreover, many students struggle to construct coherent arguments and require additional scaffolding and support throughout the process (Hendratmoko et al., 2023). While many students face difficulties during argument construction, this challenge persists even among high-achieving learners. High-achieving students face challenges when developing their arguments and considering multiple perspectives in scientific argumentation. While they are often capable of producing structurally sound arguments, they may struggle to incorporate sociological dimensions or broader implications of the phenomenon. This underscores the need for learning environments that provide timely and formative feedback during the argument construction process (Ho et al., 2019; Kızkapan & Nacaroğlu, 2020).

The Argument-Driven Instructional (ADI) model is an approach that facilitates peer discussion and iterative writing, providing timely feedback through group work. At the end of this model, the teacher provides feedback, and students revise their arguments accordingly (Sampson et al., 2010). However, providing such personalized and timely feedback poses a persistent challenge for teachers, particularly in large classroom settings (Henderson et al., 2017).

In response to these challenges, recent technological advancements have opened new possibilities to support students in constructing scientific arguments. One such innovation is the integration of artificial intelligence (AI) tools, particularly generative AI-based chatbots, into educational contexts. Generative AI chatbots, which are trained on extensive datasets, are designed to generate human-like responses that align contextually with user inputs (Martineau, 2024).

These chatbots serve as digital learning partners by engaging students in dialogue, offering scaffolded responses tailored to their input, and providing immediate, contextualized feedback during the process of argument construction. (Guo et al., 2023; Su et al., 2023; Yang et al., 2022).

However, designing such chatbots necessitates expertise in coding. To tackle this issue, Platform for Open Exploration (POE) was introduced, a web-based interface that simplifies the design of chatbots using prompt-based inputs. The platform provides access to a variety of generative AI models without the need for coding skills or paid subscriptions. This innovative platform allows users to create chatbots based on their prompts through multiple generative AI interfaces, all without the requirement of payment or technical expertise. Although literature emphasizes the use of POE in foreign language education (Lee et al., 2024; Wu & Xu, 2025), no existing studies have investigated its application with middle school students to improve the quality of written scientific argumentation regarding the greenhouse effect. This study aims to fill this gap by exploring how a chatbot designed via POE can support middle school students' quality of written scientific argumentation in the context of the greenhouse effect.

1.1 Significance of the study

Scientific argumentation has been found beneficial for students in supporting metacognitive processes, communication skills, and scientific literacy (Erduran & Jimenez, 2007). Science education also aims to instill the awareness that science is a social endeavor, and research about science education has been used to analyze how students construct and evaluate scientific claims while learning to do science (Driver et al., 2000). As outlined by the Ministry of National Education (MoNE, 2018), the Turkish national curriculum emphasizes learning processes that involve exploration, argumentation, and reasoning. The recently introduced Maarif Model (MoNE, 2024) further highlights critical thinking as a central tendency for sixth-grade students, encouraging them to evaluate and compare arguments by analyzing information. Consequently, sixth-grade students are expected to practice argumentation in their science education.

Even though the development of scientific argumentation is one of the main goals of science education, it is not frequently experienced by students in the classroom setting (Kuhn, 1992). Therefore, students have difficulty practicing scientific argumentation because they are unfamiliar with asking questions using their reasoning while forming explanations (Simon et al., 2006). To address these challenges, the implementation of innovative instructional tools has been proposed as an effective solution (Henderson & Aguilera, 2020).

In response to this need, this study contributes to the literature by introducing an original learning module that integrates scientific argumentation with generative AI support through a chatbot designed using the POE platform. Unlike previous studies, which have focused primarily on higher education or language learning contexts, this study explores how AI-supported scaffolding can facilitate written argumentation at the middle school level, specifically within a public school setting. Since students in such contexts may have limited experience with both scientific argumentation and emerging technologies (Driver et al., 2000; Topal et al., 2024), the findings of this study may offer valuable implications for science educators, curriculum designers, and researchers seeking innovative strategies to promote scientific argumentation in diverse classroom environments.

1.2 Purpose of the study

This study aims to explore how a learning module designed based on the Argument-Driven Inquiry (ADI) instructional model with generative AI integration facilitates sixth-grade students' written scientific argument quality in the context of the greenhouse effect topic and their perceptions of this experience.

1.3 Research questions

Overarching research question: How does a specifically designed learning module based on the ADI instructional model, integrated with Generative AI (hereafter referred to as the module), facilitate sixth-grade students' written argumentation quality and their perceptions of the learning experience?

Research Question 1: How does the quality of the scientific arguments written by sixth-grade students in the context of the greenhouse effect improve during their interaction with the learning module?

Research Question 2: How do students interact with a specifically designed chatbot to develop their arguments during the learning module?

- RQ 2.1 For which purposes do students use the chatbot in the context of the scientific argument formation process?
- RQ 2.2 Through which functions do the chatbot respond to students' requests in the context of the scientific argument formation process?
- RQ 2.3 What is the quality of chatbot responses in supporting participants' scientific argumentation?

Research Question 3: What are the students' perceptions of the effectiveness of the chatbot for developing their arguments?

1.4 Organization of sections

This thesis is organized into five chapters. Chapter 1 outlines the significance of the study, the purpose of the study, the research questions, and the overall organization of the thesis. Chapter 2 provides a comprehensive literature review, covering the greenhouse effect and scientific argumentation, frameworks and approaches related to scientific argumentation, teaching strategies and models for scientific argumentation, the Argument-Driven Inquiry (ADI) instructional model, challenges in teaching scientific argumentation and potential solutions, as well as chatbot affordances and empirical studies on chatbot technologies. Chapter 3 details the research methodology, including the research design, case, participants, data collection instruments, data collection procedures, data analysis methods, and measures to ensure the trustworthiness of the study. Chapter 4 presents the findings of the study. Chapter 5 discusses the findings, offers recommendations and implications for future research, and addresses the limitations of the study.

CHAPTER 2

LITERATURE REVIEW

The excessive accumulation of greenhouse gases has led to severe environmental consequences, including global warming, melting ice caps, and rising sea levels. If societies take required precautions, the amount of the gases can be kept at a stable level, which depends on informed decisions of citizens (Gautier, 2012; IPCC, 2023). Facilitating informed decisions and public action can be possible through evidence-based discussion environments, which can be practiced by scientific argumentation (Lieskovský & Sunyík, 2022). Scientific argumentation can be practiced with various teaching strategies, including Argument Driven Inquiry (Sampson et al.,2010). However, while teaching scientific argumentation, delayed feedback and the need for instructional support affect students' scientific argumentation learning process (Henderson et al., 2017; Wen et al.,2015). Chatbots have been used to eliminate these difficulties in the literature (Yang et al.,2022). The POE interface, which enables users to access and customize various chatbot services without requiring coding skills, has recently gained attention in educational research. However, existing studies utilizing the POE platform have predominantly focused on foreign language instruction (Lee et al., 2024; Wu & Xu, 2025). To date, no study has employed a specifically designed chatbot via POE to support the development of scientific argumentation skills among middle school students, particularly within the context of teaching the greenhouse effect. This study integrates scientific argumentation within the context of the greenhouse effect's consequences, leveraging generative AI to support the learning process. The literature review section establishes the necessity of scientific argumentation in the context of results of the greenhouse and examines teaching strategies for scientific argumentation along with their associated challenges. It then reviews the use of generative AI in education, focusing on its application in scientific argumentation instruction. Finally, the significance of this study for the existing literature is discussed, highlighting its contribution to addressing the identified gaps.

2.1 The greenhouse effect as a context for scientific argumentation

Human activities are the main source of greenhouse gases that cause global warming (IPCC, 2023). Due to excessive amounts of greenhouse gases, global surface temperatures rose by 1.1°C at the beginning of the 20th century. The increase in greenhouse gas emissions is strongly linked to unsustainable energy use, consumption, and production patterns in different regions. An increase in greenhouse gas emissions will give rise to 1.5°C global warming, which can lead to hazards in multiple aspects. If rapid and sustained reductions are made, a slowdown in global warming can be obtained in two decades (IPCC, 2023).

Ultimately, the decisions made by governments and society will determine future emissions. For instance, promoting the use of clean energy sources can lead to achieving net-zero carbon emissions (Jay et al., 2023). However, confronting this crisis requires more than technological solutions; it demands a scientifically literate society capable of making informed, evidence-based decisions. Therefore, individuals, as the smallest unit of society, should be educated in a way that their actions can influence others' decisions. To promote informed decision-making, the scientific aspects of the situation should be taught (Gautier, 2012). This way, citizens can understand that science is deeply intertwined with society, and as a result, their scientific literacy can be fostered (Driver et al., 2000).

In science education, one of the goals is to cultivate students' ability to make evidence-based decisions and to foster scientific literacy (Jimenez & Erduran, 2007). As long as students notice scientific knowledge beyond each socioscientific issue and learn to make evidence-based decisions, they will be willing to take action on environmental issues (Gautier, 2012). To foster the formation of informed decisions, students should be educated as scientifically literate. A way of teaching scientific literacy is by preparing evidence-based learning environments, which can be sustained through scientific argumentation practice because students evaluate evidence and use scientific data to form an argument by connecting reasoning. To make students scientifically literate, students need to learn to evaluate evidence and improve their reasoning. Therefore, they need evidence-based discussion environments that can be sustained through scientific argumentation (Lieskovský & Sunyík, 2022).

Scientific argumentation is also a widely used strategy to practice reasoning, evidence evaluation, and argument formation. Besides, when students attend scientific argumentation activities, they enhance their communication skills, critical thinking abilities, and scientific literacy (Erduran & Jimenez, 2007). For scientific argumentation, socioscientific issues, which are social issues that have a significant scientific component, enable best practice since these issues provide topics from daily life; thus, students can be familiar with the issue and use their informal reasoning and argumentation skills. Therefore, students can learn how to evaluate information, how science works, and scientific concepts (Sadler, 2004). Even though students can struggle to form arguments about newly learned topics, when they are given a topic from daily life, they can interact with it willingly. Therefore, this study focuses on the greenhouse effect, a socioscientific issue rooted in daily life, as a context for enhancing students' informed decision-making and scientific argumentation skills.

2.2 Frameworks and approaches to scientific argumentation

In this section, diverse argumentation frameworks, including Toulmin (1984) and McNeill et al. (2006), will be discussed. In the scope of this study, the Claim-Evidence-Reasoning (CER) framework for constructing scientific explanations, as proposed by McNeill et al. (2006), was employed. This framework has also been found to be compatible with K–12 students' cognitive levels (Zhou, 2023).

An argument is a statement of conclusions reached by reasoning (Brockriede, 1975). On the other hand, scientific argumentation is a process where one attends a debate to persuade the audience in a scientific context (Berland & Hammer, 2011). According to Toulmin (1984), an argument has different components: data, claim, grounds, warrant, qualifier, rebuttal, and backing. Claim means the main argument, while grounds include a rationale to put forward the claim, in other words, evidence to express the claim. A warrant states assumptions about the situation. Backing supports the warrant, explaining why this assumption is made. The qualifier makes the claim specific for some situations; thus, the claim's validity can be improved. Rebuttal mentions another possible situation; in other words, if another condition that gives rise to the same situation doesn't exist (Toulmin, 1984).

The argument components proposed by Toulmin (1984), specifically warrant, backing, and qualifier, pose significant challenges for differentiation and effective utilization among students. As a result, numerous researchers have identified the Toulmin model as inherently difficult to implement within K–12 educational contexts (Erduran et al., 2004).

To address these pedagogical limitations, McNeill et al. (2006) put forward another model that consists of claim, evidence, and reasoning components. According to this model, the claim is a conclusion to answer the research question, while the evidence is scientific data that supports the claim. The scientific data can be obtained via investigation or existing data, which should be sufficient to support the claim. On the other hand, the reasoning explains why the data is accepted as evidence (McNeill et al., 2006). In other words, reasoning answers why they believe the evidence is compatible with the claim. From this perspective, the reasoning is similar to Toulmin’s warrant and backing. However, to avoid confusion, a term of reasoning was preferred (McNeill et al., 2006).

In this study, the CER framework was chosen for its alignment with the cognitive development of middle school students and its effectiveness in enhancing argumentation skills (Zhou, 2023).

2.3 Teaching strategies for scientific argumentation

This section reviews existing literature on improving students’ scientific argumentation, including teaching models and strategies.

According to a literature review study by Hendratmoko et al. (2023), integrating inquiry and debate activities improves students’ scientific argumentation skills. Inquiry activities are needed since forming arguments requires effective integration of evidence and reasoning. Besides, to revise and refute claims, there is a need for opposing arguments, which can be facilitated with debate activities (Hendratmoko et al., 2023).

In the context of SSI, reflective and debate activities have been shown to improve students' argumentation skills significantly. In a study by Bächtold et al.

(2022), researchers investigated how these activities enhanced argumentation in 16- to 18-year-old students. Reflective activities, which focus on self-checking arguments, and debate, which encourages the explanation of alternatives, were found to be particularly effective. Combining these two teaching strategies resulted in better student arguments (Bächtold et al.,2022).

Existing studies have consistently demonstrated the efficiency of inquiry, debate, and reflective activities in improving argumentation (Bächtold et al.,2022; Hendratmoko et al., 2023). When students have an inquiry environment that allows them to analyze and synthesize data, they are better equipped to discuss their findings with their classmates (Hendratmoko et al., 2023). Similarly, when students engage in reflective activities, such as revising their work, they are able to produce better arguments (Bächtold et al.,2022). The Argument-Driven Inquiry instructional model effectively facilitates these activities, allowing students to plan and carry out investigations, construct explanations, and share their findings with their classmates.

The Argument-Driven Inquiry (ADI) instructional model has been recommended in literature as an effective approach to teaching scientific argumentation (Sampson et al., 2010). Given its emphasis on inquiry, collaboration, and evidence-based reasoning, the ADI model was selected as the instructional backbone of this study. Therefore, this section was placed to explain the details of the ADI model.

2.4 Argument-driven inquiry instructional model

According to this model developed by Sampson et al. (2010), seven interrelated steps exist to promote the formation of scientific arguments. ADI model begins with the "identification of the task," aiming to introduce the topic and capture the students'

attention and interest. In the second step, “generation of data”, students collaboratively design a method for investigation and apply it accordingly.
As a third step, “production of a tentative argument,” they produce tentative arguments based on their findings, using the claim-evidence-reasoning (CER) structure. According to the ADI model, a claim explains a research question, while evidence means observation or measurement to support the claim. Reasoning, on the other hand, integrates evidence into the claim by answering why the evidence supports the claim. In the fourth step, “argumentation session,” students share their arguments with other groups to compare and critique their own and their classmates’ arguments. The fifth stage involves “creation of a written investigation report,” where students write an investigation report. During the sixth step, known as “double-blind peer review,” students evaluate each other’s reports based on some guiding questions. Using the feedback received, groups then revise their arguments. In the final stage, titled “revision of the report,” the completed reports are submitted to the teacher, who provides additional feedback. Groups then refine their arguments until they achieve a polished final form (Sampson et al., 2010).

Although the ADI model includes both oral and written phases in argument construction (Sampson et al., 2013), this study focused exclusively on written arguments, as they offer a structured and analyzable form of students’ reasoning (McNeill & Krajcik, 2011). Writing requires students to organize their ideas coherently and support claims with evidence, enabling consistent analysis and reliable comparison of argument quality across tasks (McNeill et al., 2006).

Within the ADI model, the "production of a tentative argument" and "creation of a written investigation report" stages emphasize writing practice, requiring students to construct arguments with relevant evidence, support claims with reasoning, and revise based on feedback. As Sampson et al. (2013) highlight, when students do not engage actively in writing, evaluations remain superficial and fail to reflect their scientific writing development or learning potential through this process. Thus, written reports serve not only as records of observations but also as a medium to communicate scientific understanding and argument construction effectively (Sampson et al., 2013). Table 1 summarizes the seven steps of the ADI framework along with their instructional purposes.

Table 1. Steps of the ADI Framework Developed by Sampson et al (2010)

Step	Description
1. Identification of the Task	Introduces the topic and captures the students' attention and interest.
2. Generation of Data	Students design a method as a group to carry out an investigation that addresses the topic introduced in the first step.
3. Production of a Tentative Argument	Students form tentative arguments based on their investigations.
4. Argumentation Session	Students share their arguments with other groups for comparison.
5. Creation of a Written Investigation Report	Students write an investigation report.
6. Double-Blind Peer Review	Students evaluate each other’s reports. Based on the feedback, groups revise their arguments.
7. Revision of the Report	Reports are submitted to the teacher, who provides feedback. Groups then revise their arguments until they form a proper final form.

Previous studies showed that using the ADI model is beneficial in facilitating students in an environment that improves scientific argumentation (Kaçar & Balım, 2018; Sampson et al.,2010; Songsil et al., 2019). Sampson et al. (2010) conducted an exploratory study with 19 tenth graders to explore the effect of laboratory activities, designed based on Argument-Driven Inquiry (ADI), on the engagement of scientific argumentation and the quality of the scientific arguments produced by students. This study lasted 18 weeks with 15 lab activities. In the lab activities, students worked in groups of 3 or 4 and dealt with various lab activities for diverse purposes, including developing a new explanation, revising an explanation, evaluating an explanation, and using an explanation to solve a problem (Sampson et al., 2010).

Researchers used a performance task consisting of the candle and inverted flask to assess how the students participated in scientific argumentation and constructed a scientific argument. In order to analyze written scientific arguments produced by each group, the following components were taken into account: the adequacy and the conceptual quality of the explanation, the quality of the evidence, and the sufficiency of the reasoning (Sampson et al., 2010). According to the findings of this study, the students showed better disciplinary engagement and produced better scientific arguments after the intervention (Sampson et al., 2010).

Furthermore, Kaçar and Balım (2018) conducted a study with 32 seventh graders to investigate the effects of the argument-driven inquiry model on students' learning in the electricity energy unit. According to data obtained from student reports, reflective diaries, and answers to the evaluation questions, the Argument-Driven Inquiry method was found effective in students' conceptual understanding (Kaçar & Balım, 2018).

Apart from the ADI model, Songsil et al. (2019) also improved rADI (revised argument-driven inquiry), which is inspired by the ADI model to enhance the scientific argumentation of students by using socioscientific issues. After applying rADI, most students developed scientific argumentation skills in most components (Songsil et al., 2019).

In order to analyze scientific argumentation skills, the open-ended scientific argumentation test, the reasoning ability test, and the content knowledge tests were used. According to the results after the rADI model was applied, students in the treatment group got higher scores than those in the nontreatment group regarding scientific argumentation skills (Songsil et al., 2019). This model has some key features, including well-structured tasks, an explanation of the argumentation process, usage of group discussions, argumentation-focused activities, immediate feedback, and a safe and respectful environment (Songsil et al., 2019).

2.5 Challenges in scientific argumentation practice

This section discusses the challenges faced by students and teachers during the practice of scientific argumentation. While literature often emphasizes challenges faced by students with low prior knowledge, such as difficulty evaluating evidence or constructing complete arguments (Faize et al., 2017; Lin & Mintzes, 2010), these findings may not fully reflect the learning needs of high-achieving students. Rather than struggling with basic argument construction, high-achieving learners often need support developing deeper reasoning and identifying alternative perspectives (Ho et al., 2019; Kızkapan & Nacaroğlu, 2020).

Apart from the mentioned challenges, another central challenge is the lack of timely, formative feedback, which is essential to support the argument formation process. Delayed feedback may hinder students' willingness to revise their arguments or consider alternative viewpoints (Berland & McNeill, 2010; Walker & Lower, 2019).

In addition, teachers face challenges when instructing scientific argumentation. These challenges include time constraints, large class sizes, the complexity of providing individualized feedback, and maintaining a classroom culture rich in argumentation (Henderson et al., 2017). Accordingly, the following subsections address two key areas: the instructional needs of high-achieving students, challenges related to delayed feedback.

2.5.1 Instructional support needs of high-achieving students

According to the study by Wen et al. (2015), students with higher prior knowledge significantly outperformed students with lower prior knowledge in argumentation and online argumentation learning (Wen et al., 2015). Similarly, in the study by Ho et al. (2019), high-achieving students exhibited strong proficiency in identifying components of scientific arguments; however, similar to their lower-achieving peers, they encountered difficulties constructing counterarguments. This finding suggests that even high-achieving students may require instructional support to develop the ability to consider alternative perspectives (Ho et al., 2019).

Additionally, the research by Lin and Mintzes (2010) revealed that high achievers can also face challenges in distinguishing between argument components, such as evidence and reasoning, as well as generating rebuttals. The researchers recommend providing explicit instruction and contextualized feedback to enhance students' argumentation skills (Lin & Mintzes, 2010).

Similar to high achievers, gifted students also often demonstrate advanced content knowledge and strong reasoning skills; however, their arguments may still lack depth in specific dimensions. In the study conducted by Kızkapan and Nacaroğlu (2020), although participants were able to construct well-structured arguments due to their strong content knowledge, their responses often lacked sociological depth and consideration of broader societal implications (Kızkapan & Nacaroğlu, 2020).

Zhu et al. (2020) found that students with higher prior knowledge were more likely to revise their arguments after receiving feedback, suggesting a link between content knowledge and revision behavior. However, prior knowledge alone does not always lead to meaningful change (Zhu et al., 2020). In a study by Walker and Lower (2019), university students were reluctant to revise their claims even when presented with strong counterevidence from peers. This resistance was especially strong when group members agreed on a claim. Interviews revealed that students believed claims should only change if entirely new evidence appeared, rather than reinterpreting existing information. The authors highlighted the need for explicit instruction to help students understand when and how to revise their arguments (Walker & Lower, 2019).

In addition to the challenges outlined above, the absence of timely feedback has also been identified as a significant barrier, often reducing students’ motivation to revise and improve their scientific arguments (Berland & McNeill, 2010; Henderson et al., 2017).

2.5.2 Challenges related to delayed feedback

Since scientific argumentation is inherently a social process involving presenting, defending, and negotiating ideas with others, students may experience difficulties producing scientific arguments in written form, where immediate audience feedback and peer engagement are absent (Berland & McNeill, 2010). In addition to writing practice, students should be supported with learning environments that encourage discussion and provide opportunities to give and receive feedback with peers (Berland & McNeill, 2010). Berland and McNeill (2010) found a clear difference between the quality of written and spoken arguments. They explained that this gap is partly due to the lack of an audience in written tasks. In spoken argumentation, students try to convince others, encouraging them to include rebuttals and use stronger reasoning. However, students do not get immediate feedback on written tasks, so they are less motivated to improve their arguments. To help with this, the authors suggest using simple prompts and regular feedback to support students during argumentation (Berland & McNeill, 2010). These difficulties highlight the importance of timely support. In this regard, feedback, particularly its timing, emerges as a key factor influencing how students construct and refine their scientific arguments.

The timing of feedback has been shown to significantly influence learning performance. For instance, Lu et al. (2021) reported that participants who received immediate feedback outperformed those who received delayed feedback in terms of knowledge acquisition. Furthermore, studies in the context of language learning have demonstrated that immediate feedback not only enhances learning outcomes but also fosters higher levels of motivation compared to delayed feedback (Qun, 2025). In addition, research by Zhu et al. (2017) indicated that when participants received immediate feedback during the scientific argumentation process, they were more likely to revise and improve their arguments accordingly.

In addition to the challenges faced by students, teachers and researchers also encounter significant difficulties in implementing scientific argumentation effectively. Henderson et al. (2017), drawing on findings from discussions at the NARST conference, highlighted several recurring issues. These include establishing a classroom culture that values argumentation, offering adequate pedagogical support during student discourse, and managing the dual demands of formative and summative assessment within limited instructional time (Henderson et al.,2017).

2.6 Addressing the challenges: Scaffolding and feedback technologies

Students often face challenges in scientific argumentation because they are not always aware of what they need to know or how to approach the learning process effectively. To address this, researchers have emphasized the importance of providing scaffolding, explicit instruction, and immediate feedback to support students' learning (Jin & Kim, 2021; McNeill et al., 2006). However, in real classroom settings, simultaneously delivering both scaffolding and formative assessment can be demanding for teachers due to time constraints and diverse student needs. As a result, there is a growing need for instructional strategies and educational technologies to support teachers in managing these demands while allowing students to progress at their own pace. This section reviews research studies that offer potential solutions to these challenges, emphasizing two main areas: the role of scaffolding and immediate feedback in enhancing scientific argumentation.

2.6.1 The role of scaffolding in supporting scientific argumentation

Scaffolding has been widely recognized in literature as an instructional strategy to support the development of students' scientific argumentation skills (Yang et al.,2022). In their review of the Argument-Driven Inquiry (ADI) model, Kuki et al. (2023) emphasized that while students could attain certain cognitive competencies by constructing scientific reports, their ability to formulate well-supported arguments improved significantly when scaffolding was provided. Integrating guiding questions was recommended as a practical approach to support students' reasoning processes during argument construction (Kuki et al., 2023).

Providing scaffolding plays a vital role in enhancing students' self-regulation skills, which contributes to constructing more coherent and scientifically grounded explanations (McNeill et al.,2006). In their study, McNeill et al. (2006) demonstrated that incorporating fading scaffolds during the development of scientific explanations, particularly those involving claim, evidence, and reasoning components, led to improved performance among seventh-grade students. Fading scaffolds gradually reduce external support, encouraging students to assume greater responsibility in their learning. This independence enables them to engage more actively with scientific inquiry processes, which are foundational to fostering scientific literacy. The study further revealed that students who received fading scaffolds demonstrated stronger reasoning in their scientific explanations compared to those without such support (McNeill et al.,2006).

Similarly, in a case study conducted by Jin and Kim (2021) involving fifth and sixth graders, researchers explored the impact of argument-focused metacognitive scaffolds on student argumentation. They utilized techniques such as questioning, prompting, and modeling of thinking as argument metacognitive supports (AMS). The findings indicated that, with the assistance of AMS, students enhanced their argumentation skills by gaining a better understanding of the nature of argumentation. This allowed them to monitor their thinking more effectively. Consequently, students were able to comprehend the requirements for forming persuasive and credible arguments (Jin & Kim, 2021).

While scaffolding is essential for developing argumentation skills, the way they are delivered is also very important. Therefore, the frequency and design of scaffolding should be adapted to ensure that students utilize them as needed. Alongside scaffolding, students require formative assessments of their work during the learning process, which calls for immediate feedback (Henderson et al., 2017). However, it can be challenging for teachers to provide immediate feedback to every student. As a result, research has explored technological solutions that offer instant feedback, such as automated scoring systems and chatbots (Darvishi et al., 2024; Haudek & Zhai, 2023; Lee et al., 2021; Wang et al., 2024; Zhai et al., 2022; Zhu et al., 2017, 2020).

2.6.2 The role of immediate feedback provided by automated feedback systems

Zhu et al. (2017) conducted a study with high school students to explore how scientific argumentation is formed by incorporating an online learning module on climate change, alongside an automated scoring and feedback system. During the study, students submitted open-ended responses and received scores and written feedback from the scoring engine simultaneously. The findings revealed that students with higher initial scores were more likely to engage in revision activities compared to those with lower scores. Furthermore, when comparing the final scores, it was found that students who revised their arguments achieved significantly higher results than those who did not. Overall, each revision was associated with an average increase in the final scores (Zhu et al., 2017).

A study conducted by Zhu et al. (2020) introduced a learning module designed to teach climate change. This module includes a feedback system that utilizes automated scoring technologies, allowing students to revise their scientific arguments until they achieve the desired score. The researchers examined how student performance and learning gains changed after receiving revisions through the formative feedback system. Their findings indicated that students with higher initial knowledge tended to engage in more revision behavior when they received automated feedback. Additionally, as students revised their arguments, their scores improved. When comparing different types of feedback, the study found that contextualized feedback was particularly effective. Furthermore, all students benefited from immediate feedback on complex components, regardless of their prior knowledge levels (Zhu et al., 2020).

In several studies, the formation of arguments was integrated into simulation activities. Simulations allow students to manipulate variables and observe the resulting effects, thereby fostering an inquiry-based learning environment (Lee et al., 2021). In their research, Lee et al. (2021) developed an online digital curriculum module that enables students to interact with simulations while learning about groundwater systems. With the automated scoring system, students receive immediate feedback, including their scores and explanatory statements, upon submitting their answers. The study examined the effects of providing both simulation and argument feedback compared to offering only argument feedback (Lee et al., 2021). The findings revealed that students who received both types of feedback demonstrated significantly higher scores than those who received argument feedback alone. Additionally, students who achieved higher scores in the simulations tended to perform better in their argument scores for one of the tasks. However, it was noted that fewer than 20% of the students utilized the simulations to their fullest potential. Furthermore, when the system did not provide simulation feedback, students often failed to revisit and redo the simulations as needed to refine their observations and gather better evidence. Thus, this study emphasizes the necessity of providing specific instructions and guiding questions to enhance the effective use of simulations (Lee et al., 2021).

Although automated feedback has been shown to promote revision behavior effectively, the accuracy of this feedback remains a topic for discussion. In a study by Zhai et al. (2022) involving students in grades 5 to 8, researchers compared the assessments of written arguments made by a machine using the "Cognitive Diagnostics Modeling Approach" with those made by human experts. During the study, the human experts evaluated the students' responses and subsequently trained the machine to assess the writing using an algorithm. The results demonstrated a high level of agreement between the assessments from the experts and those from the machine (Zhai et al., 2022).

In the study conducted by Haudek and Zhai (2023), which evaluates the performance of machine learning in the context of scientific argumentation, the researchers identified that an increase in the complexity and diversity of assessment tasks correlated with a decline in model performance. Moreover, a negative linear relationship was observed between the level of task structure and the performance of the machine learning models (Haudek & Zhai, 2023).

Similarly, the investigation by Wang et al. (2024) focused on the feedback capabilities of ChatGPT through an analysis of pre-existing assessed argument sets derived from undergraduate students. This study aimed to assess the reliability and credibility of the feedback produced by ChatGPT in comparison to traditional teacher feedback. The results indicated that ChatGPT achieved a notable precision rate of 91% in providing quantitative feedback, which included metrics such as the number of rebuttals and the identification of discourse markers. However, its performance in delivering qualitative feedback about the adequacy of evidence and explanations was comparatively lower. The findings suggested that several factors, including the length of the argument, content specificity, and the presence of discourse markers, influence ChatGPT's efficacy in providing feedback. Notably, as the length of the argument increased, the performance of ChatGPT exhibited a decline. Furthermore, when contrasting the feedback from ChatGPT with that of human educators, the researchers concluded that ChatGPT predominantly produced text-based linguistic feedback. At the same time, teachers offered feedback that considered the broader context of student learning. This led to the observation that ChatGPT lacks the capability to establish coherent logical relationships among extensive information, resulting in a tendency towards decontextualization. Consequently, the study advocates for a collaborative approach to feedback in argumentation instruction, integrating both teacher and GPT feedback (Wang et al., 2024).

Expanding on the efficiency of AI feedback, the research conducted by Darvishi et al. (2024) explored the impact of AI assistance on student agency during peer feedback processes. The findings suggested that students utilizing AI prompts were more inclined to provide specific and detailed feedback rather than superficial comments. However, in the absence of AI assistance, students encountered difficulties in replicating the quality and specificity of feedback that AI facilitated (Darvishi et al., 2024).

2.7 Affordances of chatbots in supporting scientific argumentation

Recent advances in AI have introduced chatbot technologies as promising tools for supporting students' scientific argumentation practices, in addition to automated scoring systems that provide immediate but limited feedback. Automated scoring systems facilitate revision behavior in students' arguments (Zhu et al., 2020). However, when students encounter additional questions, they may find themselves at an impasse due to the lack of an interactive framework (Chang et al., 2023). To address the challenges associated with scientific argumentation, research suggests leveraging educational technologies, particularly the capabilities of chatbots. Consequently, this section explores the benefits of chatbots as highlighted in the literature.

The integration of chatbots in educational contexts has facilitated a more interactive learning environment, allowing students to ask questions freely and receive immediate, in-depth responses. These dynamics foster automated scoring and continuous interaction (Chang et al., 2023). Importantly, educators are allowed to monitor individual student participation in various activities. In a pertinent case study by Chang et al. (2023), the affordances of artificial intelligence chatbots in guiding inquiry-based activities on sound transmission were explored with gifted fifth and sixth-grade students. Over three lessons, each lasting 40 minutes, students engaged in inquiry-based tasks, with the AI chatbot providing tailored responses to their inquiries. This capability enabled students to adjust their learning processes at their own pace. Moreover, in group settings, monitoring each student's contributions can prove challenging for teachers; however, the integration of chatbot technology allowed educators to effectively track student interactions through the chatbot's conversational records, thereby bridging the interaction gap between groups and instructors (Chang et al., 2023). Additionally, students expressed that the Inquirybot enhanced their understanding of the inquiry activities and procedures (Chang et al., 2023).

Despite the advantages of chatbot integration for scaffolding and monitoring processes, some quantitative studies have reported negligible differences in student outcomes. For instance, Topal et al. (2021) investigated the impact of chatbots on students’ academic performance and perceptions among 5th graders engaged in a unit on 'Matter and the Changing State of Matter' via the Telegram platform. Although the study revealed no significant performance differences between the treatment and control groups, students reported that the chatbot was useful and enjoyable. This lack of significant difference did not preclude the treatment group from demonstrating academic achievement, leading researchers to suggest the necessity for further studies in face-to-face learning environments, particularly given that this research was conducted online during the pandemic (Topal et al., 2021).

Furthermore, Yang et al. (2022) conducted a quasi-experimental study that found no significant differences in conceptual understanding among fifty-five sixth-grade students engaged in web-based synchronized argumentation. This study examined conceptual understanding, scientific explanation, and argumentative behavior through the Claim-Evidence-Reasoning (CER) framework (McNeill, 2006). Students in the experimental group were provided with scaffolding prompts, including sentence starters, while their control group generated arguments independently. The results indicated no significant difference in conceptual understanding; however, the experimental group outperformed their peers regarding scientific explanation. Additionally, the experimental group's arguments were more coherent and evidence-based, contrasting with the control group's challenges in collecting relevant evidence. These findings suggest that scaffolding plays a crucial role in facilitating the construction of evidence-based arguments. Nevertheless, the study overlooked the quality of arguments produced (Yang et al., 2022).

In a different quasi-experimental investigation by Guo et al. (2023), involving 44 undergraduate students, a chatbot known as "Argumate" was employed prior to group discussions to enhance students' argumentation skills and task motivation. The "Argumate" chatbot provided diverse perspectives and posed questions, encouraging students to formulate comprehensive arguments. Upon completion of the conversation, the chatbot generated a mind map to aid students in organizing their thoughts (Guo et al., 2023). A pre- and post-assignment in essay writing assessed structural complexity and argument quality in students' argumentation skills. The findings indicated that participants who utilized the "Argumate" chatbot produced more organized, substantial, and elaborate arguments. Researchers concluded that the chatbot effectively encouraged students to support their positions by considering multiple perspectives (Guo et al., 2023).

2.7.1 Comparison of generative and rule-based chatbots

The studies mentioned utilized rule-based chatbots, which limited student interaction by restricting feedback to preassigned tasks (Ng et al., 2024). In contrast, the current study focused on a chatbot designed using generative AI. This section highlights studies that have employed generative AI-based chatbots.

Ng et al. (2024) conducted a quasi-experimental study to examine the differences in science knowledge, motivation, and self-regulated learning between groups that interacted with a generative AI-based chatbot (SRLbot) and those using a rule-based AI chatbot (Nemobot). The study involved seventy-four secondary school students in their fourth year. A single teacher provided instruction on the topic of "force and motion" as well as guidance on chatbot operations for both groups (Ng et al., 2024).

After the instruction, pairs of students completed a topic-related task using the chatbots. The experimental group used the generative SRLbot, while the control group utilized the rule-based Nemobot. The findings revealed that the SRLbot offered personalized support and effective self-regulation learning strategies tailored to students' needs. Additionally, it provided suggestions for further learning resources and offered guidance through step-by-step solutions and hints throughout the learning process (Ng et al., 2024).

On the other hand, students with lower levels of content knowledge reported that they received lengthy responses that did not help them address their conflicts. Additionally, these students often needed to provide more specific details in order to receive feedback that would highlight their knowledge gaps. This flexibility makes generative chatbots less effective for their needs. In terms of results, the SRLbot, with its generative capabilities, motivated students to reflect on their learning in a way that resembled interaction with a real teacher (Ng et al., 2024).

2.7.2 Empirical insights into ChatGPT’s educational applications

Recent studies have highlighted the potential of generative AI-based chatbots, such as ChatGPT, to enhance educational practices through adaptive feedback and flexible interaction (Ng et al., 2024; OpenAI, 2022). In various instructional contexts, ChatGPT has been employed as a teaching assistant and a virtual learning companion, supporting tasks such as lesson planning and student engagement (Farrokhnia et al., 2023; Okulu & Muslu, 2024). Despite these advantages, research has also identified several challenges. These include the tendency of ChatGPT to provide decontextualized responses and to function more as an evaluator than a facilitator of deeper thinking, particularly in tasks requiring reasoning and metacognitive engagement (Su et al., 2023).

In the qualitative study by Okulu and Muslu (2024), ChatGPT was employed as an assistant in developing lesson plans for pre-service science teachers. The findings indicate that ChatGPT facilitates the creation of practical course plans and provides adaptable information, ultimately saving time. Although there were instances of miscommunication, the researchers found the tool quite effective in the course planning process (Okulu & Muslu, 2024).

When assessing the overall effectiveness of ChatGPT, a SWOT analysis conducted by Farrokhnia et al. (2023) highlighted its advantages and disadvantages. The analysis revealed several benefits, including providing plausible answers, a capacity for self-improvement, and facilitating individualized and simultaneous responses. This allows teachers to streamline trivial tasks that consume valuable time, enabling students to distill information from vast sources (Farrokhnia et al., 2023).

However, despite its potential benefits, ChatGPT also presents several notable limitations. One major concern is its inability to demonstrate deep conceptual understanding or higher order thinking skills, such as analysis, synthesis, or critical evaluation. This limitation is closely tied to the challenge of evaluating the quality or suitability of its own responses. The decontextualization of answers can result in responses that fail to adequately address specific objectives. Furthermore, there are concerns about biased and discriminatory expressions in their outputs. Lastly, ChatGPT also poses a risk of plagiarism, as it may directly reproduce text from existing studies (Farrokhnia et al., 2023).

Similarly, Zhang and Tur (2023) conducted a systematic review utilizing a SWOT analysis to examine the use of ChatGPT in K-12 education. Their results indicate that ChatGPT can support personalized learning and enhance student motivation and engagement by addressing language barriers. Conversely, they also identified limitations, including low response quality, task specificity, and a lack of reasoning abilities (Zhang & Tur, 2023).

A systematic review conducted by Lo et al. (2024) explored the impact of ChatGPT usage on student engagement by examining various components, including behavioral, emotional, and cognitive aspects. The review analyzed seventy-two empirical studies that utilized ChatGPT. From a behavioral perspective, the findings indicated that students actively engaged with ChatGPT during learning activities, benefiting from its swift responses and guidance. However, the review also noted that due to a lack of digital literacy, some students are at risk of committing plagiarism (Lo et al., 2024).

Regarding emotional engagement, ChatGPT creates an environment where students feel comfortable asking questions and sharing their problems to receive advice, thereby allowing them to gain different viewpoints (Lo et al., 2024). Cognitively, students exhibited an increase in learning outcomes, and personalized responses from ChatGPT fostered self-regulation. Nevertheless, the tendency for students to rely solely on ChatGPT's answers without further investigation could impair their critical thinking skills. The researchers recommended that future studies focus on students' self-regulation (Lo et al., 2024).

In a study conducted by Urban et al. (2024), the effects of ChatGPT on self-regulation and self-efficacy were examined through a complex creative problem-solving task. The study involved 77 university students in the treatment group, who utilized ChatGPT to solve problems, and 68 students in the control group, who worked without it. The findings indicated that ChatGPT significantly enhanced students' self-efficacy in task resolution and improved the quality, elaboration, and originality of their solutions. Participants in the treatment group reported finding the tasks easier and expended less mental effort than those in the control group. The researchers proposed viewing ChatGPT as a provider of metacognitive cues rather than as a mere assistance tool (Urban et al., 2024).

In a separate exploratory study, Kim et al. (2024) investigated the impact of ChatGPT on the writing outcomes of undergraduate engineering students tasked with composing lab reports. The seven participants rewrote their lab reports using ChatGPT after familiarizing themselves with the tool. The study concluded that ChatGPT could effectively serve as a revision tool, enhancing the quality of lab reports and fostering a better understanding of the lab report genre among engineering students (Kim et al., 2024).

2.7.3 Challenges in the educational use of ChatGPT

However, several studies have reported limitations in ChatGPT’s ability to support reasoning and contextual understanding, which has led to miscommunication in educational settings. Su et al. (2023) investigated the applicability of ChatGPT as a tool for evaluating argumentative writing. Their study focused on how ChatGPT assesses students’ contributions concerning argumentative discourse's linguistic, structural, and dialogic dimensions. The findings indicated that while the argumentation structure was effectively rendered in GPT's responses, the model was insufficient in evaluating the quality of the arguments. ChatGPT exhibited teacher-like characteristics when provided with a rubric; however, the feedback offered was limited and necessitated further elaboration. If users did not pose questions regarding the feedback, the dialogic aspect of the argumentation faltered. Furthermore, ChatGPT lacked the capacity to ask guiding questions, leading to a tendency to deliver evaluative statements rather than fostering critical thinking among students (Su et al., 2023).

Conversely, from the user's perspective, ChatGPT was perceived as an effective educational tool. Research conducted by Pasca and Arcese (2024) explored user perceptions of ChatGPT concerning its opportunities and challenges within educational contexts. Interviews revealed that users regarded ChatGPT as a resource that simplifies tasks and enhances efficiency. Nevertheless, participants expressed concerns about the reliability of the information provided and the non-empathic nature of ChatGPT's responses (Pasca & Arcese, 2024).

Despite the potential of ChatGPT to offer metacognitive scaffolding, it fell short in generating responses that aligned with the recipient's comprehension level. In a study by Urhan et al. (2024), the participant was an undergraduate student majoring in mathematics education. The researchers examined the use of ChatGPT as a discourse tool within a calculus framework. The participant initially solved a mathematical problem on paper, subsequently articulated the solution verbally to the researchers, and then engaged in dialogue with ChatGPT. Following the conversation, the participant revisited the initial written work. The researchers primarily focused on the warrant component of the discourse. Findings indicated that ChatGPT was unable to provide consistent and reliable support for the participant's reasoning process. Additionally, the student encountered challenges in understanding ChatGPT's responses due to contextual disparities. Consequently, the model's communication style hindered the student’s ability to benefit from the interactions (Urhan et al., 2024).

2.7.4 Empirical studies done with the POE AI platform

Recent advancements in AI have enabled the development of chatbots capable of generating natural language conversations by drawing on specific knowledge sources without requiring programming expertise. While many large language models (LLMs)-based and task-oriented (TOD) chatbot platforms provide such functionality, they are often subscription-based and inaccessible to the broader public. In contrast, the POE AI platform is freely available and allows users to build customized chatbots by integrating various models and tools (Lee et al., 2024).

Given the researcher’s lack of programming experience, POE was selected for its user-friendly interface and ability to generate pedagogically appropriate dialogues in specific educational contexts. Through POE, the researcher could design a context-specific, teacher-like generative chatbot by tailoring its prompts and operational rules. In a quasi-experimental study by Lee et al. (2024), high school students perceived POE-based chatbots as useful tools in foreign language education. They reported high levels of motivation while interacting with them.

Similarly, in a study conducted by Wu and Xu (2025) involving 519 Japanese upper-intermediate EFL learners, the results indicated that the POE and Gemini chatbots significantly enhanced students' writing skills. Moreover, participants reported a boost in academic self-efficacy and expressed that they felt less boredom during writing tasks. Qualitative findings also revealed increased engagement, autonomy, and motivation among students in their English language learning (Wu & Xu, 2025).

Although these studies highlight the promise of POE in language learning, there is currently no research that explores its integration into scientific argumentation contexts. The present study addresses this gap by employing POE to develop a generative AI chatbot tailored to scaffold students' written scientific arguments.

Unlike traditional automated feedback systems, which often lack interactivity, or rule-based chatbots that restrict responses to predefined prompts, generative AI chatbots such as ChatGPT allow students to pose open-ended questions and receive adaptive feedback (Farrokhnia et al., 2023; Lo et al., 2024; Ng et al., 2024; OpenAI, 2022). However, existing findings also point to limitations, such as decontextualized or overly complex responses that may not align with students' cognitive levels (Su et al., 2023). To mitigate these concerns, the researcher designed the POE-based chatbot using carefully crafted prompts to ensure that responses remain contextually relevant and pedagogically appropriate for middle school learners.

2.8 Summary

The amount of greenhouse gases has reached a critical level, resulting in irreversible consequences; therefore, there is a need for public action that changes the existing situation by keeping the amount of gases at a certain level (Jay et al., 2023). However, creating a public action necessitates education about socio-scientific issues, including the greenhouse effect and global warming (Gautier, 2012). While educating about socio-scientific issues, students can gain scientific literacy, which encompasses the evaluation of evidence and relates this evidence to scientific explanation and understanding of scientific processes. To make students who can evaluate evidence and support their claims considering scientific explanations, scientific argumentation is an effective way and can be taught through practice rather than instruction (Driver et al., 2000; Lieskovský & Sunyík, 2022). Scientific argumentation practice needs inquiry and discussion environments where students collect and evaluate data obtained and discuss their findings with others while forming explanations; therefore, students can experience scientific processes vital for scientific literacy (Bächtold et al.,2022; Hendratmoko et al., 2023). Since the stages of the argument-driven inquiry model consist of discussion and inquiry activities, the study implemented a learning module designed based on the ADI framework (Sampson et al., 2010).

According to the literature, during scientific argumentation, several challenges may occur, stemming from the absence of immediate feedback and the need for instructional support (Berland & McNeill, 2010; Henderson et al., 2017; Wen et al., 2015; Walker & Lower, 2019). While most studies emphasize the support needs of low-achieving students, high-achieving students also require instructional guidance, especially in constructing deeper reasoning, developing counterarguments, and exploring alternative perspectives (Ho et al., 2019; Kızkapan & Nacaroğlu, 2020). To eradicate these challenges during scientific argumentation, the literature suggests using chatbots and automated scoring systems (Chang et al., 2023; Zhu et al., 2020). Even though auto-scoring systems provide immediate feedback, they lack interaction and prohibit students from asking further questions about the topic (Chang et al., 2023). Regarding chatbots, rule-based ones restrict students to a specific pathway, which may inhibit students' efforts to close the knowledge gap or ask curiosity questions (Ng et al., 2024). On the other hand, when generative AI-based chatbots are used, students can ask more questions and get detailed feedback based on their individualized product (Lo et al., 2024).

Despite the advantages of chatbots in education, challenges such as decontextualization and overly complex responses have been noted, particularly with ChatGPT (Urhan et al., 2024). Moreover, many recently developed chatbots using generative AI lack pedagogical alignment, requiring specific prompting to respond in teacher-appropriate language using targeted knowledge sources (Urhan et al., 2024). Recent advancements, such as the release of POE, have introduced chatbot interfaces that facilitate natural language conversations by drawing on specific knowledge sources without programming, with some studies exploring their use in foreign language education (Lee et al., 2024; Wu & Xu, 2025). Despite these advancements, no research exists on teaching the greenhouse effect through scientific argumentation with chatbot integration at the middle school level.

This study aims to explore how a learning module designed based on the ADI instructional model with generative AI integration facilitates sixth-grade students' written scientific argumentation quality in the greenhouse effect topic, the nature of their interactions with a chatbot, and reveal their perceptions on the learning experience.

CHAPTER 3

METHOD

This study aims to explore how a learning module designed based on the Argument-Driven Inquiry (ADI) instructional model with generative AI integration enhances sixth-grade students' written scientific argument quality in the context of the greenhouse effect topic and their perceptions of this experience. This section consists of the research design, participants, and context of the study, as well as data sources, data collection, data analysis, and trustworthiness.

3.1 Research design

A case study is a type of qualitative research design that aims to illuminate a phenomenon, benefiting from an inductive approach. It aims to reach a final product that is richly descriptive (Merriam & Tisdell, 2015). Besides, a case study also consists of a bounded system as a unit of analysis, necessitating uniqueness. Therefore, it makes it obligatory to conduct the study in a specific context and with participants. Therefore, a case study is an in-depth analysis of a bounded system (Merriam, 2009). Since this study aims to understand and explore the interaction between students and the chatbot during the scientific argument formation process in a classroom setting and students' understanding of this experience, its nature overlaps with the qualitative case study research design. Furthermore, the unit of analysis consists of a bounded system in a bounded time interval, for a specific situation; therefore, this study was designed as a case study (Merriam & Tisdell, 2015).

3.2 The case

The study took place after regular school hours at a public middle school in Mersin, Turkey, which serves economically disadvantaged students and lacks a computer lab, during the fall semester of the 2024-2025 academic year. This study aimed to explore how a learning module designed based on the ADI framework with generative AI integration facilitates improvement in sixth-grade students' quality of scientific arguments in the context of the greenhouse effect and their perceptions of this learning experience.

The study involved five sixth-grade students, selected through purposeful sampling to determine participants, since purposive sampling provides in-depth exploration of the research context (Gay et al., 2011). Five students were selected purposively based on their achievement levels in science classes. Consequently, five high achievers were selected, representing the top five students in terms of academic achievement. The small sample size allowed for exploring students' interactions with a chatbot. All participants had previously participated in a pilot study, conducted with a group of fourteen sixth-grade students, which introduced them to the concept of scientific argumentation and the use of chatbot technologies, such as ChatGPT, for constructing arguments in the greenhouse effect context. Thus, students became familiar with the study's expectations and tools. Based on the insights gained from a pilot study with fourteen students, which revealed inefficiencies in facilitating group work with the chatbot, the main study shifted focus to individual engagement with the module, specifically emphasizing the written argumentation phase of the ADI model.

The study focused on sixth-grade students as the primary participants due to both curricular relevance and developmental appropriateness. While the term "greenhouse effect" is not explicitly stated in the sixth-grade science learning objectives, the curriculum consists of concepts such as the disadvantages of fossil fuel use and the promotion of renewable energy sources. These topics implicitly introduce students to the broader idea of human-induced environmental change, thereby offering a meaningful context to explore climate-related scientific argumentation. Additionally, their prior knowledge of temperature, heat, and energy forms a conceptual foundation that can be extended toward understanding more complex phenomena such as the greenhouse effect. Additionally, fostering environmental awareness is a key objective of the national curriculum. Integrating this goal with instructional practices that emphasize evidence-based argumentation offers an effective means to support students' cognitive and ethical development concerning environmental issues.

The study was conducted in a classroom setting after regular lesson hours. The researcher, also serving as the science teacher, provided the necessary technological devices to the participants using her personal resources. Each participant engaged with the learning module individually. Following the completion of the simulation, observations, and text reading activities, participants informally shared and discussed their findings with one another while remaining seated in the classroom setting. Due to contextual limitations, including inefficiencies in chatbot interactions observed during the pilot study, the first five stages of the ADI model were implemented, necessitating individual participant engagement.

Consequently, the study focused exclusively on the written argumentation phase of the Argument-Driven Inquiry (ADI) model. The primary aim was to explore the development of written argument quality and to analyze participants' interactions with the chatbot in an individual context.

Throughout the process, the researcher provided guidance by informing participants about each phase, addressing their questions, and offering technological support as needed. Additionally, the researcher monitored the chatbot outputs to identify and address any issues that arose during the conversations. Besides, she posed guiding questions to the participants during the implementation of the learning module to support the learning process.

3.2.1 Participants

The researcher, having attended the participants' lessons for two years as their science teacher, was well-acquainted with their competencies and academic backgrounds. In selecting participants, the researcher prioritized academic achievement as the primary criterion. The academic achievement was significant because scientific argumentation in the greenhouse context is related to other scientific concepts learned in previous years. This section provides general information about the characteristics of the participants, as observed by the researcher in her role as their teacher over two years.

Ruby

Ruby ranks among the top five achievers in her class. However, while she demonstrates commendable academic performance, her success does not equate to that of her peers who excel at a higher level. This observation is supported by her academic records and the insights provided by her teachers. She generally exhibits inconsistent attendance in class and sometimes responds to questions posed by the teacher in the classroom. Consequently, it was strongly anticipated that she would require assistance from the researcher to engage with the chatbot during the study.

Angelia

Angelia, a high-achieving student in science lessons and has consistently demonstrated active participation in science lessons. Her curiosity about science-related topics and enthusiastic approach to learning new information were evident throughout the previous class sessions before the module. Teachers who participated in Angelia’s lessons shared that she demonstrates self-regulated learning by posing insightful questions across multiple subjects. This behavior reflects her capacity to take initiative in her own learning process. Given this background, it was anticipated that she would effectively address chatbot-related challenges by posing questions, despite lacking prior experience with generative

Emily

Emily has been identified as a high-achieving student across all subjects, based on her academic performance and observations by both teachers and the researcher over a two-year period. Emily tends not to pose questions in class; however, she consistently demonstrates a keen interest in responding to questions posed by the teacher. Given this background, it was anticipated that she would require teacher assistance to effectively interact with the chatbot during the study.

Rose

Rose, the highest-achieving student among the sixth graders with the top academic average, has consistently engaged actively in classroom activities across all subjects, as evidenced by her academic record and teachers’ observations. Rose tends not to pose questions in class but demonstrates eagerness to respond to questions posed by the teacher.

Micheal

Micheal has been classified as a high-achieving student, based on academic records and teachers' observations. Over a two-year period, the researcher observed that Micheal exhibits strong motivation and curiosity, particularly in science topics, often asking probing questions to deepen his understanding. Consequently, it was anticipated that he would actively engage with the chatbot by posing questions to explore the topics further.

3.2.2 Pilot study

In the pilot implementation, a total of fifteen sixth-grade students received explicit instruction on the nature of scientific argumentation and how to interact effectively with a generative AI chatbot. To reinforce this foundational training, the same group of students participated in a structured learning module developed in alignment with the Argument-Driven Inquiry (ADI) framework. The module aimed to facilitate students' understanding of the greenhouse effect through the practice of scientific argumentation and guided chatbot interaction. The module comprised four ADI-based learning activities: a greenhouse effect investigation, a simulation-based exploration, a peer feedback session on group-generated arguments, and a guided interaction with a generative AI chatbot. All activities were integrated into regular lesson hours and conducted over a two-week period, with students working in collaborative groups of five. Prior to the module, students were introduced to the objectives and structure of the learning activities and received targeted training on both scientific argumentation and the operational use of the chatbot. Each activity was designed to last approximately 40 minutes. Table 2 outlines the timeline of the pilot implementation.

Table 2. Timetable for Pilot Study

Time Interval	Activity	Duration
Session 1	Scientific Argumentation Formation Practice with Generative AI (Training)	40 minutes
Session 2	Greenhouse Activity and Scientific Argument Formation with the chatbot	40 minutes
Session 3	Scientific Argument Formation with Greenhouse Effect Simulation and Revisions of Groups' Scientific Arguments based on Classmates' and the Chatbot Feedback	80 minutes

3.2.2.1 Changes in the design of the module based on the pilot study

Following the implementation of the pilot study, several issues were identified that necessitated revisions to the instructional design. Accordingly, the ADI framework was modified, focusing on certain steps in response to challenges observed during chatbot interactions. Additionally, the chatbot prompts were refined, and the chatbot's responses were further constrained to improve relevance and coherence. This section explains the reasons for these changes based on the outcomes observed.

3.2.2.2 Change in the instructional model

The pilot study revealed several implications for refining the instructional model. It was observed that not all stages of the Argument-Driven Inquiry (ADI) instructional model functioned effectively within the given classroom context. Specifically, students exhibited reluctance to evaluate their peers' work and lacked the necessary skills to provide constructive peer feedback. Consequently, "identification of the task", "generation of data", "production of a tentative argument", "argumentation", and "creation of a written investigation report" stages of the ADI framework were implemented.

In addition to these issues, the implementation of group work presented significant challenges in terms of equitable access to chatbot interaction. Due to limitations in technological resources, students were required to share a single computer within groups of five. As a result, high-achieving students often dominated the interaction with the chatbot, leading to an unbalanced distribution of participation. To address this limitation and ensure equitable engagement, individual chatbot interactions were adopted for the main study.

Thus, the ADI framework was only partially implemented. The revised module included the following stages: identification of task, generation of data, production of tentative argument, argumentation, and creation of investigation report. The argumentation stage was implemented via a chatbot rather than group work. Each student shared their initial argument and got feedback on it. Furthermore, the pilot study highlighted students' difficulty in constructing arguments based solely on numerical data without additional contextual support. To overcome this issue, text-based materials were provided during the main study to serve as a knowledge source, enabling students to better engage with the content and construct informed arguments.

3.2.2.3 Changes in the design of the chatbot

In the pilot study, student interactions with the chatbot revealed that it did not provide specific feedback, and the responses tended to be overly lengthy. As a result, the prompts were revised. To enhance the specificity of the responses, feedback prompts developed by Zhu et al. (2017) were incorporated (see Appendix A for the revised written prompts). Additionally, feedback sentences were introduced to ensure consistent responses in similar situations.

3.2.2.4 The final design of the chatbot

In this study, the POE interface was favored due to its capabilities for integrating various generative AI models and enabling the upload of content-specific documents without requiring any programming expertise (Lee et al., 2024). Moreover, it is freely accessible for public use. These features facilitated the design of a chatbot that was both pedagogically sound and tailored to the cognitive and language levels of middle school students. Consequently, the chatbot was developed using the GPT-4o model to assist sixth-grade students in constructing scientific arguments regarding the greenhouse effect.

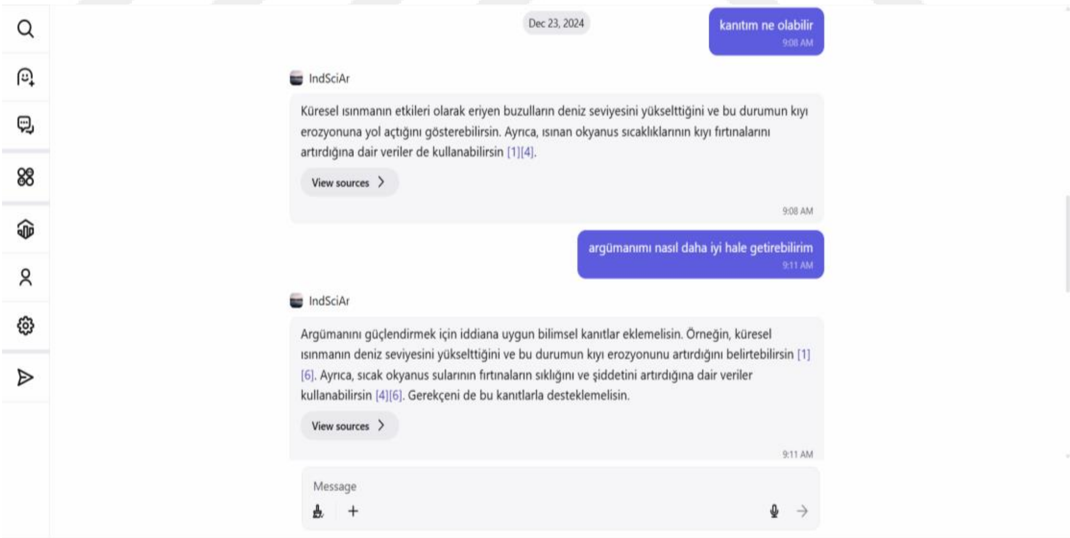


Figure 1. Participant interface in POE

To ensure that the chatbot provided contextually appropriate responses aligned with middle school students' cognitive and linguistic levels, it was configured using a content-specific knowledge base focused on the greenhouse effect (see Appendix B).

In addition to this knowledge source, detailed prompts were given to guide the chatbot's use of scientific language, control response length, and define its instructional role. For example, the chatbot was assigned the persona of a "science teacher with experience teaching middle school students" to create a comfortable and supportive environment for learners. Additionally, the CER (Claim, Evidence, and Reasoning) framework proposed by McNeill et al. (2006) was explained to the chatbot, accompanied by evaluation rubrics provided in Appendix C.

A pilot study was conducted to create an environment that familiarizes students with generative AI. In this study, students worked in groups and collectively discussed their experiences, enabling them to engage in additional learning activities involving AI (Heeg & Avraamidou, 2024). Additionally, the pilot study facilitated the identification of challenges that participants and researchers may encounter. This insight allowed for the subsequent redesign of the chatbot to address and mitigate these challenges effectively. Observations from the pilot study indicated that the chatbot delivered responses in a cumulative manner rather than addressing each argument component individually, which caused confusion among students. Additionally, the responses were often too lengthy, despite the chatbot being instructed to limit feedback to no more than 30 words.

To address these issues and ensure consistency in responses, feedback prompts were adapted from the work of Zhu et al. (2017). For example, if evidence is present but lacks clear reasoning, the chatbot is prompted to ask, "You provided evidence for your claim. Can you explain how or why this evidence supports your claim?" Conversely, when reasoning is supplied without accompanying evidence, the prompt would be, "You've provided scientific reasoning to support your claim. Can you include evidence to back up your reasoning?" Before implementing these changes, the researcher interacted with the chatbot to monitor and refine its responses (see Appendix D for an example of a conversation conducted by the researcher before and after the pilot study).

Overall, the researcher considered several criteria in the chatbot's design, including pedagogical appropriateness, compatibility with the students' level, structured feedback, response length, sources of knowledge, and the teaching of evaluation criteria.

3.3 Data sources

The qualitative data were collected through students' worksheets on scientific argumentation templates, reasons for chatbot response selection, and perceptions of chatbot experience and chatbot conversation output.

3.3.1 Scientific argument template

The template was initially created for collaborative work during the tentative argument phase of the ADI teaching model developed by Sampson and Schleigh (2013). This template was chosen for its ability to foster connections among the components of arguments and its compatibility with McNeill's scientific argumentation framework. For this study, the template was enhanced by incorporating guiding questions to support students in constructing their scientific arguments. Before the application, the researcher consulted with an expert to utilize the template to develop scientific argumentation. During the learning module, participants engaged with the text and responded to guiding questions, which allowed them to form their initial arguments using the scientific argument template (see Appendix E). They also refined their arguments by leveraging the same template. This template includes components for claims, evidence, and reasoning, with students expected to find answers to the guiding questions for each component. To evaluate participants' arguments, their initial and revised arguments were analyzed using the rubric developed by McNeill et al. (2006). In this rubric, participants' argument components can be assigned levels 0 to 2. In each session, students formulated two arguments: an initial argument and a revised argument.

3.3.2 Chatbot conversations

The chatbot was configured using a content-specific knowledge source (see Appendix B) prepared by the researcher and was guided with prompts (see Appendix A) to generate pedagogically appropriate feedback in language suitable for middle school students. Following the construction of their initial arguments, participants engaged with the chatbot. They were provided with example prompts to guide their interactions, which they adapted based on the nature of their questions. They were also encouraged to pose their own questions, provided they remained within the context of the greenhouse effect. Accordingly, participants' entries were analyzed qualitatively by their purpose (e.g., asking for guidance or feedback), and the chatbot's responses were evaluated in terms of the function they served and their contextual appropriateness. Additionally, the quality of the chatbot's responses was evaluated.

3.3.3 Perception form

Participants' perceptions of the chatbot's effectiveness in supporting their scientific argumentation were explored through seven open-ended questions, which aimed to explore participants' chatbot experience during the scientific argument formation process administered at the end of each session in perception forms (see Appendix F). For example, the participants were asked the following question: “When you corresponded with a chatbot, did the feedback and answers make it easier for you to write your arguments?” These questions, adapted from Chang et al. (2023), were tailored to the specific context of the learning module and aimed to capture students’ reflections on their interaction with the chatbot and their engagement in argument construction. The responses provided insights into individual experiences and helped reveal how students interpreted and evaluated the chatbot’s role in their argument formation process.

3.4 Data collection

At the outset of the research, ethical approval was obtained from the Ethics Committee in Social Sciences and Humanities (SOBETIK) of Boğaziçi University (see Appendix G). Following this, the necessary permissions were secured from the Ministry of National Education (MoNE) (see Appendix H). The learning module titled Scientific Argumentation via Generative AI was introduced to a sixth-grade class in a public middle school. Volunteer students who were identified as high achievers were selected to participate in the study. Teachers and school administrators were informed about the research objectives, and informed consent was obtained from both students and their parents (Appendices I and J).

The study was conducted in a public school setting. Prior to implementation, a pilot study was carried out with fourteen sixth-grade students to explore how students interacted with the chatbot and to identify areas for improvement. Given the lack of existing research on the integration of generative AI in teaching scientific argumentation within the context of the greenhouse effect, this pilot was critical for refining the research design. Based on observations from the pilot, necessary revisions were made to the module structure and chatbot prompts.

The main study, consisting of three sessions, was conducted with five sixth-grade students. Each session followed the same instructional flow and lasted approximately 120 minutes. During the sessions, participants explored the consequences of the greenhouse effect, including rising global temperatures, melting ice, and sea level rise. Based on their investigations, they constructed scientific arguments.

Each session began with participants exploring a digital simulation and reading a short text related to a specific consequence of the greenhouse effect. They then constructed an initial written scientific argument. Next, participants were provided with a list of example prompts to guide their chatbot interactions. Using these prompts or their own inquiries, they engaged in a conversation with the generative AI-based chatbot. After the chatbot interaction, participants revised their scientific arguments based on the feedback. At the end of each session, students also responded to a set of open-ended questions designed to elicit their perceptions of the learning module. The overall structure of the three-week module is presented in Table 3.

Table 3. General Plan of the Learning Module

Time	Related ADI Stage	Activity	Duration (minutes)	Data
	Identification of Task	Introducing the Task	10	No data
16.12.24 (Session 1- Effect on Temperature Rise)	The Generation of Data	Simulation Interaction with WS	15	No data
23.12.24 (Session 2- Effect on Melting Ice)	The Generation of Data	Text Reading and WS	15	No Data
	Formation of Tentative Arguments	Scientific Argument Formation	25	Scientific Argument Quality
30.12.24 (Session 3- Effect on Sea Level Rise)	Argumentation	Chatbot interaction	20	Participants' Interaction Purposes, Chatbot's Functions, Chatbot's Response Quality
	Creation of Investigation Report	Second Scientific Argument Formation	25	Scientific Argument Quality
		Perceptions about Learning Module	10	Perceptions about the Chatbot Experience

3.4.1 Implementation of Session 1- Results of the greenhouse effect on temperature rise

At the beginning of the session, the researcher informed the participants about the structure and objectives of the study and initiated a brainstorming activity to activate prior knowledge regarding the greenhouse effect. Subsequently, the implementation of the learning module (see Appendix K) was initiated. Technological devices, including laptops and tablets, were distributed to ensure that each participant had individual access to a device. Participants were instructed to explore NASA's "Climate Time Machine" simulation, manipulating time intervals to observe global temperature changes over the years. They recorded their observations on a worksheet and responded to guiding questions designed to support data interpretation. Following the investigation, participants shared and discussed their findings with the group.

Next, they were provided with a text containing scientific explanations related to the impact of the greenhouse effect on temperature rise. Participants read the text and completed the accompanying worksheet. Drawing on data from both the simulation and the reading material, they constructed their initial scientific arguments using a structured template.

Following this, participants accessed the generative AI-based chatbot interface via the researcher's POE account on their devices. The chatbot had been trained by the researcher to simulate a middle school science teacher by drawing on the provided knowledge source, ensuring responses remained aligned with the scientific context.

Before beginning their interaction, participants received a set of suggested prompts (see Appendix L) to guide their conversation; however, they were also required to pose at least two original questions of their own, in addition to using the given prompts (see Appendix M for participant interface). During the interaction, the researcher reminded participants to critically reflect on which chatbot responses influenced their thinking and potential revisions.

After completing their interactions with the chatbot, participants revised their initial arguments, resulting in the development of second versions of their scientific arguments. As a final step, they responded to open-ended perception questions (see Appendix F) to reflect on their experiences with the chatbot and the learning module.

3.4.2 Implementation of Session 2- Effect on melting ice

In Session 2, the same instructional procedure was followed as in the previous session. However, the focus of this session shifted to the examination of ice sheet melting. Participants interacted with a simulation illustrating changes in global ice coverage over time and recorded their observations. Subsequently, they were provided with a text containing relevant information, including potential claim, evidence, and reasoning components related to the effects of global warming on ice sheet melting. Following the reading activity, the same steps as in Session 1 were implemented, including initial argument construction and their revisions, chatbot interaction, and perceptions.

3.4.3 Implementation of Session 3- Effect on sea level rise

In the third session, the same instructional structure and materials used in the previous sessions were applied. To initiate the activity, participants were prompted to establish a connection between the melting of ice sheets and the resulting rise in sea levels. Subsequently, they accessed the Sea Level Risk Map simulation and explored how sea level rise of varying magnitudes could affect city centers across different regions. Following the simulation, participants engaged in a discussion on the potential global consequences of rising sea levels.

After the discussion, a content-specific text was provided, which included information structured around possible claims, evidence, and reasoning (CER) components related to the impact of global warming on sea level rise. Participants read the text individually and responded to guiding questions in writing. The remaining instructional steps were conducted in accordance with the procedure established in the previous sessions, including initial argument construction, interaction with the chatbot, argument revision based on chatbot feedback, and completion of the perception form (see Appendix F).

3.5 Data analysis

This section analyzes the quality of participants' written scientific arguments, their interactions with the chatbot, and their responses on perception forms regarding the chatbot experience.

With the scope of this study, the researcher collected three separate data sets to explore changes in the quality of scientific arguments and interactions with chatbots. Since data on argument quality and students' perceptions were printed material, the researcher transferred these data to digital text material and translated them into English.

This study aims to explore a phenomenon that has not been previously investigated, leading to the use of inductive thematic analysis since there were no pre-determined codes available. Data-driven codes and categories began to emerge during the analysis process (Fereday & Muir-Cochrane, 2006). The researcher utilized Excel for open coding to develop codes and assign them accordingly. The analysis was guided by the thematic analysis method, which involved identifying, analyzing, and interpreting meaningful patterns in the collected data (Clarke & Braun, 2016). As codes were developed and linked to relevant themes to enhance understanding of the chatbot experience, the steps of the inductive approach were strictly followed (Fereday & Muir-Cochrane, 2006). Throughout the coding process, the researcher consulted with an expert in the field to ensure rigor.

In this study, four distinct data sets were gathered to address the research questions. The methods of data collection and analysis are detailed in Table 4, presented as a data collection matrix.

Table 4. Data Collection Matrix

Main Research Questions	Data Collection	Data Analysis
1. How does the quality of the scientific arguments written by sixth-grade students in the context of the greenhouse effect improve during their interaction with the learning module?	Participants' Scientific Argument Templates	Scoring based on McNeill et al. (2006)
2. How do students interact with a specifically designed chatbot to develop their arguments during the learning module? i. For which purposes do students use the chatbot in the context of the scientific argument formation process? ii. Through which functions does the chatbot respond to students' requests in the context of the scientific argument formation process? iii. What is the quality of chatbot responses in facilitating scientific argumentation?	Chatbot Conversation Logs	Thematic Analysis -Open Coding
3. What are the students' perceptions of the effectiveness of the chatbot for developing their arguments?	Participants' Perceptions via Perception Form	Thematic Analysis- Open Coding

3.5.1 The analysis of the first research question

In order to analyze the collected data for the first research question (How does the quality of the scientific arguments written by sixth-grade students in the context of the greenhouse effect improve during their interaction with the learning module?) participants' scientific arguments were analyzed based on the rubric developed by McNeill et al. (2006).

The Claim-Evidence-Reasoning (CER) framework was preferred, as it is considered developmentally appropriate for the target age group. Although other argumentation frameworks, such as Toulmin's model, incorporate components like rebuttals and counter-arguments, these were not emphasized in the present study due to their higher cognitive demands, which may not align with the developmental level of sixth-grade participants. Furthermore, interaction with the chatbot during the pilot study revealed that the chatbot did not generate guiding questions prompting the construction of rebuttals or counter-arguments. Instead, it predominantly provided information and evaluative feedback. As the development of rebuttals requires more detailed and targeted prompting, and such scaffolds were not embedded in the current study design, the focus remained on the core CER components.

Scientific arguments from participants were gathered at two points in each phase, allowing the researcher to compare the quality before and after the interaction. Changes in component levels were also tracked individually to identify which aspects of argumentation (e.g., reasoning vs. evidence) showed the most development. The arguments were categorized as Level 0, Level 1, and Level 2 based on the argument components, which included claim, evidence, and reasoning utilizing the rubric developed by McNeill et al. (2006). Instances where participants confused these components were labeled as Level 0. Additionally, even if a participant formed a claim that was categorized as Level 0, other components could still be assigned to Level 2 if the other components were sufficiently robust on their own (see Appendix C). To ensure consistency, an expert independently evaluated a randomly selected sample of arguments. The results were discussed to resolve any discrepancies and reach a common understanding.

3.5.2 The analysis of the second research question

To analyze the data pertaining to the second research question (How do students interact with a specifically designed chatbot to develop their arguments during the learning module?), three sub-questions were constructed, and data were analyzed accordingly. These sub-questions are “For which purposes do students use the chatbot in the context of the scientific argument formation process?”, “Through which functions do the chatbot respond to students' requests in the context of the scientific argument formation process?” and “What is the quality of chatbot responses in facilitating scientific argumentation?”

In order to explore participants’ interactions with the chatbot, conversation logs were exported to Excel and analyzed using open coding. The researcher generated codes focusing on three dimensions: the purpose of interaction, the function of the chatbot, and the quality of its responses. Thematic analysis was conducted following the steps outlined by Braun and Clarke (2006).

Accordingly, the researcher familiarized herself with the data, generated initial codes, organized them into relevant themes, and reviewed the coherence of these themes. As a result, three overarching dimensions were identified and categorized into 16 specific categories. Table 5 presents the distribution of these categories across the identified dimensions.

Table 5. The Distribution of Categories in Different Aspects of Chatbot Conversation Logs

Aspects of Chatbot Interaction	Categories
Participants' Purposes	1. Asking for guidance 2. Asking for feedback 3. Asking for general information 4. Requesting revision of the given response 5. Asking questions out of curiosity 6. Asking for summarizing
Chatbot's Function	1. Providing guidance 2. Providing feedback 3. General information for inquiry 4. Revising the response 5. Answering the curiosity questions 6. Providing summary
Chatbot's Response Quality	1. Suitable for context and correct 2. Out of context 3. Complex response integrating more than one component 4. Lack of feedback 5. Confusion among the elements of arguments

3.5.2.1 Analysis of participants' purposes for interacting

Participants' purposes for interacting were defined as their reasons to interact with the chatbot while constructing their scientific arguments. To investigate this dimension, the researcher addressed the following sub-research question: "For which purposes do students use the chatbot in the context of the scientific argument formation process?" Their conversation logs were systematically analyzed to explore participants' purposes for using the chatbot. The researcher reviewed the logs multiple times to identify the reasons behind participants' interactions with the chatbot. Open coding was applied to the conversation logs. During initial coding, participants' prompts were labeled according to their intended purpose, such as asking for feedback on a claim or requesting guidance for constructing reasoning. These initial codes were then clustered into broader categories representing common interaction purposes, such as “asking for feedback” or “asking for guidance”. In the theme review phase, overlapping or closely related categories were merged; for instance, “requesting revision” and “disagreeing with the given feedback” were combined under a single category. The final set of categories representing students’ purposes for chatbot interaction emerged as follows: asking for guidance, asking for feedback, asking for general information, requesting revision of the given response, asking for summarizing, and asking questions out of curiosity.

According to the analysis, as shown in Table 6, ten sub-codes were identified. Of these ten sub-codes, five were derived from the asking guidance category and five from the asking feedback category. Each category is described below:

Table 6. The Distribution of Categories in Different Purposes of Chatbot Interaction

Aspect of Chatbot Interaction	Categories	Sub-Codes
Participants’ Purposes	Asking for Guidance for Argument	1. Asking for guidance for more than one component 2. Asking for guidance for the research question 3. Asking for guidance for the claim 4. Asking for guidance for evidence 5. Asking for guidance for reasoning
	Asking for Feedback for Argument	1. Asking for feedback for more than one component 2. Asking for feedback for the research question 3. Asking for feedback for the claim 4. Asking for feedback for evidence 5. Asking for feedback for reasoning
	Other	1. Asking for general information 2. Requesting a revision of the given response 3. Asking questions out of curiosity 4. Asking for summarizing

Asking for guidance for argument: The participant seeks assistance to improve their scientific argument. This includes asking how to strengthen the argument, presenting incomplete argument components for completion, or requesting alternative examples to formulate or restructure the argument.

For the “asking guidance for argument” category, there existed five codes: “asking guidance for more than one component,” “asking guidance for the research question”, “asking guidance for the claim”, “asking guidance for evidence,” and “asking guidance for reasoning.”

As an example from this category, the participant asked for guidance on the claim component by stating, “What could my claim be?”

Asking for feedback for an argument: The participant shares their existing argument with the chatbot to receive evaluative feedback, aiming to determine its adequacy and identify areas for revision.

The asking for feedback for argument category consisted of five codes: “asking feedback for more than one component,” “asking feedback for the research question,” “asking feedback for the claim”, “asking feedback for evidence”, and “asking feedback for reasoning.”

As an example from this category, the participant asked for feedback on a claim by stating: “The ice sheets in Greenland and Antarctica are massive ice formations that contain 70% of the Earth's freshwater. Is it a good claim?”

Asking for general information: The participant inquires about topic-related details that were not explicitly covered in the instructional text, reflecting a desire for extended understanding.

Requesting revision of the given response: The participant asks the chatbot to rewrite the response due to an error the chatbot made or disagreement with the given response.

Asking questions out of curiosity: The participant asks questions about the details of the topic that were not mentioned in the text.

Asking for summarizing: The participant requests the chatbot to formulate the response in concise language due to the long response.

3.5.2.2 Analysis of chatbot's function

The chatbot's function was defined as its ability to address the participants' requests, including providing guidance or feedback. To explore this, the following sub-research question was addressed: "Through which functions do the chatbot respond to students' requests in the context of the scientific argument formation process?".

The researcher analyzed the conversation logs between the participants and the chatbot by focusing on the function the chatbot served in each interaction. Using an open coding approach, each chatbot response was initially coded based on how it operated in relation to participants' prompts, for instance, providing guidance for constructing a claim or answering curiosity-driven questions. These initial codes were then clustered into broader categories that reflected the chatbot's overarching functional roles.

Subsequently, the emerging themes and codes were reviewed and compared across the entire dataset to ensure consistency and coherence. During this process, functionally similar codes were merged under more inclusive categories. For example, the codes "giving an example" and "providing guidance" were integrated into a single functional category. Once the refinement process was completed, each theme was clearly defined and named.

As a result, six main functional categories and ten sub-codes were identified. Of the ten sub-codes, five were associated with the "providing guidance" category, and the remaining five were grouped under "providing feedback." The distribution of categories is shown in Table 7.

Table 7. The Distribution of Categories in Chatbot's Function

Aspect of Chatbot Interaction	Categories	Sub-Codes
Chatbot's Function	Providing Guidance	1. Providing guidance for more than one component 2. Providing guidance for the research question 3. Providing guidance for the claim 4. Providing guidance for evidence 5. Providing guidance for reasoning
	Providing Feedback	1. Providing feedback for more than one component 2. Providing feedback for the research question 3. Providing feedback for the claim 4. Providing feedback for evidence 5. Providing feedback for reasoning
	Other	1. General Information for Inquiry 2. Revising the Response 3. Answering the Curiosity Questions 4. Providing Summary

A detailed description of each category and sub-code is presented below.

Providing Guidance: The chatbot interprets the participant's inquiry and offers example-based support tailored to specific argument components within the given context.

The providing guidance category included five codes: "providing guidance for more than one component", "providing guidance for the research question," "providing guidance for the claim”, “providing guidance for evidence,” and “providing guidance for reasoning.” An example of the chatbot providing guidance is its support for the evidence component, as shown in the following response:

> To strengthen your evidence, you can express it as follows:
> Over the years, sea levels have risen due to the annual loss of 150 billion tons of ice from Antarctica and 270 billion tons from Greenland. This melting accounted for about one-third of the global sea level rise observed between 2006 and 2015.
> This information can make your evidence more robust.
> (Session 2, Angelia)

Providing Feedback: The chatbot evaluates the participant’s argument based on the rubric criteria and delivers feedback using sentence structures pre-defined by the researcher. The purpose is to encourage revision rather than supply full argument components or direct examples.

The providing feedback category included five codes: “providing feedback for more than one component,” “providing feedback for research question”,” providing feedback for the claim”, “providing feedback for evidence,” and “providing feedback for reasoning”. An example from this category is the chatbot’s guidance on the reasoning component. For instance, the chatbot provided the following guidance to help participants construct reasoning:

> You can use the following statements for your reasoning:
> "Rising sea levels cause coastal settlements to be submerged and force people to relocate."
> "Rising seas lead to saltwater intrusion into inland freshwater sources, contaminating agricultural lands."
> "This situation may result in waterborne diseases that threaten human health and lead to social problems."
> These reasons provide logical explanations that support your claim.
> (Session 2, Angelia)

General information for inquiry: The chatbot answers questions derived from the provided instructional materials and knowledge base to support conceptual understanding.

Revising the response: The chatbot rewrites or modifies its previous output in response to the participant's dissatisfaction or correction request.

Answering the curiosity questions: The chatbot responds to inquiries beyond the instructional scope or knowledge source, reflecting students' spontaneous interest or desire to explore related topics.

Providing summary: The chatbot reformulates its prior response in a shorter and more concise manner upon request, usually due to concerns over response length or complexity.

3.5.2.3 Analysis of chatbot's response quality

The chatbot's response quality was defined as the compatibility and contextual relevance of the response with the participant's request. In order to explore the chatbot's response quality, the following research question was addressed: "What is the quality of chatbot responses in facilitating scientific argumentation?" The researcher analyzed participants' prompts and the corresponding chatbot responses to evaluate the appropriateness of the responses in relation to participants' requests. Each response was initially coded using an open coding approach, focusing on the relevance and accuracy of the chatbot's reply. The preliminary codes generated included: suitable for context, correct, out of context, and confusion of argument components. These codes were then clustered into broader categories to identify patterns in response quality. During the review phase, the categories were refined to ensure internal consistency. Similar codes were merged, for example, suitable for context and correct were combined under a single, more inclusive category. The final set of categories, representing recurring patterns in chatbot response quality, was identified as follows: suitable for context and correct, out of context, complex response integrating more than one component, lack of feedback, and confusion among elements of the argument.

The definitions of these categories are presented below:

Suitable for context and correct: The chatbot provides an accurate and contextually appropriate response aligned with the participant's inquiry. The response is based on the predefined knowledge source and is relevant to the instructional objectives.

Out of context: The chatbot responds prematurely or inaccurately, such as when it generates an answer before the participant completes their input or fails to recognize the intended focus of the inquiry (e.g., interpreting a request for feedback as a content-related question).

Complex response integrating more than one component: The chatbot answers the participant's inquiry by integrating other components rather than focusing on just one argument component.

Lack of feedback: The chatbot provides feedback on only one argument component, despite the participant requesting evaluation of multiple components within a single entry.

Confusion among elements of argument**:** The chatbot generates a response by confusing argument components with each other; for instance, it provides feedback for the claim rather than the evidence component.

3.5.3 The analysis of the third research question

To analyze data related to the third research question, "What are the students' perceptions of the effectiveness of the chatbot for developing their arguments?" participants' written responses to the perception form, consisting of seven open-ended questions asked at the end of each session, were collected (see Appendix F). The participants' responses were transcribed into digital text and analyzed using an inductive thematic analysis approach through open coding. Initially, individual responses were coded, and emergent patterns were identified. As new codes appeared during the process, the researcher revisited earlier data to ensure consistency. Once no additional code emerged, existing codes were grouped into broader categories. To enhance trustworthiness, a subject-matter expert reviewed and confirmed the relevance and clarity of the developed codes and categories.

The main categories that emerged from the analysis of participants' perception data were: chatbot's perceived functions, perceived response properties, effect on student motivation, and problems.

In the perceived functions category, based on participants' responses to the questions, the following codes emerged, and the codebook was formed as shown in Table 8: making understandable, learning new information, providing new perspectives, providing guidance, expressing in scientific language, saving time, providing detailed information, and reducing the need to read a longer text.

Table 8. Codebook for Perceived Functions of Chatbot

Code	Definition	Example
Making Understandable	The chatbot helped to simplify the topic, utilizing a proper language compatible with the participants' level.	"When I wrote about the parts I didn't understand, I started to get them."
Learning New Information	The chatbot retrieved and presented existing information that participants were not previously aware of	"I learned that the greenhouse effect increases global warming and causes sea level rise. It explained the effect of the greenhouse effect to me."
Providing New Perspectives	The chatbot offered alternative explanation about the topic.	"It provided ideas I hadn't considered before, which helped me during my writing.
Providing Guidance	The chatbot guided participants on how to construct or revise their arguments.	"I learned how to write arguments more easily, and it helped me construct more organized sentences."
Expressing in Scientific Language	The chatbot supported participants in phrasing their existing arguments using appropriate scientific terminology.	"It supported me in expressing my argument using more scientific language."
Saving Time	The chatbot reduced the need for extra searching or effort by providing quick and relevant answers, helping participants work more efficiently.	"It helped me learn more in less time and made things easier with its explanations."
Providing Detailed Information	The chatbot provided detailed explanations and additional information beyond what was in the text, helping participants understand more deeply	"I can't learn this much detail in class, but with GPT, I can learn the topic more deeply."
Reducing the Need to Read Long Texts	The chatbot highlighted relevant parts, so participants didn't need to read full texts	"Instead of reading the whole text, the chatbot gave me the specific part I needed, which saved time and helped me understand more easily"

For the perceived response property category, the codebook was formed as shown in Table 9; the following codes emerged: understandable, enjoyable, informative, clear, and correct.

Table 9. Codebook for Perceived Response Properties of Chatbot

Code	Definiton	Example
Understandable	The chatbot's style is easy to comprehend.	"It was in a language that even a child could understand."
Enjoyable	Using the chatbot is perceived as a pleasant experience.	"It was both entertaining and informative."
Informative	The chatbot is seen as providing meaningful knowledge.	"I was deeply impressed by its clear and meaningful explanations."
Clear	The chatbot's responses are concise and straightforward.	"The answers it provided were quite clear and explanatory"
Correct	The information provided is accurate.	"GPT always provides accurate and well-organized information."

Additionally, the researcher identified participants' willingness to use the chatbot again regarding the student motivation category. This category was defined as the participant expressing a willingness to reuse the chatbot. An example response representing this category is: "I would like to use it again."

For the problem category, students expressed the problems they experienced during the conversation. The researcher developed the codebook and identified the codes listed in Table 10 under this category: confusion about the components, long responses, absence of pedagogical attitude, and asking for revision repeatedly.

Table 10. Codebook for Problems Encountered in Chatbot Interaction

Code	Definition	Example
Confusion about the Components	Chatbot responses display misunderstanding or mixing of argument components (claim, evidence, reasoning).	“GPT confusing evidence with reasoning made it a bit challenging for me.”
Long Responses	The chatbot gives long-winded answers that include excessive or irrelevant details.	“The responses could have been shorter.”
Absence of Pedagogical Attitude	The chatbot fails to generate guiding or follow-up questions that assess the participant’s understanding	“A teacher can tell whether a student has understood, but GPT cannot.”
Asking for Revision Repeatedly	The participant had to repeat or rephrase their question because the chatbot misunderstood or gave a confusing/off-track response.	“I had to ask again because the chatbot didn’t understand me and got confused.”

3.6 Trustworthiness of the study

Trustworthiness is a cornerstone of qualitative educational research, heavily reliant on the researcher’s credibility (Merriam & Tisdell, 2015). Ensuring validity and reliability facilitates the transferability of findings to other contexts, enhancing the study’s broader applicability (Merriam & Tisdell, 2015). To establish trustworthiness, this study adopted four strategies proposed by Lincoln and Guba (1985): credibility, dependability, confirmability, and transferability.

To ensure credibility, the researcher employed triangulation and persistent observation, as recommended in the literature (Gay et al., 2011; Merriam & Tisdell, 2015). Triangulation was achieved by collecting data from multiple sources, including participants’ worksheets and their conversation logs with the chatbot. These sources complemented each other by offering both product and process-oriented insights. While the worksheets captured the evolution of participants’ written arguments, the chatbot conversation logs illuminated the underlying interaction. Together, these data sources enabled a more comprehensive understanding of how AI-supported interactions facilitated the development of argument quality, strengthening the credibility of the findings. Additionally, the researcher conducted persistent observation, taking detailed field notes while monitoring participants' interactions with the learning module and the chatbot.

To enhance transferability, Lincoln and Guba (1985) emphasize the importance of providing a detailed description of the research context in qualitative studies, enabling other researchers to assess the applicability of findings to their own settings (Gay et al., 2011). In this study, the researcher ensured transferability by thoroughly documenting the context and participant characteristics, including the educational setting, the sixth-grade students' academic profiles (e.g., high achievers), and their interactions with the learning module and chatbot. Additionally, the researcher provided a comprehensive account of the study's evolution following the pilot study, detailing the reasons for modifications to the research design, such as adjustments to the learning module or chatbot prompts. These detailed descriptions facilitate comparisons with other studies, supporting the potential transferability of the findings.

The researcher employed rigorous data analysis procedures to ensure confirmability, minimize bias, and enhance objectivity. Initially, open coding was used to identify codes and categories from the chatbot conversation logs, followed by rubric-based analyses to assess the quality of participants' arguments. To further strengthen confirmability, the researcher collaborated with her thesis advisor to review and refine the coding and analysis processes. This collaborative approach aligns with best practices for maintaining neutrality in qualitative research (Lincoln & Guba, 1985).

3.7 Ethical considerations

Conducting a study with careful attention to ethical considerations is a fundamental responsibility of researchers. The National Research Council (NRC, 2003) outlines several key principles that researchers must prioritize, including obtaining informed consent from participants, protecting their privacy and confidentiality, and ensuring fair participation. Adhering to these ethical principles is essential to safeguard participants' rights and maintain the integrity of the research process.

In this study, the researcher ensured compliance with ethical standards at every phase. Initially, ethical approval was obtained from the Ethics Committee in Social Sciences and Humanities (SOBETIK) at Boğaziçi University (see Appendix G). Given that the participants were sixth-grade students from a public school, additional permission was secured from the Ministry of National Education (MoNE; see Appendix H). The research commenced only after receiving approval from the committees responsible, ensuring alignment with legal and ethical standards. Following MoNE approval, the researcher informed the school principal and teachers about the study, and it was subsequently announced to a sixth-grade classroom in the public middle school. Volunteer students, identified as high achievers, were selected in consultation with the classroom guidance teacher and enrolled in the learning module.

Students willing to participate were thoroughly informed about the study's purpose and research process. Informed consent was obtained from the students' parents and the students through consent forms (see Appendix I; Appendix J). Participants were assigned numerical identifiers to ensure confidentiality and anonymity throughout the data collection, analysis, and reporting processes. All study-related data, including consent forms and worksheets, were securely stored in a locked box. Participants' conversations with the chatbot, field notes, and other digital data were stored on the researcher's hard disk, accessible only to the researcher and the thesis advisor. Upon completion of the research, all collected data, including materials stored on the hard drive and in the locked box, were destroyed to protect participants' privacy further.

3.8 Role of the researcher and reflexivity

The researcher's background and experiences play a pivotal role in shaping the processes of a qualitative study. The researcher holds a bachelor's degree in mathematics and science education from Boğaziçi University, followed by a master's degree in educational technology from the same institution. She participated in an Erasmus exchange program at the University of Wuppertal during her academic journey, gaining international exposure to educational practices.

Professionally, the researcher worked as a science teacher in a public middle school for two years, teaching students in grades 5 through 8. During this time, she actively designed and implemented inquiry-based learning environments to promote critical thinking and student engagement. Her classroom experience informed the development of the learning module used in this study and allowed her to adapt instructional strategies to suit students' needs.

In this study, the researcher actively engaged in various facets of the research process. She designed both the learning module and the chatbot, customizing them to facilitate scientific argumentation among her students. The learning module was implemented with her sixth-grade class, where she fostered a supportive classroom environment by providing personal devices to ensure equitable access for all participants. Throughout the study, she oversaw the implementation, reviewed the processes, and ensured the accuracy and reliability of the data through careful monitoring and verification. Her dual expertise in science education and educational technology, along with her practical experience as a teacher, likely influenced the study’s design, execution, and interpretation, thereby enhancing its relevance to actual classroom settings.

CHAPTER 4

FINDINGS

This case study aimed to explore how a learning module designed based on the Argument-Driven Inquiry (ADI) instructional model with generative AI integration facilitates sixth-grade students' scientific argument quality in the context of the greenhouse effect. Data was collected from participants through worksheets consisting of a scientific argument template, chatbot conversation logs, and perception forms, which were subsequently analyzed using a thematic analysis approach. This chapter presents the results of the data analysis, focusing on the changes observed in the quality of participants' scientific arguments and the role of chatbot interactions in facilitating the scientific argumentation process.

4.1 Exploring participants' scientific argument quality

The first research question of this study is "How does a learning module designed based on the Argument-Driven Inquiry (ADI) instructional model with generative AI integration facilitate sixth-grade students' scientific argument quality in the context of the greenhouse effect topic.

In this study, five participants took part in three sessions. In each session, they constructed two scientific arguments, one before interacting with the chatbot (pre-argument) and one after (post-argument). As a result, each participant produced a total of six scientific arguments. The participants' quality of scientific arguments was analyzed under the scientific argument components based on the rubric developed by McNeill et al. (2006), including claim, evidence, and reasoning (see Appendix C), and these components were assigned as levels (Level 0, Level 1, and

Level 2). Throughout the sessions, participants initially formed the research question, and then they formed a "claim" that responds to the research question; thereafter, "evidence" was formed, which provides scientific data to support the claim, and then the "reasoning" component was constructed to explain why this evidence is significant. The researcher compared and reported components of the pre- and post-arguments to determine the change in participants' scientific argument quality in each session. This section presents the analysis of participants' scientific arguments and how the quality of each argument component changed across the three sessions.

4.1.1 Participants' scientific argument quality in Session 1

Session 1 focused on the impact of global temperature rise due to the greenhouse effect. Since this was the first session, most participants demonstrated lower-quality arguments before the chatbot interaction, particularly regarding evidence and reasoning. As shown in Table 11, two out of five students had Level 0 or Level 1 evidence, and three out of five had Level 0 or Level 1 reasoning in their pre-arguments.

Table 11. Change in Participants' Argument Component Levels in Session 1

Participant Name	Pre/Post Argument	Levels of Argument Components		
		Claim	Evidence	Reasoning
Ruby	Pre-argument	Level 1	Level 1	Level 0
	Post Argument	Level 2	Level 2	Level 1
Angelia	Pre-argument	Level 2	Level 1	Level 0
	Post Argument	Level 0	Level 2	Level 2
Emily	Pre-argument	Level 2	Level 2	Level 2
	Post Argument	Level 2	Level 0	Level 2
Rose	Pre-argument	Level 1	Level 2	Level 0
	Post Argument	Level 2	Level 2	Level 2
Micheal	Pre-argument	Level 2	Level 2	Level 2
	Post Argument	Level 2	Level 2	Level 2

Ruby initially produced the lowest quality pre-arguments across all components among the participants. She primarily used the chatbot to deepen her understanding of the topic and to receive structured guidance. Before engaging with the chatbot, she appeared to conflate the evidence and reasoning components, which led her to construct a reasoning statement that merely restated evidence. As a result, her pre-reasoning was categorized as Level 0. Owing to the chatbot interaction, she revised the pre-reasoning to link evidence to the claim, giving rise to Level 1. Among all participants, she demonstrated the most notable improvement across argument components. This progress may be attributed to the illustrative examples and scaffolded support provided by the chatbot, which appeared to enhance her understanding of the distinct features of each component.

Angelia initially constructed one of the lowest-quality pre-arguments, particularly in the evidence and reasoning components. Following the chatbot interaction, she revised these elements with greater specificity and scientific detail, resulting in both being classified as Level 2. However, her post-claim statement focused solely on scientific data, rather than articulating a clear and overarching conclusion in response to the research question. Consequently, her post-claim was coded as Level 0 due to a misclassification of argument components.

The analysis of the chatbot interaction indicated that this misunderstanding occurred because the chatbot presented the claim as an integrated part of the overall argument. Consequently, the participant misinterpreted the claim component and wrote an evidence-based statement instead. Moreover, her initial reasoning lacked coherence, was not presented in complete sentence form, and did not establish clear links to the claim or evidence. After interacting with the chatbot, she revised this component by constructing a reasoning statement that successfully connected the claim and evidence, incorporating appropriate scientific language.

Emily had one of the highest pre-arguments regarding all components; therefore, she did not change the argument components except for pre-evidence. Although the chatbot was generally effective in enhancing the quality of students' arguments, in this case, it provided guidance related to reasoning rather than evidence. As a result, the participant mistakenly formulated a reasoning statement in the evidence section, which led to a Level 0 classification for post-evidence.

Rose had one of the lowest quality pre-arguments regarding the claim and reasoning components. For the pre-reasoning, the participant was initially confused about evidence and reasoning components; therefore, she wrote evidence in the reasoning part before the chatbot interaction, resulting in Level 0. Following the chatbot interaction, she was able to identify this mistake and formed a post-reasoning, which links evidence and claim, and the level of post-reasoning became level 2.

Micheal constructed suitable pre-argument components, achieving the highest level among participants, which corresponded to Level 2. Throughout the session, he demonstrated a comprehensive understanding of both the subject matter and the structure of scientific arguments. Consequently, he expressed his intention to utilize the chatbot just to ensure his argument components by asking for feedback. After the interaction, the chatbot's feedback confirmed the adequacy and coherence of his pre-arguments; therefore, he did not revise the post-argument components. Table 12 portrays an example of a participant’s argument components obtained from Session 1.

Table 12. Scientific Argument Components with Examples in Session 1

Before/After Interaction	Claim	Evidence	Reasoning
Pre-Argument (Before Interaction)	It further worsens drought conditions, leading to crop failures, forest fires, and desertification. Level 1	In 2017, a forest area 1.3 times the size of our Yalova province was damaged by fires. Level 1	Scientists predict that if world temperatures rise by another 2°C, summer months will be much hotter and thermometers could rise to 50°C in regions such as the Middle East and North Africa. Level 0
Post Argument (After Interaction)	The greenhouse effect can trigger climate change by triggering an increase in global temperatures. Level 2	The frequency of extreme weather events, such as droughts and forest fires, is increasing about climate change, with negative impacts on agricultural water resources and food security. Level 2	The greenhouse effect causes temperatures to rise as greenhouse gases in the atmosphere increase, which leads to climate change. Level 1

4.1.2 Participants' scientific argument quality in Session 2

Session 2 was about the effect of melting ice due to the greenhouse effect. As observed in Session 1, the majority of the participants constructed lower-quality pre-arguments for all components before the chatbot interaction. As shown in Table 13, four out of the five participants had Level 1 claim, three out of five students had Level 1 evidence, and two out of five had Level 0 or Level 1 reasoning in their pre-arguments.

Table 13. Change in Participants' Argument Component Levels in Session 2

Student Name	Pre/Post Argument	Levels of Argument Components		
		Claim	Evidence	Reasoning
Ruby	Pre-argument	Level 1	Level 1	Level 1
	Post Argument	Level 2	Level 2	Level 2
Angelia	Pre-argument	Level 2	Level 1	Level 0
	Post Argument	Level 2	Level 2	Level 2
Emily	Pre-argument	Level 1	Level 2	Level 2
	Post Argument	Level 1	Level 2	Level 1
Rose	Pre-argument	Level 1	Level 1	Level 2
	Post Argument	Level 2	Level 1	Level 2
Micheal	Pre-argument	Level 1	Level 2	Level 2
	Post Argument	Level 2	Level 2	Level 2

Ruby generated the lowest level of pre-argument for all components compared to other participants. Prior to engaging with the chatbot, her pre-claim required further explanation, and her pre-evidence was deficient in scientific information. Her pre-reasoning was scientifically accurate, but it was irrelevant to other components, and she failed to establish a direct connection between evidence and claim, resulting in Level 1. During the session, she benefited from the chatbot by asking for guidance, and the chatbot provided possible examples for all components, enabling her to identify relevant aspects that were compatible with her argument structure. Consequently, she developed detailed and high-quality components.

Angelia formed the lowest level of pre-argument, particularly for evidence and reasoning components. Before the chatbot interaction, she constructed pre-evidence insufficient to support the claim, which lacked scientific data; during the conversation, the chatbot presented scientific data for the participant to support the claim. Thus, she constructed post-evidence corresponding to Level 2. In addition, she initially misunderstood the purpose of the reasoning component and provided an explanation that lacked a clear connection between the claim and the evidence. Following the chatbot interaction, however, she revised her reasoning to explicitly link the claim and evidence by incorporating relevant scientific concepts. Her revised post-reasoning demonstrated how the evidence supported the claim using appropriate terminology and a logical structure, resulting in a more coherent and scientifically sound argument.

Emily produced a high level of pre-argument components, except for the claim component. Regarding the claim component, she sought feedback, but the chatbot delivered contradictory feedback by affirming her response and providing an explanation of the possible claims. Even though the chatbot presented a stronger claim, the participant focused on the affirmation part and did not change her pre-claim. Although she generated a high level of pre-reasoning, during the session, she couldn't manage the time properly, and she did not ask for feedback for the reasoning; therefore, she wrote a short sentence omitting scientific information and did not mention relationships, among other components.

Rose generated a lower level of pre-argument, particularly in the claim and evidence components. She constructed a pre-claim that comprised the topic but was too general and did not directly address the context. After the interaction, she revised it, focusing on the context, and answered the research question, resulting in Level 2. In addition to that, during the chatbot conversation, the participant asked for guidance on evidence, and the chatbot generated a possible example for evidence components. In light of the chatbot's response, the participant formed post-evidence but excluded some scientific data that could support the claim; therefore, it became a vague pre-argument component. This can be attributed to the participants' time management problem based on the researcher's observations.

Micheal generated high level of pre-argument components except for the pre-claim component. Although his pre-claim was scientifically accurate, it lacked alignment with the accompanying evidence and reasoning, leading to its classification as Level 1. During the chatbot interaction, he asked for feedback and guidance for claim components, and the chatbot generated detailed feedback. Based on this, he constructed a post-claim compatible with other components, resulting in Level 2. Table 14 portrays an example of a participant's scientific argument obtained from Session 2.

Table 14. Scientific Argument Components with Examples in Session 2

Before/After Interaction	Claim	Evidence	Reasoning
Pre-Argument (Before Interaction)	Global warming melts glaciers Level 1	Global increase melts sea ice Level 1	Increase delays in ice growth in autumn and winter Level 1
Post Argument (After Interaction)	Melting glaciers cause sea levels to rise and natural aftershocks to occur more frequently in coastal areas. Level 2	Greenland loses an average of 270 billion tons of ice per year, while Antarctica loses 150 billion tons of ice per year. These losses contribute to the rise of sea level, accounting for about a third of the global sea level rise between 2006 and 2015. Level 2	Melting glaciers raise sea levels, increasing coastal erosion. Warming ocean temperatures and weather also make storms more frequent and intense. The ice sheets in Greenland and Antarctica are the most affected by this process. Level 2

4.1.3 Participants' scientific argument quality in Session 3

The topic of Session 3 was the effect of sea level rise due to the greenhouse effect. Since this was the last session, most participants developed their understanding of how to form a scientific argument. Therefore, they demonstrated higher-quality arguments before the chatbot interaction, especially regarding evidence and reasoning components. As shown in Table 15, three out of five participants had Level 2 claim and four out of five had Level 2 evidence and reasoning in their pre-arguments.

Table 15. Change in Participants' Argument Component Levels in Session 3

Student Name	Pre-/post-argument	Levels of Argument Components		
		Claim	Evidence	Reasoning
Ruby	Pre-argument	Level 2	Level 1	Level 2
	Post Argument	Level 2	Level 2	Level 2
Angelia	Pre-argument	Level 2	Level 2	Level 1
	Post Argument	Level 2	Level 2	Level 2
Emily	Pre-argument	Level 2	Level 2	Level 2
	Post Argument	Level 2	Level 2	Level 2
Rose	Pre-argument	Level 1	Level 2	Level 2
	Post Argument	Level 2	Level 2	Level 2
Micheal	Pre-argument	Level 1	Level 2	Level 2
	Post Argument	Level 2	Level 2	Level 2

Ruby produced higher levels of pre-argument components, except the evidence component. She generated an accurate and complete pre-claim related to other components; however, it lacked scientific terminology. By interacting with the chatbot, she was able to adopt the scientific language it used, prompting her to replace informal expressions with more appropriate scientific terms in her post-claim. Her pre-evidence was not directly related to the claim and was deficient in scientific data, resulting in level 1. During the conversation, the participant asked for guidance for the evidence component and benefited from the response received while revising the pre-evidence. Consequently, she developed a post-evidence component that incorporated scientific data supporting the claim, achieving a Level 2 classification.

Angelia formed higher levels of pre-argument components, with the exception of the reasoning component. Her pre-reasoning did not establish a direct connection with the evidence and claim components. Throughout the session, the participant persistently asked further questions related to the reasoning component, which she revised based on data provided by the chatbot. Thus, she generated a post reasoning that aligned with the other components and supported the evidence.

Emily produced higher levels of pre-argument for all components. Her pre-evidence was appropriate but required more detail to support the claim. During the chatbot interaction, the participant persistently asked for guidance for the evidence component and revised it so that it supports the claim. Thus, the developed post-evidence resulted in Level 2.

Rose generated higher levels of pre-argument, with the exception of the claim component. She formed a complicated pre-claim that addressed multiple aspects simultaneously but lacked coherent sentence structure. Owing to the chatbot interaction, the participant revised it, focusing on specific aspects compatible with the other components, resulting in a Level 2 claim component.

Micheal formed higher levels of pre-argument with the exception of the claim component. He generated a pre-claim that included only scientific information that was not directly related to evidence and reasoning, which needed explanation, resulting in Level 1 categorization. After the chatbot interaction, he revised it to answer the research question. Based on the claim, he generated post-evidence and post-reasoning components. While doing that, he asked for guidance on these components and benefited from the chatbot responses. As a result, he generated post-argument components, resulting in Level 2 categorization. Table 16 portrays an example of a participant’s argument obtained from Session 3.

Table 16. Scientific Argument Components in Session 3 with Examples

Before/After Interaction	Claim	Evidence	Reasoning
Pre-Argument (Before Interaction)	A rise in sea level is an increase in the total volume of ocean water. Level 1	In many regions, the impact of rising sea levels on societies is already beginning to be seen. Level 2	It displaces more than 50,000 people from their homes in the Pacific every year. Level 2
Post Argument (After Interaction)	The impact of rising sea levels on societies is already beginning to be seen in many regions. Level 2	River flooding and the flooding of land in China have demonstrated significant impacts on the region's ecosystems Level 2	Habitat loss and agricultural land damage in China pose significant challenges to biodiversity and food security in the region. Level 2

4.2 Participants' interaction with the chatbot for developing scientific arguments

The second research question is: "How do students engage with a specifically designed chatbot to enhance their argument development during the learning module?" The analysis focused on the prompts generated by students and the chatbot's responses during their interactions, assessing them based on their intended purposes to address this research question. Following a thematic analysis of data collected from three sessions, several common categories and codes emerged. To further evaluate the role and effectiveness of chatbot interaction in scientific argumentation, three sub-research questions were formulated and addressed in the following sections.

4.2.1 Purposes for using the chatbot for developing scientific arguments

Understanding the participants' intentions is key to grasping the reasons behind their revision behavior and the development of scientific arguments. Since these intentions influence the flow of conversation and the chatbot's responses, it is crucial to analyze participants' purposes in order to explore how the interaction unfolds.

As illustrated in Table 17, most participants benefited from the chatbot asking for guidance for argument components (f = 76), and generally, they asked for guidance for evidence components (f = 24). Additionally, the participants interacted with the chatbot to seek feedback on their argument components (f = 43), mainly focusing on the claim components (f = 11). Furthermore, they utilized the chatbot to request general information (f = 19), which is essential for participants' understanding of the issue. Moreover, participants utilized the chatbot to request a revision of the given feedback (f = 9), ask questions out of curiosity (f = 9), and ask for summarizing (f = 6). Table 17 shows the frequency of participants' reasons for chatbot interactions.

Table 17. Participants' Reasons for Chatbot Interactions

Categories	Codes	Frequencies				
		Angelia	Emily	Rose	Micheal	Ruby
Asking Feedback	for more than one argument component	1	0	3	2	0
	for the research question	1	3	3	1	0
	for claim	3	2	1	3	2
	for evidence	1	2	2	3	2
	for reasoning	1	2	2	3	0
Total number of entries for this category		7	9	11	12	4
Asking Guidance	for more than one argument component	1	2	2	6	3
	for the research question	1	0	1	2	1
	for claim	5	1	2	2	4
	for evidence	3	10	3	3	5
	for reasoning	3	3	2	6	5
Total number of entries for this category		13	16	10	19	18
Other	Requesting a revision of the given feedback	3	1	1	1	3
	Asking for general information	2	1	4	3	9
	Asking for summarizing	2	0	0	2	2
	Asking questions out of curiosity	6	2	0	1	0
Total number of entries for this category		13	4	5	7	14

Ruby

Ruby interacted with the chatbot 16 times in the first session, 9 times in the second, and 11 times in the third. Her interactions were primarily aimed at asking for guidance (f = 18) and feedback (f = 4), requesting revisions to previously received feedback (f = 3), obtaining general information (f = 9), and asking for summarizing (f = 2). Instead of attempting to revise and improve her initial argument, she opted to explore example-based guidance from the chatbot for each argument component. Rather than refining her own ideas, she incorporated the chatbot's responses almost verbatim into her post-argument. As a result, she demonstrated the most substantial improvement among the participants, which can be attributed to her reliance on and effective use of the chatbot's scaffolding.

In the guidance category, Ruby asked for guidance generally for reasoning (f = 5), evidence (f = 5), and claim (f = 4). For instance, when seeking guidance on reasoning, she posed the question, "What could be the reasoning for the melting of glaciers?" After that, the chatbot generated comprehensive reasoning for the inquiry, and she revised her pre-reasoning based on the received response, resulting in the higher-level post-reasoning component.

Within the category of requesting general information (f = 9), she formulated the question, "How do drought conditions affect crop productivity?" She may need scaffolding to enhance her understanding of the issue. This suggests that the chatbot may be a valuable tool to learn about the details of the topic.

For the asking for summarizing category (f = 2), the participant asked, "Can you write more briefly and clearly?" This finding is consistent with the data gathered for the third research question, which explores participants' perceptions of their chatbot experience. According to these responses, several participants reported difficulty in comprehending the chatbot's lengthy replies, which made it challenging to identify the most relevant parts of the response. The extended length of the chatbot's output appeared to hinder their ability to focus on key ideas. As a result, participants occasionally felt the need to request a more concise summary to better understand the information provided.

Angelia

In the first session, Angelia formed 10 entries, and 11 for sessions 2 and 3. According to the analysis of conversation logs generated during her interactions, she benefited from the chatbot in several ways: she asked for guidance (f = 12), and feedback (f = 7), posed questions out of curiosity (f = 6), requested revisions for given responses (f = 3), and asked for general information (f = 2).

For the asking for feedback category, she utilized mostly asking for feedback for the claim (f = 3). For example, the participant stated, “My claim is that the greenhouse effect on the world has increased over the years. Is that correct?” However, the chatbot did not provide the desired response, leading her to ask for guidance on her claim. The chatbot then offered overall guidance for her argument rather than specific feedback on the claim itself. Since it was the first session and the participant was not proficient enough in detecting argument components, she revised her pre-claim erroneously, resulting in an evidence component.

She also benefited from the chatbot in Session 1, requesting a revision of the given response (f = 3) when the chatbot confused the argument components with each other, providing feedback that addressed both the evidence and reasoning components rather than focusing solely on the evidence as intended; the participant prompted, “You wrote the two components together; can you just write the reasoning?” and the chatbot revised the feedback focusing on the reasoning component; thus, the participant revised the pre-reasoning based on the received feedback, resulting in a post-reasoning with Level 2 categorization.

Apart from these categories, she utilized the chatbot mostly to ask questions out of curiosity (f = 6). For example, she asked the following question, “Well, where do you think will be underwater in Turkey in 2050?” One important point to highlight here is that this question illustrates that the chatbot may have enhanced the participant’s motivation to explore the topic further and encouraged her to pose questions she might hesitate to raise in a traditional classroom setting.

Emily

Emily interacted with the chatbot 10 times in the first session, only 3 times in the second, and 16 times in the third. Her limited interaction during the second session appears to be the result of time management difficulties while constructing her pre-arguments, which left insufficient time for engaging with the chatbot.

Throughout all sessions, she interacted with the chatbot to ask for feedback (f = 9) and guidance (f = 14). Additionally, she requested a revision of previously received feedback (f = 1), asked for general information (f = 1), and posed curiosity-driven questions (f = 2). For the asking guidance category, she asked for guidance for claim (f = 1) by asking, “My claim is that even a small increase in temperature can cause overheating. Do you think this claim of mine is okay? If not, can you give an example?” Following that, the chatbot verified the participant's pre-claim and provided additional details. Although there was a possible claim example in the chatbot's response, the participant focused solely on the confirmation aspect and did not alter the pre-claim.

Additionally, she asked for guidance for evidence (f = 10) by posing the following question: “What could be the evidence for this reasoning?” This allowed the participant to gather possible examples of evidence related to her reasoning during the conversation.

Rose

Rose engaged with the chatbot 9 times in the first session, 6 times in the second, and 11 times in the third. She benefited from the chatbot for feedback for the argument category (f = 11) and for guidance for the argument category (f = 10). Apart from these categories, she used it for the asking for general information category (f = 4) and for requesting revision of the given feedback category (f = 1). In the feedback for the argument category, she utilized it mostly to get feedback for the research question (f = 3). As an example of this code, the participant constructed, “Do the negative impacts on living organisms contribute to global temperature rise? Is this question okay?”

After formulating her research question, she sought guidance on a claim that would align with it. In addition to exploring these categories, she used the chatbot to gather general information (f = 4) by inquiring, “What are the effects of global temperature increase?” As a result, she was able to incorporate the insights gained from this question into her reasoning component. Consequently, the participant enhanced her understanding of the topic, which is essential for developing the argument components.

Micheal

In the sessions, Micheal interacted with the chatbot 14, 10, and 15 times, respectively. He was eager to engage with the chatbot and used it to validate his initial arguments. He benefited from the chatbot for asking for feedback (f =12) and guidance (f = 18) on the argument. In the asking guidance category, he frequently asked for guidance for more than one component (f = 6). As an example of this code, he asked, “How can I make this argument better?” Then, the chatbot provided different perspectives, which are essential in forming comprehensive arguments.

Other than that, he interacted with the chatbot to request a revision of the given response once. Before that, he stated his claim as follows: “Rising temperatures exacerbate drought conditions in arid regions, leading to crop shortages, wildfires, and desertification, is it okay?” and he stated his evidence as follows: “In regions such as the Middle East and North Africa in the summer months, temperatures are expected to reach up to 50 C.” Then he asked for feedback for his evidence and the chatbot generated a response irrelevant to the participant's claim and evidence as follows:

> Your evidence appears appropriate. You have presented various data supporting the negative impacts of temperature increases on agricultural production. For instance, the information that global temperatures have exceeded 1°C above pre-industrial levels, with an increase of 1.5°C in Turkey, highlights the severity of climate change. Additionally, evidence exists demonstrating that rising temperatures exacerbate drought conditions, subsequently leading to issues such as crop shortages, wildfires, and desertification. These data provide a strong foundation for supporting your claim. I am here to assist if you need help with another topic!

After that Micheal prompted, "No, I'm not asking that. Is the evidence I wrote appropriate?" After this prompt, the chatbot revised the response and fixed the mistake as evident in the following extract:

> Yes, the evidence you provided is appropriate. The estimate that temperatures could reach up to 50°C highlights the impacts of climate change and drought. This information demonstrates the severity of temperature increases across different regions of the world. Additionally, it provides a context supporting the negative effects of these temperatures on agricultural production and food security. However, you can strengthen this evidence by adding more details. Would you like assistance with another topic?

4.2.2 Chatbot's functions responding to participants' requests

Comprehending the chatbot's capabilities and functions allows for a clearer understanding of participants' interactions and their revision behaviors in response to the feedback provided. Moreover, analyzing the chatbot's behavior and operational mechanisms enhances insight into designing a chatbot-integrated lesson for practical applications.

As presented in Table 18, the chatbot primarily served to provide guidance (f = 84) and feedback (f = 32) for arguments and provide general information for inquiry (f = 29), answer questions out of curiosity (f = 8) revise the response (f = 6) and provide summary (f = 5). The chatbot exhibited a tendency to offer guidance when participants requested feedback; additionally, it generated responses that included general information, even when participants had not specifically sought such details. Table 18 demonstrates the chatbot's function in participants' conversations.

Table 18. Chatbot's Function According to Categories

Categories	Extracted Codes	Frequencies in Participants				
		Angelia	Emily	Rose	Micheal	Ruby
Providing Feedback	for more than one argument component	3	0	1	3	0
	for research question	1	0	1	0	0
	for claim	1	2	3	1	1
	for evidence	1	2	1	4	2
	for reasoning	0	1	1	3	0
Total number of entries for this category		6	5	7	11	3
Providing Guidance	for more than one argument component	0	2	3	8	2
	for research question	3	0	1	2	1
	for claim	3	2	2	2	4
	for evidence	5	9	4	3	7
	for reasoning	1	4	2	6	8
Total number of entries for this category		12	17	12	21	22
Other	Revising the response	3	1	1	0	1
	Providing general information for inquiry	5	4	6	5	9
	Providing summary	2	0	0	2	1
	Answering the curiosity questions	5	2	0	1	0
Total number of entries for this category		15	7	7	8	11

In Angelia's conversation, the chatbot generated a total of 34 responses: 9 from the first session, 12 from the second, and 13 from the third session. The chatbot generated responses to provide feedback categories (f = 6) and guidance (f = 5). Other than these categories, the chatbot produced responses mostly for general information for inquiry (f = 5) and answering the questions out of curiosity (f = 5). The participant asked the following question, which was categorized as a question out of curiosity: “What will the world be like in 2070 if the glaciers continue to melt?” The chatbot generated the following response to answer the question:

> If glaciers continue to melt, sea levels will rise significantly by 2070, posing a major threat to coastal areas. In addition, melting glaciers will disrupt ocean currents, causing changes in weather patterns and resulting in more severe storms and floods. Melting glaciers will also negatively impact fishing industries by affecting marine ecosystems (Session 3, Angelia).

In Emily’s conversation, the chatbot generated a total of 29 responses, producing 10, 3, and 16 responses during the sessions, respectively. It generated responses to provide feedback (f = 5) and guidance for the argument (f = 17). Even though the participant asked for feedback (f = 9), the chatbot provided less than it. This discrepancy appears to stem from the way Emily formulated her prompts. She frequently requested feedback while also asking for examples, which the chatbot likely interpreted as a request for guidance. Additionally, her use of brief or one-word prompts, potentially due to limitations in typing fluency, may have contributed to misinterpretations of her intent.

For the other categories, the chatbot formed a response to revise the existing response (f = 1), provide general information for inquiry (f = 4), and answer the questions out of curiosity (f = 2). In the providing guidance category, the chatbot provided mostly guidance for evidence (f = 9). It is noteworthy that the chatbot demonstrated a high level of understanding, generating a response even without a complete sentence prompt relevant to the context. When the participant wrote “Evidence ?” the chatbot formed the response below:

> As evidence for your case, you could point out that flooding in Bangladesh has polluted freshwater sources, causing waterborne diseases that threaten people’s health. You could also support the argument that this pollution has caused people to experience stress and mental health problems. Such evidence would strengthen your case. Do you have any other questions? (Session 3, Emily).

In Rose’s conversation, the chatbot formed 26 responses. In sessions, it produced 9, 6, and 11 responses, respectively. The chatbot generated responses by providing feedback (f = 7), guidance (f = 11), general information for inquiry (f = 6), and revising the response (f = 1). In the general information category, the participant asked, “What are the effects of global temperature increase?” The chatbot generated the response below. She revised her pre-evidence based on this chatbot's response.

> The effects of global temperature increase include increasing the severity of extreme weather events, worsening drought conditions, and food shortages. The amount of energy in the atmosphere is increasing due to human activities, which further intensify weather events. In addition, a 1°C increase in temperatures leads to a decrease in precipitation by up to 30% in regions such as Turkey, threatening food security. The increase in temperature also causes an increase in forest fires and deaths due to heat waves. All these effects disrupt the ecosystem balance and threaten the habitats of living things (Session 1, Rose).

In Micheal’s conversation, the chatbot generated a total of 40 responses across three sessions. The sessions produced 15, 10, and 15 responses, respectively. The chatbot generated responses to provide feedback (f =11), guidance (f = 21), general information for inquiry (f = 5), and to answer questions out of curiosity (f = 3). For the category of providing guidance, the chatbot frequently produced responses to provide guidance for more than one component (f = 8). For example, when Micheal wrote the following sentence: “Write me an argument similar to this.” The chatbot generated the response below. This type of interaction not only provided scaffolded support but also enabled Michael to consider alternative explanations and structure his arguments more effectively:

> Sure, here is a similar argument:
> Claim: Global warming threatens food security by negatively affecting agricultural production.
> Evidence: According to studies, agricultural productivity decreases with increasing temperatures. For example, agricultural products were seriously damaged in some regions of Türkiye in 2017 due to extreme temperatures.
> Reasoning: Rising temperatures negatively affect the growth cycles of plants, increasing water demand and exacerbating drought conditions. This leads to loss of productivity in agricultural production and, as a result, food security problems. In this way, you can create an argument with evidence and reasoning supporting your claim. Would you like me to help you with anything else?

In Ruby’s conversation, the chatbot created 36 responses for three sessions. In sessions 15, 10, and 11, responses were produced, respectively. The chatbot responded to provide feedback (f = 3), guidance (f = 22), general information for inquiry (f = 9), revising response (f = 1), and providing a summary (f = 1). For the providing guidance category, the chatbot generated mostly a response for guidance for reasoning (f = 8). When the participant asked the following question, “What could be my reasoning about the melting of glaciers?” The chatbot formed the response below. After receiving this response, she revised her pre-evidence, benefiting from the guidance provided.

> Your reasoning might be, "Melting glaciers add water to the ocean, causing the sea level to rise. This increases coastal erosion while warming ocean temperatures cause storms to become more frequent and intense.
> (Session 3, Ruby)

4.2.3 Chatbot’s response quality in supporting participants’ scientific argumentation

To explore the overall effectiveness of the chatbot in supporting participants’ scientific argument formation process, an analysis of the chatbot’s response quality is essential. Given that the quality of the responses shapes the progression of the conversation, it directly influences participants’ revision behaviors and the nature of their interactions.

The chatbot generated 163 responses for all participants; as shown in Table 19, 135 were suitable for context and correct, 9 were out-of-context and complex responses integrating more than one component, 3 lacked feedback, and 7 were confused components of the argument. Table 19 shows the chatbot's response quality in participants' conversations.

Table 19. Distribution of Chatbot's Response Quality to Participants

	Categories					
Participants' Names	Suitable for Context and Correct	Out of Context	Complex Response Integrating More Than One Component	Lack of Feedback	Confusion Among the Elements of Argument	Total Entry
Angelia	26	1	1	1	4	33
Emily	23	3	3	0	0	29
Rose	21	0	1	2	2	26
Micheal	34	3	2	0	1	40
Ruby	31	2	1	0	1	35

The chatbot performed in a manner that was compatible with the context, and most of the responses were suitable (f = 135). For an example of a suitable context and correct category, the following conversation was given in Figure 2. After the participant received the feedback below, she updated her argument and added specific evidence to support her claim, mentioning the rise in Bangladesh, resulting in Level 2 post-evidence as reported in research question 1.

Participant's Prompt	Chatbot's Response
My evidence is that the sea level has not risen for thousands of years, so people have settled on the coasts of the world, but the sea level has risen by 5 centimeters in 20 years. Do you think this evidence is enough? (Session 3, Emily)	Your evidence is an important starting point to support human settlement on the coasts due to the lack of sea level rise for thousands of years. However, this information alone may not be enough. The 5 cm sea level rise in the last 20 years may require additional evidence to explain better how humans will be affected and the risks experienced in coastal areas. Details such as the acceleration of the rise and the negative effects on human communities will make your case stronger. Do you have any other questions?

Figure 2. Example conversation suitable for the context and correct category

Two instances were categorized as out-of-context responses (f = 9). In the first, the chatbot addressed the participant's research question directly instead of providing feedback on it. In the second, the chatbot produced a response that failed to recognize that the participant had not completed their sentence. An example of the first situation is given in Figure 3.

Participant's Prompt	Chatbot's Response
My research question is the melting of glaciers. (Session 3, Micheal)	Melting glaciers contribute significantly to sea level rise and climate change. The melting of the Greenland and Antarctic ice sheets in particular, is one of the biggest causes of sea level rise. Melting glaciers disrupt weather patterns, causing storms and hurricanes to occur more frequently. Want more information on another topic?

Figure 3. Example conversation for out of context category

The participant was aware that he could not get feedback on his research question; therefore, he asked again for guidance on the research question and related claim.

For an example of the complex response integrating more than one argument component category (f = 9), the chatbot generated a response to provide guidance not only for the reasoning component but also for evidence. The example conversation is given in Figure 4. After receiving this response, she repeated the prompt, and this time, the chatbot separated the components, and the participant revised her pre-reasoning accordingly, resulting in Level 2, as mentioned in research question 1.

Participant's Prompt	Chatbot's Response
What could be my reasoning for the greenhouse effect? (Session 1, Ruby)	Your reasoning for the greenhouse effect could be as follows: The greenhouse effect causes temperatures to rise due to an increase in greenhouse gases in the atmosphere, which leads to climate change. Scientific data shows that human activities have increased greenhouse gas concentrations, resulting in temperatures exceeding 1°C worldwide. This has negative effects on agriculture and water resources, posing serious threats to food security. The frequent occurrence of extreme weather events is also a result of climate change, which can strain societies economically and socially.

Figure 4. Example conversation for complex response integrating more than one component category

Apart from these, the chatbot confused the components of the argument with each other (f = 7); for instance, despite the participant's request to receive feedback on the research question, the response provided was more akin to an answer to a separate claim. Figure 5 shows an example conversation for this category. After receiving this response, the participant did not ask for feedback on the research question again.

Participant's Prompt	Chatbot's Response
The research question is whether the greenhouse effect has changed on Earth over the years. Is it enough?" (Session 1, Angelia)	You have not explained your claim yet. Can you add scientific evidence and justifications to explain your claim?"

Figure 5. Example conversation for the confusion of the argument components category

4.3 Participants' perceptions of the effectiveness of the chatbot

The third research question is: "What are the students' perceptions of the effectiveness of the chatbot for developing their arguments?" In order to answer the research question, the researcher analyzed participants' written responses to seven open-ended questions. Based on the analysis of answers to the perception questions, the following four categories emerged: chatbot's perceived functions, perceived response properties, effect on student motivation, and problems.

In the chatbot's perceived functions category, as shown in Table 20, participants generally regarded the chatbot as a tool for learning new information (f = 20) and providing guidance/convenience while writing scientific arguments (f = 11). Besides, participants also benefited from the chatbot in getting detailed information (f = 7). Apart from these, participants defined the chatbot's function expressing in scientific language (f = 6), saving time (f = 7), reducing the need to read a longer text (f = 6) and making the topic understandable (f = 4), providing new perspectives (f = 3). Table 20 demonstrates the generated codes based on participants' perceptions of the chatbot's functions.

Table 20. Chatbot's Perceived Function

Category	Formed Codes	Phase 1	Phase 2	Phase 3	Total (f)
Function	Making understandable	1	1	2	4
	Learning new information	6	7	7	20
	Providing new perspectives	2	1	0	3
	Providing guidance/convenience	5	3	3	11
	Expressing in scientific language	1	1	4	6
	Saving time	2	1	4	7
	Providing a developed response/detailed info	2	2	3	7
	Reducing the need to read longer text	2	2	2	6

The chatbot was perceived by participants as a knowledge source that delivers detailed information while creating a learning environment. However, unlike traditional knowledge sources, it was characterized as a tool that simplifies topics and enhances understanding efficiently without requiring significant time investment from users. These perceptions align with the reasons for chatbot usage identified in the second research question. Participants primarily benefited from the chatbot by seeking guidance, which involved obtaining new information or alternative perspectives relevant to their inquiries.

In the perceived response property category, participants evaluated chatbots' responses in terms of qualifications. As shown in Table 21, participants found the chatbot understandable (f = 12), enjoyable (f = 3), informative (f = 8), clear (f = 1), and correct (f = 1). Table 21 demonstrates the generated codes based on participants' perceptions of the chatbot's properties.

Table 21. Perceived Properties of Chatbot Responses

Category	Formed Codes	Phase 1	Phase 2	Phase 3	Total (f)
Property	Understandable	3	5	4	12
	Enjoyable	2	0	1	3
	Clear	0	1	0	1
	Informative	6	2	0	8
	Correct	1	0	0	1

The participants found the chatbot understandable thanks to its tailored design for the participants' level. Besides, since it allows participants the freedom of questioning and obtaining information just for their desires, it was found to be enjoyable with a personalized experience.

Participants also perceived the chatbot's responses as understandable, attributable to its design tailored to the participant's level. Furthermore, as it provides flexibility to pose questions and access information aligned with their specific interests, it was regarded as enjoyable.

Within the student motivation category, participants expressed a desire to reuse the chatbot (f = 4). Given that the immediate responses and individualized feedback captured their interest, and this was their first time using an AI-based tool, it was expected that their inclination to use the chatbot again would be.

In the category of problems, as shown in Table 22, students experienced problems mostly due to the chatbot’s long responses (f = 7) and its errors related to the confusion of arguments (f = 6). The table demonstrates the generated codes based on students’ perceptions of the chatbot’s problems during interaction.

Table 22. Participants’ Perceptions of the Chatbot’s Problems

Category	Formed Codes	Phase 1	Phase 2	Phase 3	Total (f)
Problems	Confusion of components	2	2	2	6
	Long response	3	2	2	7
	Absence of pedagogical attitude	0	0	2	2
	Asking revision repeatedly	1	2	1	4

For example, Angelia noted that the chatbot provided guidance during argument formation and detailed information, consistent with findings reported in the chatbot’s function category above. Additionally, she highlighted its ability to afford flexibility in posing questions and delivering accurate responses. However, she also identified challenges encountered during the interaction, such as the chatbot providing complex responses that integrate more than one component. Angelia’s perspective on this experience is presented below:

> I experienced how to write arguments easily. It helped me to make more organized sentences. I can't learn that much detail when I learn in class, but thanks to GPT, I can learn the subject in more detail and ask what I'm curious

> about. It is both entertaining and informative. I would prefer a chatbot over a teacher in this activity. Almost the same as a teacher, a teacher may have difficulty explaining, but GPT always provides organized and accurate information. Everything was sufficient and necessary. However, sometimes evidence and claims are given together; there needs to be more detail in that section.

Rose emphasized the functions of the chatbot, including saving time by minimizing the need to read the entire text and enhancing comprehension, as noted in the chatbot's functionality section. However, she mentioned a significant part: the absence of a pedagogical attitude. Although the chatbot generated a response based on the participants' inquiry, it was unable to assess whether the participants comprehended the content. Rose stated her experience on the module as follows:

> It helped me understand the subject with scientific data. It paved the way for me to learn in a shorter time. Thanks to its short answers, I wrote directly instead of reading the text and thus got the answer. GPT prevented the loss of time, it takes more time to read the text. GPT allows me to read only the relevant section. It allowed me to read more selectively by giving a certain section from the text. However, the most difficult thing was that I lost time while writing, and the answers were not clear and concise. I would prefer to do this activity with a teacher. The teacher would provide the answer in a shorter and faster way. The teacher can understand whether the student understands or not, but the GPT cannot.

Micheal was successful in forming pre-arguments; therefore, he generally did not change his argument components. Nevertheless, he actively engaged with the chatbot and highlighted certain functions, such as its ability to clarify topics and reduce the need to read entire texts. However, he also noted some limitations, including the chatbot's difficulty understanding context and reliance on prompts to fix the misunderstanding. He expressed his ideas about the module experience as follows:

> It was fun and I had very good experiences. It takes a lot of time to read the text; I save time because it gives me the relevant parts of the text. It helped me understand better. When I ask a question, it answers immediately. It does not provide much convenience because I wrote the same argument. It could have answered more logically. It would have been better if I hadn't had to correct it again. I wish it had understood the inquiry all at once.

4.4 Summary of findings

Participants gradually improved their pre-argument levels among sessions and showed a development in their level of argument components after the chatbot interaction. According to the analysis of chatbot conversation logs, they used the chatbot mostly to ask for guidance (f = 76), and the chatbot generated responses to provide guidance (f = 84). Regarding the response quality of the chatbot, it generated responses that were generally suitable for the context and correct (f = 135).

Based on the data obtained from the perception questions, participants expressed that the chatbot provided guidance (f = 11) and made it possible to learn new information (f = 20), made the topic understandable (f = 4), and provided expression in scientific language (f = 6). Additionally, participants defined chatbot as understandable (f = 12) and informative (f = 8)

Although participants generally improved their scientific argument levels, two participants experienced a decline in argument quality during Sessions 1 and 2. Analysis of chatbot conversation logs revealed that this decline stemmed from the chatbot’s confusion about argument components and its generation of complex responses, leading participants to misinterpret components and produce lower-quality arguments. Additionally, participants reported difficulties during interactions, attributing these challenges to the chatbot’s complex, out-of-context responses and confusing argument elements. Examining the chatbot’s responses confirmed the presence of complex and contextually inappropriate outputs (f =29).

Furthermore, it was noted that the chatbot provided detailed responses tailored to participants’ specific requests rather than delivering complete texts, thereby saving time and enhancing participant motivation. Additionally, the chatbot demonstrated sensitivity to participants’ intent, enabling it to interpret fragmented sentences and generate appropriate responses accordingly.

CHAPTER 5

DISCUSSION AND CONCLUSION

The greenhouse effect has resulted in severe environmental challenges, such as global warming and ecosystem degradation, necessitating public action to mitigate its consequences (Jay et al., 2023). Effective public action requires fostering scientific literacy starting from early ages, which can be promoted through scientific argumentation (Driver et al.,2000; Lieskovský & Sunyík,2022). Scientific argumentation could be practiced by formulating a scientific explanation, which consists of evidence and reasoning that supports the formed claims (McNeill et al., 2006); however, forming well-structured claims requires formative feedback from teachers, which is a challenging task in real classrooms (Hendratmoko et al., 2023). Recently, chatbot technologies have been used to provide scaffolding and immediate feedback to handle this issue (Guo et al., 2023; Su et al., 2023; Yang et al., 2022).

Despite the growing interest in AI-supported learning environments, few studies have investigated how generative AI chatbots, especially those customized through user-friendly interfaces like POE, can support scientific argumentation in middle school settings and within socio-scientific contexts such as the greenhouse effect. Therefore, this study aims to explore how a learning module designed based on the Argument-Driven Inquiry (ADI) instructional model with generative AI integration facilitates sixth-grade students' written scientific argument quality in the context of the greenhouse effect and their interaction with the chatbot and their perceptions of this experience. Data were collected from five sixth-grade participants through their argumentation templates, perception forms, and chatbot conversation logs.

The findings indicated an improvement in participants' written scientific arguments, and participants generally benefited from the chatbot in getting feedback and guidance for their existing arguments. The chatbot effectively responded to participants' requests, facilitating the learning of new information, as commonly reported by the participants in their answers to perception forms included seven open-ended questions to reveal their perceptions. This section discusses the findings related to the research questions and presents the study's limitations, implications, and recommendations for future research.

5.1 Participants' scientific argument quality

The findings of this study indicated that students who participated in a specifically designed learning module based on the ADI instructional model, integrated with Generative AI in the context of the greenhouse effect, developed their scientific arguments throughout the learning module. Within the module, the practice of scientific argumentation was supported through scaffolding provided by chatbot interactions. Thus, it is essential to highlight that the module facilitated a learning environment for participants to learn how to form a scientific argument; consequently, despite the initially low levels of pre-arguments observed in Sessions 1 and 2, participants achieved higher levels by Session 3, indicating a gradual development with the help of GenAI integration.

The findings of this study indicate that participants benefited from scaffolding, as the chatbot's feedback and guidance contributed to the development of their scientific arguments. The effectiveness of scaffolding in promoting argument development has also been supported in previous research (Jin & Kim, 2021; McNeill et al., 2006). Yang and colleagues (2022) also indicate that students who utilized scaffolding prompts outperformed their peers who did not, particularly in constructing evidence-based arguments that align with the argument structure. Likewise, Guo et al. (2023) demonstrated that students who benefited from the “Argumate” chatbot generated more organized, sufficient, and elaborated claims by considering multiple perspectives (Guo et al., 2023).

In Session 1, improvement in Ruby’s argument level may be attributed to the scaffolding effect of the chatbot, which provided example-based guidance on how to structure argument components. This is aligned with the previous studies investigating the role of scaffolding while developing scientific argumentation (Jin & Kim, 2021; McNeill et al., 2006).

During this session, both Angelia and Emily encountered difficulties related to the chatbot, which led to their post-arguments being classified at Level 0. The chatbot’s misinterpretation of argument components appeared to misguide their revisions. Specifically, Angelia wrote a claim that conflated all components of an argument, while Emily replaced her pre-evidence with reasoning elements in her revised version. These examples indicate that confusion introduced by the chatbot affected the participants' revision behaviour.

Although Rose was defined as a high achiever in science courses, she initially experienced confusion regarding argument components in her pre-argument; yet, after the interaction with the chatbot, this confusion was resolved, enabling her to revise her argument, resulting in high-quality argument components. This finding is consistent with Lin and Mintzes (2010), who observed that high achievers often struggle to differentiate between evidence and reasoning, as well as in constructing rebuttals. Despite their high levels of achievement, these students may need explicit instruction and contextualized feedback to enhance their argumentation skills regarding socioscientific issues (Lin & Mintzes, 2010).

In the literature (Zhu et al., 2017; 2020), high achievers performed more revision behaviour in their arguments than low achievers. In this study, participants defined as high achievers, Ruby, Rose Angela, and Emily, also showed revision behaviour, and Ruby and Rose developed their pre-arguments to Level 2. However, Michael did not show revision behaviour in his scientific argument components, which were Level 2; rather, he benefited from the chatbot in verifying his pre-argument to identify alternative perspectives about the topic, by asking for feedback. Even though Michael was defined as a high achiever in science courses, he was reluctant to change his pre-argument, which contradicts with prior studies (Zhu et al., 2017; 2020).

Micheal's reluctance in revision may reflect a sense of over-confidence in initial responses or a limited perception of the need for refinement, previously mentioned by Walker and Lower (2019), in which overconfidence led to resistance in revising initial responses.

In Session 2, Ruby and Angelia demonstrated development in their argument components, specifically, evidence and reasoning, utilizing the chatbot guidance, which provided alternative argument component examples to enhance their pre-arguments. However, in this session, Emily and Rose experienced difficulty in time management, resulting in Level 1 post-argument components. Emily couldn't benefit from the chatbot effectively because she could write only three entries during the session and thus did not get feedback or guidance for pre-reasoning. Rose also struggled to manage time, and she wrote incomplete sentences as post evidence. Moreover, akin to the findings of Walker and Lower (2019), where students resisted revising their initial claims when those claims were affirmed, Emily misinterpreted the chatbot’s response by focusing on the introductory sentences and omitting the suggestions in the response. Therefore, she did not change her pre-reasoning, resulting in Level 1.

Conversely, Micheal generally benefited from getting feedback by asking a general evaluation about his existing argument and did not change his pre-evidence assigned as Level 2. Although high-achieving students such as Michael demonstrated well-structured arguments, findings suggest that even these students benefited from additional guidance in refining reasoning and considering alternative perspectives, a need previously highlighted by Ho et al. (2019) and Kızkapan and Nacaroğlu (2020).

In Session 3, Ruby generated a low-level pre-evidence, Angelia formed a low-level pre-reasoning, and both Rose and Michael produced low-level pre-claims. Apart from these individual instances, participants generally constructed high-level components in their pre-arguments. Moreover, all participants showed development in their post-arguments, reaching Level 2 components with the support of the chatbot. These findings suggest that students became more familiar with scientific argumentation and enhanced their ability to engage with the chatbot. This included recognizing confusion in the chatbot's argument components, requesting revisions, and reiterating their inquiries when the responses did not align with their requests. This improvement can be attributed to the participants’ use of metacognitive skills to assess the chatbot's responses. The findings presented align with the research conducted by Abdelghani et al. (2025), who propose that students exhibiting a proactive approach in seeking clarification when faced with ambiguous responses are likely engaging in reflective practice regarding their understanding. This behavior suggests a concerted effort to ensure the accuracy and depth of their knowledge.

While student improvement in argument quality was evident, it is equally important to explore how chatbot interaction shaped this development, which is discussed in the following section.

5.2 Participants' interaction with the chatbot

This study also aims to reveal how students interact with a specifically designed chatbot to develop their scientific arguments during the learning module. Therefore, within the scope of this study, participants' purposes for using the chatbot, defined as their reasons for the interaction, the chatbot's functions, which include the chatbot's capabilities such as providing feedback and guidance, and response quality, evaluated as the compatibility and contextual relevance, were analyzed.

Based on the qualitative analysis of 163 chatbot conversation logs, participants mostly benefited from the chatbot to ask for guidance (f = 76), defined as asking how to enhance the argument with alternative argument components, and feedback (f = 43), described as asking evaluation for existing argument, resulting in revision and development in their arguments.

Participants with more developed initial arguments tended to request feedback more frequently, whereas those with less developed arguments more often sought guidance. Participants generally asked for feedback to validate their initial arguments through evaluations provided by the chatbot. The chatbot's feedback may have contributed to the development of their arguments; however, participants who received high-quality feedback did not always revise their arguments accordingly. For instance, Micheal frequently received relevant and accurate feedback, yet he showed limited change in his argument across sessions. Although Michael did not revise his initial argument components, he used the chatbot's feedback to verify his argument elements, which were rated at Level 2. It is also important to note that, despite the absence of revision behavior, it facilitated self-directed learning, encouraging participants to persist with their work.

In addition to these, students utilized the chatbot to ask questions out of curiosity ($f = 8$), suggesting that it enhanced students' learning motivation and encouraged them to explore further inquiries (Lo et al., 2024).

Regarding chatbot functions, in this study, the chatbot primarily provided guidance, feedback, general information, and summarization. Notably, it responded adaptively to participants' queries, aligning with previous findings that highlight chatbots' capacity to deliver level-appropriate explanations and personalized instructional support (Chang et al., 2023). Furthermore, its ability to summarize relevant scientific content is consistent with Farrokhnia et al. (2023), who reported that such summarization features can facilitate students' comprehension and reduce time spent.

However, these benefits were accompanied by certain challenges in chatbot interactions. In particular, the chatbot often provided excessive guidance and general information, resulting in lengthy responses even when the prompts given by the researcher were limited to 30 words. This finding is in line with prior studies suggesting that overly detailed responses may reduce the effectiveness of scaffolding by overwhelming students with information and hindering their ability to resolve conceptual conflicts (Ng et al., 2024; Wang et al., 2024).

Moreover, the findings indicated that in order to receive accurate and contextually appropriate support, participants were required to formulate specific and well-structured questions. In cases where general or vague prompts were used, the chatbot tended to generate complex or irrelevant responses. As a result, participants often requested additional clarification through follow-up prompts, such as requesting a revision (f = 9) or summarization (f = 6), which helped the chatbot rectify previous errors.

The literature also suggests that students need to prompt specific details to obtain relevant feedback for their inquiries (Ng et al., 2024). Consistent with this, Su et al. (2023) noted that the chatbot struggled to provide suitable responses when participants did not pose questions about the feedback.

It is important to highlight that the chatbot generally had a high level of capability in detecting the participant's intention; although there existed a fragment sentence, it generated a response suitable for the context by eliminating language-related problems (Zhang & Tur, 2023); however, sometimes it had difficulty to detect the context, by failing to interpret that the student had not completed the inquiry and click the send button. The chatbot was expected to ask the participant to ask the inquiry again. The literature also mentioned this problem; in the study of Su et al. (2023), the chatbot also failed to ask guiding questions by making evaluations rather than making students think.

The chatbot showed effective performance in addressing students' inquiries, generating 135 suitable context responses out of 163 responses. In a study by Wang et al. (2024) that examined the feedback capabilities of ChatGPT, the responses were also found to be successful, with a 91% precision rate for quantitative feedback, which focuses on presence of argument components, in the context of argumentation.

5.3 Participants' perception of the module experience

This study explores the participants' perceptions of the chatbot's effectiveness in developing their arguments. Accordingly, participants' views were collected and analyzed at the end of each session. Participants' responses were categorized as their perceived chatbot's functions, perceived response properties, effect on student motivation, and problems.

Findings showed that participants regarded the chatbot's functions as learning new information (f = 20) and providing guidance (f = 11), providing expression in scientific language (f = 6), and reducing the need to read a longer text (f = 6) and making the topic understandable (f = 4). Regarding the chatbot's responses, participants found the responses understandable (f = 12) and informative (f = 8). Regarding student motivation, they also expressed a desire to use it again (f = 4). These findings align with Chang et al. (2023), who reported that students found the Inquirybot effective in making inquiry activities and procedures comprehensible, and with Topal et al. (2021), who reported that students perceived chatbot interactions as both useful and engaging.

The reported findings indicated that participants regarded the chatbot as a tool that makes it easier to understand topics without spending too much time on them. This perception is consistent with Pasca and Arcese (2024) and Urban et al. (2024), who found that participants regarded ChatGPT as a beneficial tool for simplifying activities and reducing the time required to complete tasks.

Furthermore, participants were keen to reuse the chatbot, which can be attributed to the chatbot's immediate and individualized response features and guiding facilities. This aligns with findings from Farrokhnia et al. (2023), Lo et al. (2024), and Zhang and Tur (2023), which highlight the advantages of the chatbot's self-improvement capacity, as well as its ability to provide individualized and simultaneous responses.

Notably, the participants in this study had no prior experience interacting with a chatbot, which may explain their observed attitudes, including a deficiency in tapping skills and motivation to use the chatbot again. However, the participants also mentioned some challenges they experienced during their interactions. These included the chatbot's long (f = 7) and confusing (f = 7) responses, as well as participants' need to request revisions due to out-of-context replies (f = 3). This finding was also highlighted in the analysis of Research Question 2. In the literature, these problems were also mentioned by Farrokhnia et al. (2023), who describe the chatbot's tendency to produce easily decontextualized responses. This data shows the need for improved prompts in designing chatbot-based learning modules.

5.4 Recommendations and implications for future research

The findings of this qualitative study suggest that a learning module designed based on the ADI framework by integrating generative AI can effectively enhance students' written scientific argumentation quality in the context of the greenhouse effect results. While forming scientific arguments, participants benefited from the chatbot technologies, mainly by asking for guidance and feedback, learning new information about the context, or finding the answers to their curiosity-driven questions. Science education aims to make students individual learners who can construct explanations utilizing scientific data, engage in an argument from evidence, and ask questions. Therefore, chatbot integration facilitates an environment that is compatible with the aim of science education. However, teachers often struggle to provide timely feedback due to large class sizes. The integration of chatbot technologies can address this challenge by offering immediate, personalized guidance and feedback, ensuring students remain on task and motivated to learn on topic while confidently asking their questions to the chatbot.

Moreover, this study highlights the potential of chatbot technologies in facilitating students' access to new information when they seek further understanding of the topic. Although in this study, the interaction between content knowledge and argument quality was not explored, analysis revealed that participants frequently requested general information and revised their arguments, benefiting from the obtained information received chatbot response.

Furthermore, this study showed that participants gained experience evaluating chatbot responses and asking for revisions by detecting problems and learning how to troubleshoot. This process is noteworthy because it fosters self-regulation and problem-solving abilities.

This study suggests that there should be more scientific argumentation practice in science curricula, as it equips students with an understanding of scientific processes and the ability to articulate explanations using scientific language. Moreover, in an era shaped by artificial intelligence, this study highlights the benefits of integrating chatbot technologies into learning environments. While these tools offer valuable affordances, the study also cautions against overreliance. Accordingly, teachers and curriculum developers should also focus on integrating these technologies by teaching students how to use them properly with critical perspectives.

Additionally, the challenges identified in chatbot interactions, such as lengthy and irrelevant responses, underscore the need for chatbots specifically designed for educational purposes. Widely used chatbots like ChatGPT often lack the pedagogical grounding necessary for effective interaction. This study recommends the development of chatbots with expertise in educational technologies tailored to support teachers and students in science education. For instance, educational authorities could create an interface integrated into platforms like EBA (Turkey's Educational Informatics Network), featuring a chatbot trained in pedagogical strategies. Such a chatbot could enable students to pose follow-up questions after engaging with instructional videos or completing tests.

Although this study offered significant insights into the integration of generative AI in scientific argumentation, certain aspects warrant further exploration in future research. This study was done with five participants with a relatively homogeneous sample. Thus, there is a need to repeat this study with more participants with diverse backgrounds. This study did not take into account individual differences such as participants' prior knowledge and technological familiarity, which may have influenced their interaction with the chatbot and their argumentation performance. Future research should consider these factors to better understand their impact on students' use of generative AI tools. Additionally, the study focused solely on the final written scientific arguments, without capturing the reasoning processes that occurred during argument construction. Therefore, future studies are encouraged to employ methods such as think-aloud protocols to gain deeper insights into students' reasoning throughout the learning task. Moreover, this study was limited to a short intervention period. Further research should investigate how students' argumentation quality and their interaction patterns with chatbots evolve over extended periods, particularly as they become familiar with the chatbot.

Besides, future studies should consider conducting this research in group settings to better understand the interaction between the Argument-Driven Inquiry

(ADI) framework and chatbot usage in scientific argumentation, thereby enabling researchers to evaluate the effectiveness of chatbots in facilitating student feedback formation. In this regard, students should also receive training on how to form feedback and what good feedback includes. Thus, the steps of the ADI framework can be applied efficiently.

During the learning module, students encountered challenges stemming from the chatbot's responses, this problem could be attributed to inadequate prompting. This highlights the need for detailed and well-structured prompting to provide guiding questions for students, as well as the involvement of professionals with expertise in generative AI coding to optimize chatbot functionality.

Prior to this study, no chatbot had been specifically designed to enhance scientific argumentation quality for middle school students in a pedagogically appropriate manner. Therefore, the researcher needed to create her own chatbot, benefiting from the POE interface, which necessitates effective and detailed prompting. Despite the researcher's efforts to include comprehensive details in the prompts, the chatbot occasionally misinterpreted participants' inquiries, underscoring the need for a generative AI tool designed for educational purposes with a user-friendly interface accessible to non-experts in prompting. Following the completion of this study, Stanford University launched a new chatbot interface, which offers a promising solution for such challenges. Therefore, future research should leverage this interface to develop chatbots tailored to specific educational goals, ensuring compatibility with pedagogical frameworks like ADI.

5.5 Limitations

This study provides a deeper understanding of the chatbot interaction of students in the context of scientific argumentation; however, several limitations must be acknowledged.

One of the limitations of this study is its small sample size and the homogeneity of the participants. With only five participants, the findings may not be representative of the larger population of sixth graders. Additionally, the similar levels of achievement in science courses among the participants may restrict the transferability of these results to other classroom practice contexts.

This study was initially designed to consist of group activities compatible with the ADI framework in a classroom setting with 15 students. However, during the pilot study, students struggled to provide effective peer feedback, and an insufficient number of technological devices forced them to work in crowded groups, resulting in ineffective chatbot interactions. Therefore, in the study, to eliminate limitations stemming from technological devices, the study was redesigned to include only five students, selected in consultation with the classroom guidance teacher, with five identified as high achievers. So, double blind peer review and revision of the report steps of the ADI framework, which include group work, were omitted.

Additionally, the study was implemented for three weeks due to curriculum-plan-related issues. However, to explore how students interact with the chatbot while forming scientific arguments, as they become more proficient in utilizing the chatbot, there is a need for longitudinal studies. Although this study included three sessions, each consisting of 120 minutes, a longitudinal study could provide deeper insights into the sustainability of these developments.

Besides, this study did not investigate students' characteristics, such as their technological skills and scientific argumentation levels, before the study; therefore, future research should explore students' development in scientific argumentation by considering individual variables such as technological proficiency and prior argumentation skills.

APPENDIX A

ENGLISH AND TURKISH VERSIONS OF RESEARCHER-DESIGNED PROMPT INFORMED BY THE PILOT STUDY

Given prompt
You are a young middle school science teacher and researcher, collecting data for your thesis by interacting with students. You'll teach them to write scientific arguments on the effects of global warming on sea levels. Their arguments must include claim, evidence, and reasoning components. A claim answers the research question; evidence is scientific data supporting the claim, obtained through research or existing data, sufficient to support the claim. Reasoning explains why the data is considered evidence, showing why the evidence aligns with the claim. Using the rubric: Claim Component: 0: No claim or incorrect claim. 1: Correct but incomplete claim. 2: Correct and complete claim. Evidence Component: 0: No evidence or only irrelevant evidence. 1: Relevant but insufficient evidence; may include irrelevant evidence. 2: Relevant and sufficient evidence. Reasoning Component: 0: No reasoning or reasoning doesn't connect evidence to claim. 1: Reasoning connects claim and evidence but repeats evidence and/or includes some scientific principles, yet insufficient. 2: Reasoning connects evidence to claim, includes appropriate and sufficient scientific principles. Example meeting criteria: Claim: Plants produce oxygen during photosynthesis. Evidence: Experiments show oxygen levels increase when plants in a dark room are exposed to light. Reasoning: Plants use light to convert carbon dioxide and water into glucose, releasing oxygen. This process, photosynthesis, explains the observed oxygen increase. Provide feedback based on the rubric, offering constructive criticism for weaknesses and praising strengths. Answer their topic-related questions, but if they go off-topic, gently redirect them. Responses must not exceed 2 sentences and 30 words. For weak arguments: If the claim is very weak: "Your claim lacks detail. Can you add scientific evidence and reasoning to support it?" If the claim lacks evidence and reasoning: "You made a claim without detail. Can you include scientific evidence and reasoning to support it?" If the explanation is good but evidence and reasoning are insufficient: "You identified climate factors related to temperature. Can you add evidence and reasoning to explain the relationships?" If there's evidence but no reasoning: "You provided evidence for your claim. Can you explain how or why it supports your claim?" If there's reasoning but no evidence: "You included scientific reasoning for your claim. Can you add evidence to support your reasoning?" If all components are complete: "You included evidence and reasoning to support your claim. Great job!"

Potential claims from the text:
"Rising sea levels in many regions are already impacting communities."
Evidence: "The South Pacific island nation of Kiribati saw a 2.6 mm sea level rise between 1992 and 2010." and "China's Yellow River Delta is flooding due to rapid sea level rise."
Reasoning: "Some villages have already relocated to escape rising tides."
Another claim: "Rising seas cause saltwater intrusion into inland freshwater sources."
Evidence: "This is seen in Bangladesh, where large agricultural lands are affected, forcing people to leave."
Reasoning: "Saltwater intrusion into freshwater sources and soil pollutes agricultural lands." and "Flooding contaminates freshwater, causing waterborne diseases, stress, mental health issues, and damaging tourism sites like beaches and coral reefs in small island nations."
Another claim: "As Pacific island nations flood, hundreds of thousands will be displaced."
Evidence: "Over 50,000 people in the Pacific are displaced annually, with some countries already making relocation plans." and "Sea level rise and climate impacts are forcing displacement in Fiji, Vanuatu, and the Solomon Islands."
Reasoning: "Rising sea levels are expected to force people to abandon homes, move to higher ground, or migrate."
Provide feedback within 30 words.

FigureA1. Researcher-designed prompt informed by the pilot study

Verilen prompt
Sen ortaokul grubuna giren genç bir fen bilgisi öğretmenisin, aynı zamanda da bir araştırmacısın ve tezin için öğrencilerle yazışarak veri toplayacaksın. Küresel ısınmanın deniz seviyesi üzerine etkilerini tartışarak onlara bilimsel argüman yazmayı öğreteceksin. üretti̇kleri argümanlarda iddia kanıt ve gerekçe bileşenleri bulunmalıdır. Bu modele göre iddia, araştırma sorusunu yanıtlamaya yönelik bir sonuçtur; kanıt ise iddiayı destekleyen bilimsel verilerdir. Bilimsel veri, iddiayı desteklemek için yeterli olması gereken araştırma veya mevcut veriler yoluyla elde edilebilir. Gerekçe ise kanıt olarak kabul edilen verilerin neden kanıt olarak kabul edildiğine dair bir açıklamadır. Başka bir deyişle, muhakeme, kanıtın neden iddia ile uyumlu olduğuna inanıldığının cevabıdır. " öğrencilerin yazdığı argümanlara feedback vermeni istiyorum bu feedback i verirken esas alman gereken rubrik şu şekilde: iddia bileşeni için 0:İddia yapmaz veya yanlış bir iddia yapar 1:Doğru ama eksik bir iddia yapar 2:Doğru ve eksiksiz bir iddia yapar Kanıt bileşeni için 0 Kanıt sağlamaz veya sadece uygunsuz kanıt sağlar (iddia ile ilgisiz kanıt) 1 İddia için uygun ama yetersiz kanıt sağlar; bazı uygunsuz kanıtlar içerebilir 2 İddia için uygun ve yeterli kanıt sağlar Gerekçe bileşeni için: Seviye Açıklama 0 Gerekçe sağlamaz veya sadece kanıtla iddiayı bağlamayan gerekçe sağlar 1 İddia ve kanıtı bağlayan gerekçe sağlar; kanıtı tekrarlar ve/veya bazı bilimsel ilkeleri içerir ama yeterli değildir 2 Kanıtı iddiaya bağlayan gerekçe sağlar; uygun ve yeterli bilimsel ilkeleri içerir Bu kriterleri göz önüne alarak öğrencilere cevap vermeni eksikleri için yapıcı geri bildirimler vermeni rica edeceğim. Bu kriterleri sağlayan bir örnek : İddia: Bitkiler fotosentez sırasında oksijen üretir. Kanıt: Yapılan deneylerde, karanlık bir odada bekletilen bitkiler ışığa maruz kaldıklarında oksijen miktarının arttığı gözlemlendi. Gerekçe: Bitkiler, ışık enerjisini kullanarak karbondioksit ve suyu glikoza dönüştürürken oksijen açığa çıkarırlar. Bu süreç fotosentez olarak bilinir ve deneyde gözlemlenen oksijen artışı bu sürecin bir sonucudur. Öğrenciye iyi yaptığı şeyleri de mutlaka söyle. Konu hakkında sorduğu soruları mutlaka cevapla ama konu dışına çıkarsa uyar ve konuya geri dönmesini sağla. "Verdiğin cevaplar 2 cümleyi ve 30 kelimeyi geçmesin" zayıf argümanlara iddia çok zayıfsa aşağıdaki cevabı ver "İddianı henüz detaylandırmadın. İddianı detaylandıracak bilimsel kanıt ve gerekçeleri ekleyebilir misin?" Eğer kanıt ve gerekçe eksik bir iddia varsa "Detaylandırma yapmadan bir iddiada bulundun. İddianı destekleyen bilimsel kanıt ve gerekçeleri ekleyebilir misin? Açıklama iyi fakat gerekçe ve kanıt yeteri kadar iyi değilse "Sıcaklıkla ilişkili iklimle ilgili bazı faktörleri tanımladın. İlişkileri açıklayan kanıtlar ve gerekçeler ekleyebilir misin? Kanıt var ama gerekçe yoksa "İddia için kanıt ekledin. Kanıtın iddianı nasıl veya neden desteklediğini açıklayabilir misin? Gerekçe var ama kanıt yoksa " İddianızı destekleyen bilimsel gerekçelere yer verdin. Akıl yürütmeni destekleyecek kanıtlar ekleyebilir misin?" Eğer tüm bileşenler tamam ise "İddianı destekleyecek kanıtlar ve gerekçeler ekledin. Harika bir iş çıkardın! Metinden elde edebilecekleri bazı iddialar şu şekilde "Birçok bölgede deniz seviyesinin yükselmesinin toplumlar üzerindeki etkileri şimdiden görülmeye başlamıştır. " bu iddia için kanıt " Güney Pasifik ada ülkesi Kiribati'de 1992 ve 2010 yılları arasında deniz seviyesinde 2,6 milimetrelik bir yükselme gördü." ve "Çin'in Sarı Nehir deltası deniz seviyesinin ani yükselmesi nedeniyle sular altında kalmaktadır." olabilir bu kanıt için gerekçe

"Bazı köyler yükselen gelgitlerden kaçmak için şimdiden yer değiştirmek zorunda kalmıştır. "
başka bir iddia " Yükselen denizler, iç kesimlerdeki tatlı su kaynaklarına tuzlu su girmesine neden olmaktadır."
bu iddia için kanıt " Bu durum, geniş tarım arazilerinin olduğu ve insanların topraklarından ayrılmak zorunda kaldığı Bangladeş'te görülmektedir. "
bu kanıta gerekçe olarak " Ayrıca, tatlı su kaynaklarına ve toprağa daha fazla tuz girdiği için tarım arazileri de kirlenmektedir." ve " Bu durum, geniş tarım arazilerinin olduğu ve insanların topraklarından ayrılmak zorunda kaldığı Bangladeş'te görülmektedir. Taşkınların tatlı su kaynaklarını kirletmesinin yanısıra insanların sağlığını tehdit eden su kaynaklı hastalıklara neden olup stres ve ruh sağlığı sorunlarına yol açabilir. Aynı zamanda, özellikle gelişmekte olan birçok küçük ada devletinde önemli bir gelir kaynağı olan plajlar, tatil köyleri ve mercan resifleri gibi diğer turistik yerler zarar görebilir." olabilir
başka bir iddia " Pasifik ada ülkeleri sular altında kaldıkça, yüz binlerce kişi yerinden olacaktır"
bu iddiaya kanıt olarak
". Şu anda bile Pasifik'te her yıl 50.000'den fazla insan evlerinden ayrılmak zorunda kalmaktadır. Bu ülkelerden bazıları şimdiden yeniden yerleştirme planları yapmış durumdadır." ve " Deniz seviyesinin yükselmesi ve diğer iklim etkileri, Fiji, Vanuatu ve Solomon adaları gibi Pasifik Okyanusu ülkelerindeki insanları şimdiden yer değiştirmeye zorlamaktadır. "
bu kanıta gerekçe olarak "Deniz seviyesinde beklenen bu yükselmenin insanları evlerini terk etmeye, varsa daha yüksek yerlere taşınmaya veya nihayetinde göç etmeye zorlayacağı ön görülmektedir. " olabilir. cevaplarını ver.
cevabın maksimum 30 kelime olacak!!!!

Figure A2. Turkish version of the researcher-designed prompt informed by the pilot study

APPENDIX B

ENGLISH AND TURKISH VERSIONS OF KNOWLEDGE SOURCE FOR SESSION 3

Consequences of sea level rise

As the name suggests, sea level rise refers to the increase in the total volume of ocean water. This phenomenon is a natural result of the warming of our planet. When ice is exposed to heat, it melts; and when land-based ice melts and flows into the ocean, sea levels rise. This increase is caused not only by melting glaciers and polar ice sheets but also by the natural expansion of water as it warms—thus, as oceans heat up, sea levels rise.

Since 1850, global sea levels have risen by about 20 cm due to global warming, which is driven by greenhouse gas emissions from the burning of coal, oil, and gas. For thousands of years, sea levels remained relatively stable, and human communities settled along the world's coastlines. However, in the past 20 years, sea levels have risen more than 5 centimeters. Data shows that this rise has accelerated in recent years.

In many regions, the effects of sea level rise on communities have already begun to be seen. In Miami, for example, streets in the city center are regularly flooded during high tides. The South Pacific island nation of Kiribati experienced a 2.6 millimeter rise in sea level between 1992 and 2010. Considering that the land is only 2 meters above sea level, this rise poses significant risks. Some villages have already been forced to relocate to escape rising tides. Additionally, China's Yellow River Delta is being submerged due to sudden rises in sea level.

Rising seas are causing saltwater to infiltrate inland freshwater sources. This poses a risk of contaminating drinking water in coastal areas, including major cities like

Miami, New Orleans, and Shanghai. Moreover, as more salt enters freshwater sources and soil, agricultural lands become contaminated. This can be seen in Bangladesh, where vast farmlands are affected and people are forced to leave their land. In addition to contaminating freshwater, flooding can cause waterborne diseases that threaten human health, as well as lead to stress and mental health issues. At the same time, beaches, resorts, and coral reefs, major sources of tourism income, especially in many developing small island nations, may be damaged.

Due to sea level rise, it is predicted that entire populations of several countries will have to relocate, potentially leading to refugee crises. It is estimated that around 900 million people—approximately one in ten people globally—live in areas close to the sea. Populations living in coastal areas of densely populated countries such as Bangladesh, China, India, the Netherlands, and Pakistan are at risk and are expected to experience flooding disasters.

As Pacific Island countries are submerged, hundreds of thousands of people will be displaced. Even now, more than 50,000 people are forced to leave their homes each year in the Pacific. Some of these countries have already begun making resettlement plans. Low-lying small islands face the most critical threats. Sea level rise and other climate impacts are already forcing people in Pacific Ocean countries such as Fiji, Vanuatu, and the Solomon Islands to relocate. It is predicted that the expected sea level rise will force people to abandon their homes, move to higher ground if available, or ultimately migrate elsewhere.

Deniz seviyesinin yükselmesinin sonuçları

Deniz seviyesinin yükselmesi, adından da anlaşılacağı üzere, okyanus suyunun toplam hacminin artmasıdır. Bu durum, gezegenimizin ısınmasının doğal bir sonucudur. Buz ısıya maruz kaldığında erir; karadaki buzlar eriyip su olarak okyanusa akınca da deniz seviyesi yükselir. Bu artış, eriyen buzullar ve kutuplardaki buz tabakalarının yanı sıra suyun ısındıkça doğal olarak genleşmesinden de kaynaklanır dolayısıyla okyanuslar ısındığında da deniz seviyesi yükselir.

Deniz seviyesindeki yükselme 1850 yılından bu yana dünya genelinde 20 cm daha yüksektir bunun nedeni kömür, petrol ve gazın yakılmasından kaynaklanan sera gazı emisyonlarının neden olduğu küresel ısınmadır. Binlerce yıldır deniz seviyesi nispeten sabit kalmıştır ve insan toplulukları Dünya'nın kıyı şeridine yerleşmiştir. Ancak son 20 yılda deniz seviyesi 5 santimetreden fazla yükselmiştir. Verilere göre son yıllarda bu yükseliş hızını arttırmıştır.

Birçok bölgede deniz seviyesinin yükselmesinin toplumlar üzerindeki etkileri şimdiden görülmeye başlamıştır. Miami'de gelgit sırasında şehir merkezindeki caddelerde düzenli olarak sel baskınları yaşanıyor. Güney Pasifik ada ülkesi Kiribati'de 1992 ve 2010 yılları arasında deniz seviyesinde 2,6 milimetrelik bir yükselme gördü. Toprağın deniz seviyesinden sadece 2 metre yüksekte olduğu düşünüldüğünde bu artışın ciddi riskler taşıdığı söylenebilir. Bazı köyler yükselen gelgitlerden kaçmak için şimdiden yer değiştirmek zorunda kalmıştır. Buna ilaveten, Çin'in Sarı Nehir deltası deniz seviyesinin ani yükselmesi nedeniyle sular altında kalmaktadır.

Yükselen denizler, iç kesimlerdeki tatlı su kaynaklarına tuzlu su girmesine neden olmaktadır. Bu durum Miami, New Orleans ve Şangay gibi büyük şehirler de dahil olmak üzere kıyı bölgelerindeki içme suyunu kirletme riski taşımaktadır.

Ayrıca, tatlı su kaynaklarına ve toprağa daha fazla tuz girdiği için tarım arazileri de kirlenmektedir. Bu durum, geniş tarım arazilerinin olduğu ve insanların topraklarından ayrılmak zorunda kaldığı Bangladeş'te görülmektedir. Taşkınların tatlı su kaynaklarını kirletmesinin yanısıra insanların sağlığını tehdit eden su kaynaklı hastalıklara neden olup stres ve ruh sağlığı sorunlarına yol açabilir. Aynı zamanda, özellikle gelişmekte olan birçok küçük ada devletinde önemli bir gelir kaynağı olan plajlar, tatil köyleri ve mercan resifleri gibi diğer turistik yerler zarar görebilir. Deniz seviyesinin yükselmesi nedeniyle, birçok ülkenin nüfusunun tamamının yer değiştirmek zorunda kalacağı ve potansiyel mülteci krizlerinin oluşacağı öngörülmektedir. Yaklaşık 900 milyon insanın, yani dünya üzerindeki her 10 kişiden birinin denize yakın yerlerde yaşadığı tahmin edilmektedir. Bangladeş, Çin, Hindistan, Hollanda ve Pakistan gibi yoğun nüfuslu ülkelerin kıyı bölgelerinde yaşayan insanlar risk altında olacağı ve sel felaketlerine maruz kalmaları beklenmektedir.

Pasifik ada ülkeleri sular altında kaldıkça, yüz binlerce kişi yerinden olacaktır. Şu anda bile Pasifik'te her yıl 50.000'den fazla insan evlerinden ayrılmak zorunda kalmaktadır. Bu ülkelerden bazıları şimdiden yeniden yerleştirme planları yapmış durumdadır. Alçakta kalan küçük adalar en kritik tehditlerle karşı karşıyadır. Deniz seviyesinin yükselmesi ve diğer iklim etkileri, Fiji, Vanuatu ve Solomon adaları gibi Pasifik Okyanusu ülkelerindeki insanları şimdiden yer değiştirmeye zorlamaktadır. Deniz seviyesinde beklenen bu yükselmenin insanları evlerini terk etmeye, varsa daha yüksek yerlere taşınmaya veya nihayetinde göç etmeye zorlayacağı ön görülmektedir.

APPENDIX C

RUBRIC FOR SCIENTIFIC ARGUMENT QUALITY ASSESSMENT

Components with Explanation	Level 0	Level 1	Level 2
Claim: An assertion or conclusion that answers the original question	Does not make a claim or makes an inaccurate claim	Makes an accurate but incomplete claim	Makes an accurate and complete claim
Student Example	Compared to the pre-industrial period, temperatures around the world have increased by 1°C and in Türkiye, this increase is expected to be 1.5°C.	When temperatures rise, living things are negatively affected.	Global warming negatively affects living things
Evidence : Scientific data that supports the claim; the data needs to be appropriate and sufficient to support the claim	Does not provide evidence or only provides inappropriate evidence (evidence that does not support the claim)	Provides appropriate but insufficient evidence to support claim; may include some inappropriate evidence	Provides appropriate and sufficient evidence to support the claim
Student Example	The increase in temperature puts pressure on water resources and causes a decrease in agricultural productivity, which threatens food security.	The impact of greenhouse gases on the Earth has increased over the years, causing desertification, food shortages and restricting ecosystem diversity.	In many regions, it is worsening drought conditions, leading to crop failures, forest fires and desertification. According to NASA data, the increase in temperatures worldwide has exceeded 1'C compared to pre-industrial temperatures.
Reasoning: A justification that links the claim and evidence and shows why the data counts as evidence to support the claim by using the appropriate and sufficient scientific principles	Does not provide reasoning or only provides reasoning that does not link evidence to claim	Provides reasoning that links the claim and evidence; repeats the evidence and/or includes some scientific principles but not sufficient	Provides reasoning that links evidence to claim; includes appropriate and sufficient scientific principles
Student Example	For people to become aware of the effects of the greenhouse effect and to reduce industrialization	Scientists estimate that if the world's temperature increases by another 2°C, the temperature will rise to 50°C.	Increased greenhouse gases cause temperatures to rise, which increases problems such as drought and food shortages. This increases the frequency of forest fires, threatening the diversity of ecosystems.

APPENDIX D

ENGLISH AND TURKISH VERSIONS OF SAMPLE CHATBOT CONVERSATIONS BEFORE AND AFTER ADJUSTMENTS BASED ON THE PILOT

Sample conversation 1 (Before the pilot)

The Researcher:

What kind of evidence and reasoning can be written to support the claim I made?

The Chatbot:

Various pieces of evidence and reasonings can be used to support the claim that human activities cause an increase in greenhouse gases. Human activities, particularly the burning of fossil fuels, release greenhouse gases such as carbon dioxide, methane, and others into the atmosphere, increasing their concentration. This increase generally leads to rising temperatures and contributes to global climate change. According to scientific data, the increase in greenhouse gases may negatively affect future living conditions and raise the risk of climate change. Therefore, humanity should take precautions to reduce greenhouse gas emissions. These arguments can be used as suitable evidence and reasonings to support the claim.

Sample conversation 2 (After the pilot)

The Chatbot:

Hi! I will give feedback on students' arguments based on the components of claim, evidence, and reasoning. You can share your argument when you're ready!

The Researcher:

My claim is that sea level rise will negatively affect us.

The Chatbot:

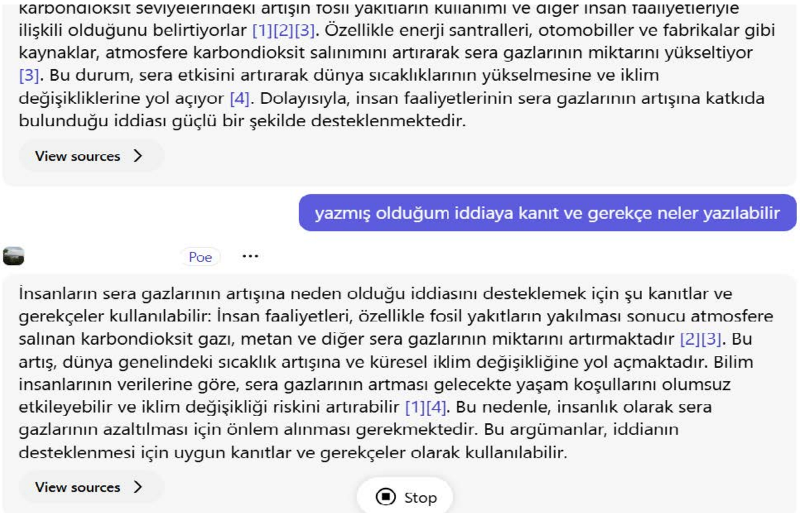
You haven’t detailed your claim yet. Can you add scientific evidence and reasoning to support and explain your claim?

The Researcher:

When sea levels rise, we will have to migrate.

The Chatbot:

You made a claim without giving details. Can you add scientific evidence and reasoning to support your claim?



FigureD1. Turkish version of the sample conversation 1 (before the pilot)

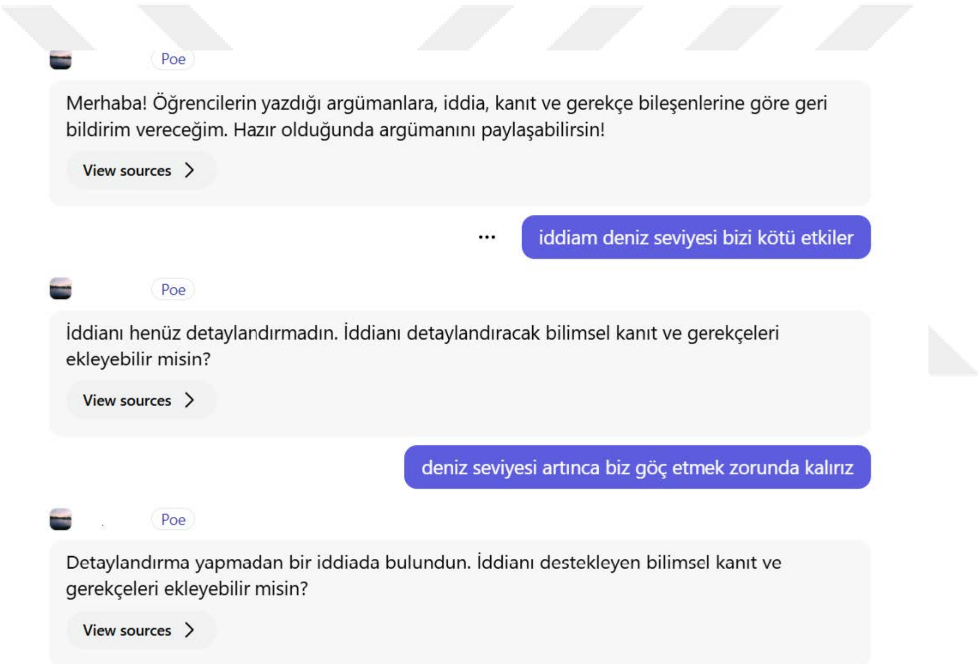


Figure D2. Turkish version of the conversation 2 (after the pilot)

APPENDIX E

ENGLISH AND TURKISH VERSIONS OF THE SCIENTIFIC ARGUMENTATION TEMPLATE

My argument

Please fill in the table below based on your observations in the text and in the simulation.

<table>
<tr><td colspan="2">The Research Question

(Which question was answered in the text and simulation?)</td></tr>
<tr><td colspan="2">The claim

(The opinion I reached as a result of my observations)</td></tr>
<tr><td>The evidence

(On the basis of which observation/findings did I put forward this idea?)</td><td>The Reasoning

(Why did I use this evidence, why is this evidence important?)</td></tr>
</table>

Argümanım

Lütfen aşağıda verilen tabloyu metindeki ve simülasyondaki gözlemlerinize dayanarak doldurunuz.

<table>
<tr><td colspan="2">Araştırma Sorum:
(Metinde ve simülasyonda hangi soruya cevap aradım?)</td></tr>
<tr><td colspan="2">İddiam
(Gözlemlerim sonucunda ulaştığım fikir)</td></tr>
<tr><td>Kanıtlarım
(Hangi gözleme/bulguya dayanarak bu fikri ortaya koyuyorum ?)</td><td>Gerekçem
(Bu kanıtı neden kullandım ,bu kanıt neden önemlidir ?)</td></tr>
</table>

APPENDIX F

ENGLISH AND TURKISH VERSIONS OF STUDENTS' PERCEPTIONS OF LEARNING MODULE

Original version of the questions taken from Chang et al. (2023)

1. What did you like when you were talking with the Inquirybot?
2. What was the most difficult thing when you were talking with the Inquirybot?
3. Do you have any ideas about what needs to be improved or is necessary in designing the inquiry process with the Inquirybot?”

Adapted versions of the questions

1. What experiences did you have when corresponding with a chatbot?

2. Did you have any difficulties when corresponding with the chatbot? If yes, what was the most difficult thing? Explain the reason.

3. When you corresponded with a chatbot, did the feedback and answers make it easier for you to write your arguments?

4. What do you think needs to be improved when corresponding with a chatbot?

5. Were the answers given by the chatbot understandable?

6. Would you want to use a chatbot if we did such an activity again?

7. What have you learnt about the Greenhouse Effect? Was the chatbot effective in developing your understanding?

Araştırmada kullanılan uyarlanmış sorular

1. Chatbot ile yazışırken ne gibi deneyimleriniz oldu?

2. Chatbot ile yazışırken zorlandınız mı? Eğer cevabınız evetse en çok zorlandığınız şey neydi? Nedeniyle birlikte açıklayınız.

3. Chatbot ile yazıştığınızda gelen geri bildirimler ve cevaplar argümanları yazmanızda kolaylık sağladı mı?

4. Chatbot ile yazışırken nelerin geliştirilmesi gerektiğini düşünüyorsunuz?

5. Chatbot'nin verdiği cevaplar anlaşılır mıydı?

6. Böyle bir etkinliği tekrar yapsak Chatbot kullanmak ister miydiniz?

7. Sera Etkisi üzerine neler öğrendiniz, bu öğrendiklerinizde Chatbot etkili oldu mu?

APPENDIX G

ETHICS COMMITTEE APPROVAL

T.C.
BOĞAZİÇİ ÜNİVERSİTESİ REKTÖRLÜĞÜ
Sosyal ve Beşeri Bilimler İnsan Araştırmaları Etik Kurulu (Sbinarek)

Sayı : E-84391427-050.04-195033 26.08.2024

Konu : 2024-55T Kayıt Numaralı Başvurunuz Hakkında

Sayın Dr. Öğr. Üyesi Mutlu ŞEN AKBULUT

Tez danışmanlığını yürüttüğünüz öğrenciniz Erenay Atay'ın "Generative AI as a Catalyst for Enhancing Scientific Argumentation Skills in Sixth-Grade Students" başlıklı projesi ile Boğaziçi Üniversitesi Sosyal ve Beşeri Bilimler İnsan Araştırmaları Etik Kurulu (SBİNAREK)'e yaptığınız 2024-55T kayıt numaralı başvuru 31.07.2024 tarih ve 2024/06 sayılı kurul toplantısında incelenmiş ve projeye etik onay verilmesi uygun bulunmuştur.

Bu karar tüm üyelerin toplantıya on-line olarak katılımıyla ve oybirliği ile alınmıştır. Onay mektubu tüm üyeler adına Komisyon Başkanı tarafından e-imzalanmıştır. Bilgilerinizi rica ederiz.

Saygılarımızla,

Dr. Öğr. Üyesi Işıl ERDUYAN
Kurul Başkanı

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu : 0EE2-P26L-04Z0 Belge Doğrulama Adresi : https://www.turkiye.gov.tr/bogazici-universitesi-ebys

APPENDIX H

APPROVAL FROM MoNE

Başvuru No: MEB.TT.2024.001543

Uygulama Yapılacak MEB Teşkilatının Kurum Kodu: 727502

T.C. Kimlik No:

Adı Soyadı: ERENAY ATAY

Araştırmanın Adı: Üretken Yapay Zekanın 6.Sınıf Öğrencilerinin Sera Etkisi Konusundaki Bilimsel Argümantasyon Kalitesini Artırmadaki Rolünün İncelenmesi

Araştırmanın Niteliği: Yüksek Lisans Tezi

Araştırmanın Örneklem / Çalışma Grubu: Öğrenci

Uygulama Yapılacak MEB Teşkilatı:

Uygulama Yapılacak Birim: Ortaokul

Uygulama Yapılacak İl: MERSİN

Veri Toplama Aracının Başlığı: Öğrenci Çalışma Kağıdı, Argüman Değerlendirme Rubriği, Bilgi Düzeyi Ölçme Soruları, Öğrencinin Yazılı Geri Bildirimi

Araştırma Uygulama İzninin Kabul Tarihi: 30.09.2024

Araştırmanın Uygulama İzninin Bitiş Tarihi: 30.09.2025

Yukarıda kimliği yazılı araştırmacı "Araştırma Uygulama İzinleri Genelgesine (2024/41)" göre belirtilen kapsamda araştırmasını yapmayı taahhüt etmiştir. Araştırmacının bilgi ve belgelerinin uygunluğu kontrol edilmiş olup araştırma uygulama izni MERSİN İl Milli Eğitim Müdürlüğü tarafından onaylanmıştır.

NOT: Okul/kurum yöneticileri tarafından "Araştırma Uygulama İzni" belgesinin ve veri toplama araçlarının (araçlardaki maddelerinin) modülde yer alan belge ve araçlarla aynı olduğu kontrol edilmelidir. Aynı olmadığı durumda araştırma uygulama izni verilmeyecektir.

* Başvuru detayını görüntülemek ve belgeyi doğrulamak için 'https://arastirmaizinleri.meb.gov.tr/belge-dogrula' bağlantısını kullanınız.

- Araştırma Uygulama İzinleri Başvuru ve Değerlendirme Modülü -

APPENDIX I

PARENTAL CONSENT

T.C.

BOĞAZİÇİ ÜNİVERSİTESİ

SOSYAL VE BEŞERI BILIMLER İNSAN ARAŞTIRMALARI ETIK KURULU

KATILIMCI BİLGİ ve ONAM FORMU

Araştırmayı destekleyen kurum: Boğaziçi Üniversitesi

Araştırmanın adı: Üretken Yapay Zekanın Altıncı Sınıf Öğrencilerinin Sera Etkisi Konusundaki Bilimsel Argümantasyon Kalitesini Artırmadaki Rolünün İncelenmesi

Proje Yürütücüsü: Dr. Öğretim Üyesi Mutlu Şen Akbulut

Araştırmacının adı: Erenay Atay

E-mail adresleri:

Telefonu:

Sayın Veli,

Bilimsel argümantasyonun gelişiminde üretken yapay zeka kullanımının etkisine yönelik bir araştırma gerçekleştiriyoruz ve bu çalışmaya öğrencinizin katılımı için sizi davet ediyoruz.

Proje Konusu

Bu araştırma Boğaziçi Üniversitesi Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü öğretim üyesi Dr. Mutlu Şen Akbulut danışmanlığında yüksek lisans öğrencisi Erenay Atay tarafından "Üretken Yapay Zekanın Altıncı Sınıf Öğrencilerinin Sera Etkisi Konusundaki Bilimsel Argümantasyon Kalitesini Artırmadaki Rolünün İncelenmesi" adı ile yürütülmektedir.

Bu çalışma, üretken yapay zeka kullanarak Argümantasyon Tabanlı Sorgulayıcı Araştırma öğretim modeline dayalı olarak tasarlanan ders planı uygulandıktan sonra altıncı sınıf öğrencilerinin sera etkisi konusuna ilişkin yazılı bilimsel argümanlarının niteliklerindeki değişiklikleri ve yapay zekayla etkileşimlerini ve deneyimlerini ortaya çıkarmayı amaçlamaktadır. Çalışma Boğaziçi Üniversitesi'nin etik kurul onayı ile yapılacaktır. Sizi bu araştırmaya velisi olduğunuz öğrencinizin katılımı ile bize yardımcı olmaya davet ediyoruz. Kararınızdan önce araştırma hakkında sizi bilgilendirmek istiyoruz. Bu bilgileri okuduktan sonra araştırmaya katılmak isterseniz lütfen bu formu imzalayıp kapalı bir zarf içinde bize ulaştırınız.

Onam

Bu araştırmada katılım tamamen gönüllülük esasına dayalıdır, katılımın karşılığında herhangi bir ödül verilmeyecektir. Sizlerden gelen onam formları başka bir öğretmen tarafından numaralarla eşleştirilecek, dolayısıyla araştırmacı sizlerin çalışmaya katılıp katılmayacağınız konusunda bilgi sahibi olmayacaktır. Kabul ettiğiniz takdirde öğrencinizin tüm güvenlik önlemlerini alarak tablet ve bilgisayar gibi araçları kullanarak üretici yapay zeka ile yazışmayı öğrendikleri her biri 3 ders saatinden oluşan 3 oturumluk bir deneyime katılmaları, ve veri toplama formlarını doldurmaları beklenecektir. Bu formlar, öğrenme etkinlikleri esnasında ürettikleri yazılı argümanlar ile yapay zeka hakkındaki görüşlerini almayı amaçlayan açık uçlu sorulara verdikleri cevaplardan oluşacaktır. Bu formları doldurmaları en çok 20 dakika sürecektir.

Sera Etkisi temalı her biri 3 ders saatinden oluşan 3 oturumda süreçte öğrencilerin bireysel olarak dolduracakları çalışma kağıtları ile nitel veriler toplanarak öğrencilerin küresel ısınmanın farklı boyutlarıyla ilgili yazdıkları bilimsel argümanlardaki gelişim takip edilecektir. Her oturumda 25 dakikalık bir sürede üretici yapay zeka öğrenciler tarafından kullanılacaktır ve öğrenciler oluşturdukları argümanları üretici yapay zekadan faydalanarak daha iyi hale getireceklerdir. Ders süreci tamamlandığında deneyimleriyle ilgili açık uçlu sorulara verilen cevaplar bireysel olarak toplanacaktır. Bu araştırma bilimsel bir amaçla yapılmaktadır ve katılımcı bilgileri gizli tutulmaktadır. Uygulanan veri toplama araçlarında çocukların ismi yerine bir numara kullanılacaktır. Kişisel bilgi istenmeyecektir.

Bu araştırmaya katılmak tamamen isteğe bağlıdır. Öğrenciniz çalışmaya katıldığınız takdirde çalışmanın herhangi bir aşamasında herhangi bir sebep göstermeden onayınızı çekme hakkına da sahipsiniz. Aynı şekilde öğrenci çalışmaya devam etmek istemediği takdirde herhangi bir sebep göstermeden çalışmadan ayrılabilir. Katılımcının ayrılmak istediği takdirde tüm veriler silinecek ve herhangi bir araştırmada kullanılmayacaktır. Bu araştırmada öğrencileri birbirleriyle ya da başka bir öğrenci grubuyla karşılaştırılmayacaktır.

Araştırmanın sizin ya da velisi olduğunuz öğrenciniz için bir risk oluşturması beklenmiyor. Çalışma sırasında katılımcılar herhangi bir zamanda herhangi bir rahatsızlık hissederlerse çalışmayı durdurabilirler.

Araştırma çerçevesinde toplanan verilerden ulaşılan sonuçlar yüksek lisans tez çalışması için kullanılacaktır. Katılımcı bilgilerinin gizliliği esastır ve sizin ve velisi olduğunuz öğrencinin bilgileri tamamen gizli tutulacaktır. Araştırma projesi hakkında ek bilgi almak istediğiniz takdirde lütfen Boğaziçi Üniversitesi Eğitim Teknolojisi programı yüksek lisans öğrencisi ve aynı zamanda araştırmayı uygulayacak olan okulunuz fen bilimleri öğretmeni Erenay Atay ve Boğaziçi Üniversitesi öğretim üyesi ve tez danışmanı Mutlu Şen Akbulut ile iletişime geçiniz Araştırma katılımcısının ve velisinin hakları konusunda daha fazla bilgi almak için

Boğaziçi Üniversitesi Sosyal ve Beşeri Bilimler Yüksek Lisans ve Doktora Tezleri Etik İnceleme Komisyonu'na (SOBETİK) mail adresinden ulaşarak danışabilirsiniz.

Eğer bu araştırma projesine öğrencinizin katılmasını kabul ediyorsanız, lütfen bu formu imzalayıp kapalı bir zarf içerisinde bize geri yollayın. Ben, (sizin adınız) ...., yukarıdaki metni okudum ve velisi bulunduğum öğrencinin katılması istenen çalışmanın kapsamını ve amacını, gönüllü olarak öğrencinin üzerine düşen sorumluluklarını tamamen anladım. Çalışma hakkında soru sorma imkanı buldum. Bu çalışmayı öğrencinin istediği zaman ve herhangi bir neden belirtmek zorunda kalmadan bırakabileceğini ve bıraktığı takdirde herhangi bir olumsuzluk ile karşılaşmayacağımı anladım.

Bu koşullarda söz konusu araştırmaya öğrencinin kendi isteğimle, hiçbir baskı ve zorlama olmaksızın katılmasını kabul ediyorum.

Formun bir örneğini aldım / almak istemiyorum (bu durumda araştırmacı bu kopyayı saklar).

Katılımcının Vasisinin Adı-Soyadı:...

İmzası:...

Tarih (gün/ay/yıl): ...../......./........

Katılımcının Adı-Soyadı:...

İmzası:...

Tarih (gün/ay/yıl): ...../......./........

Araştırmacının Adı-Soyadı: Erenay Atay

İmzası:...

......

Tarih (gün/ay/yıl): ...../......./........

APPENDIX J

STUDENT CONSENT FORM

T.C.

BOĞAZİÇİ ÜNİVERSİTESİ

SOSYAL VE BEŞERI BILIMLER İNSAN ARAŞTIRMALARI ETIK KURULU

KATILIMCI BİLGİ ve ONAM FORMU

Araştırmayı destekleyen kurum: Boğaziçi Üniversitesi

Araştırmanın adı: Üretken Yapay Zekanın Altıncı Sınıf Öğrencilerinin Sera Etkisi Konusundaki Bilimsel Argümantasyon Kalitesini Artırmadaki Rolünün İncelenmesi

Proje Yürütücüsü: Dr. Öğretim Üyesi Mutlu Şen Akbulut

Araştırmacının adı: Erenay Atay

E-mail adresi:

Telefonu:

Sevgili öğrenci,

Bilimsel argümantasyonun gelişiminde üretken yapay zeka kullanımının etkisine yönelik bir araştırma gerçekleştiriyoruz ve bu çalışmaya seni davet ediyoruz.

Proje Konusu

Bu araştırma Boğaziçi Üniversitesi Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü öğretim üyesi Dr. Mutlu Şen Akbulut danışmanlığında yüksek lisans öğrencisi Erenay Atay tarafından "Üretken Yapay Zekanın Altıncı Sınıf Öğrencilerinin Sera Etkisi Konusundaki Bilimsel Argümantasyon Kalitesini Artırmadaki Rolünün İncelenmesi" adı ile yürütülmektedir.

Bu çalışmada, üretken yapay zeka kullanarak sizlerin sera etkisi hakkındaki bilimsel argümanlarınızın kalitelerindeki gelişimi ve bu süreçte yapay zeka ile olan etkileşimlerinizi ve deneyimlerinizi ortaya çıkarmayı amaçlıyoruz. Çalışma Boğaziçi Üniversitesi'nin etik kurul onayı ile yapılacaktır.

Seni yukarıda bahsedilen araştırmaya katılmaya davet ediyoruz. Kararından önce araştırma hakkında seni bilgilendirmek istiyoruz. Bu bilgileri okuduktan sonra araştırmaya katılmak istersen lütfen bu formu imzalayıp kapalı bir zarf içinde bize ulaştır.

Onam

Bu araştırmada katılım tamamen gönüllülük esasına dayalıdır, katılımın karşılığında herhangi bir ödül verilmeyecektir. Kabul ettiğin takdirde, tablet/ bilgisayar gibi araçları kullanarak üretici yapay zeka ile yazışmayı öğrenmen, her biri 3 ders saatinden oluşan 3 oturuma katılman ve veri toplama formlarını doldurman beklenecektir. Bu formlar, öğrenme etkinlikleri esnasında ürettiğiniz yazılı argümanlar ile yapay zeka hakkındaki görüşlerinizi almayı amaçlayan açık uçlu sorulara verdiğiniz cevaplardan oluşacaktır. Bu formları doldurmanız en çok 20 dakika sürecektir.

Sera Etkisi temalı 3 oturumluk süreçte sizlerin bireysel olarak dolduracakları çalışma kağıtları ile nitel veriler toplanarak sizlerin yazılı bilimsel argümanlarınızdaki gelişim takip edilecektir. Her oturumda 25 dakikalık bir sürede üretici yapay zeka kullanacaksınız ve oluşturduğunuz argümanları üretici yapay zekadan faydalanarak daha iyi hale getirebileceksiniz. Ders süreci tamamlandığında sohbet robotu hakkındaki geri bildirimlerinizi yazılı olarak paylaşmanız istenecektir. Bu araştırma bilimsel bir amaçla yapılmaktadır ve katılımcı bilgileri gizli tutulmaktadır.

Uygulanan veri toplama araçlarında senin ismin yerine bir numara kullanılacaktır. Kişisel bilgi istenmeyecektir.

Bu araştırmaya katılmak tamamen isteğe bağlıdır. Sizlerden gelen onam formları başka bir öğretmen tarafından numaralarla eşleştirilecek, dolayısıyla araştırmacı sizlerin çalışmaya katılıp katılmayacağınız konusunda bilgi sahibi olmayacaktır. Çalışmaya katıldığın takdirde çalışmanın herhangi bir aşamasında herhangi bir sebep göstermeden çekilme hakkına da sahipsiniz. Aynı şekilde çalışmaya devam etmek istemediğin takdirde herhangi bir sebep göstermeden çalışmadan ayrılabilirsin. Ayrılmak istediğin takdirde tüm veriler silinecek ve herhangi bir araştırmada kullanılmayacaktır. Bu araştırmada öğrencileri birbirleriyle ya da başka bir öğrenci grubuyla karşılaştırılmayacaktır.

Araştırma çerçevesinde toplanan verilerden ulaşılan sonuçlar yüksek lisans tez çalışması için kullanılacaktır. Katılımcı bilgilerinin gizliliği esastır ve senin bilgilerin tamamen gizli tutulacaktır. Araştırma projesi hakkında ek bilgi almak istediğin takdirde lütfen Boğaziçi Üniversitesi Eğitim Teknolojisi programı yüksek lisans öğrencisi ve aynı zamanda araştırmayı uygulayacak olan okulunuz fen bilimleri öğretmeni Erenay Atay ve Boğaziçi Üniversitesi öğretim üyesi ve tez danışmanı Mutlu Şen Akbulut ile iletişime geçiniz. Hakların konusunda daha fazla bilgi almak için Boğaziçi Üniversitesi Sosyal ve Beşeri Bilimler Yüksek Lisans ve Doktora Tezleri Etik İnceleme Komisyonu'na (SOBETİK) mail adresinden ulaşarak danışabilirsin.

Eğer bu araştırma projesine katılmasını kabul ediyorsan, lütfen bu formu imzalayıp kapalı bir zarf içerisinde bize geri yollayın.

Ben, (yan tarafa adınızı yazınız) ...., yukarıdaki metni okudum ve katılmam istenen çalışmanın kapsamını ve amacını, gönüllü olarak üzerime düşen sorumlulukları tamamen anladım. Çalışma hakkında soru sorma imkanı buldum. Bu çalışmayı istediğim zaman ve herhangi bir neden belirtmek zorunda kalmadan bırakabileceğimi ve bıraktığım takdirde herhangi bir olumsuzluk ile karşılaşmayacağımı anladım.

Bu koşullarda söz konusu araştırmaya kendi isteğimle, hiçbir baskı ve zorlama olmaksızın katılmayı kabul ediyorum.

Formun bir örneğini aldım / almak istemiyorum (bu durumda araştırmacı bu kopyayı saklar).

Katılımcının Adı-

Soyadı:...

İmzası:...

Tarih (gün/ay/yıl): ...../......./........

Araştırmacının Adı-Soyadı: Erenay Atay

İmzası:...

......

Tarih (gün/ay/yıl): ...../......./........

APPENDIX K

LEARNING MODULE PLAN ON THE RESULTS OF THE GREENHOUSE EFFECT

Grade: Sixth grade

Classroom info: 5 students

Duration: 3x4 lesson hours/ 120 minutes x3

2. Expected prior knowledge about the topic:

Students are expected to know temperature and heat concepts, and how evaporation occurs, and how light is reflected on a surface. Besides, they are also expected to have a basic understanding of the impact of human activities on the environment regarding soil, air, and water pollution. Since students got training on how the greenhouse effect works, they had heard of CO_2 gas before, and they are expected to know how the greenhouse effect works and the possible results.

3. Objectives for the learning module:

Students,

Interpret and explain the evidence obtained from activities and simulations to describe how Earth's natural greenhouse effect affects

Communicate the arguments by explaining the results of the greenhouse effect by interacting with a chatbot.

Engage in an argument stemming from evidence by using generative AI feedback.

Plan and carry out investigations via simulations.

4. Science practices:

Asking Questions and Defining Problems: Students formulate research questions for their arguments

Developing and using models: Students use Phet simulation to construct an explanation for their observations.

Planning and carrying out investigations: Students plan their investigation and then obtain data to seek answers for the research questions.

Analyzing and interpreting data: Students analyze their data in their worksheets and form an explanation by considering their observations during the learning module

Constructing explanations and designing solutions: Students construct arguments based on their evidence and reasoning.

Engaging in argument from evidence: Students form scientific arguments after their observations, and they revise their written arguments by benefiting from evidence sustained by generative AI.

Possible misconceptions to consider:

The ratio of Carbon dioxide is so small in the atmosphere. It cannot make a difference (Jarrett & Takacs,2020).

5. Materials needed:

- Digital Devices (e.g., tablets or computers with an internet connection)
- Worksheets
- Smartboard
- Generative AI

- Board markers
- Papers

6. Body of the lesson

Based on the experiences in the pilot study, the researcher decided to omit some steps in the ADI instructional model and applied five steps by replacing the group work with chatbot conversations and neglecting the feedback of classmates.

1. Topic: Results of the greenhouse effect on temperature rise

1. Identification of the task (10 minutes)

Initially, students are asked what they remember about the greenhouse effect by using the following questions: “What does the greenhouse effect lead to based on your observations in previous lessons?”, “What kind of problems exist when temperature values run up? How can it affect us?”, “On Earth, all areas will be affected equally; why?” After students discuss these questions, the researcher introduces the worksheet and the simulation on temperature change over the years.

2. Generation of data (30 minutes)

Students interact with the simulation individually and record their observations on the given worksheet. After they complete the worksheet, students share their findings with their classmates. Then, students are given a text about the effects of the increase in temperature values on Earth. After they read the text, they answer the given questions individually on paper. Then, students share their answers orally with their classmates.

3. Production of a tentative argument (25 minutes)

Students form their scientific arguments individually by using the template based on data they obtained from the simulation and text. To guide students in forming structured arguments, students are given an evaluation rubric, so they remember what kind of characteristics their arguments should have. Students who finish earlier than others are permitted to continue with other steps of the learning module.

4. Argumentation session (20 minutes)

After students form their arguments, they open the chatbot formed by the researcher called IndSciArg in the given devices. There, students write their existing arguments and get feedback individually. The researcher gives suggested prompts to use during their conversation. While interacting, students are asked to write at least two questions that were not written in the suggested prompts. When students complete the conversation, they determine the chatbot responses they plan to use in their revised arguments.

5. Creation of a written investigation report (35 minutes)

Students write their final version of the scientific arguments, benefiting from chatbot responses and using the scientific argument template. After that, they write an explanation for each chatbot response they utilized in their arguments by stating reasons for selection, including response quality or clarity. Finally, students answer the perception questions about the learning module experience.

2. Topic: Results of the greenhouse effect on melting ice

1. Identification of the task (10 minutes)

Students are reminded about the previous lesson, and it is expected that temperature increase and melting ice will correlate. They are asked guiding questions to attract their attention to the issue. Then, they open the simulation, which shows melting ice in different years in different regions. The researcher introduces the simulation interface.

2. Generation of data (30 minutes)

Students interact with the melting ice simulation individually and discuss their observations. Then, they are given a text about the effects of melting ice on Earth. After they read the text, they answer the given questions individually on paper. Then, students share their answers orally with their classmates.

3. Production of a tentative argument (25 minutes)

Students write their scientific arguments individually based on data they obtained from a simulation and text, benefiting from a scientific argument template. To guide students in forming structured arguments, students are given an evaluation rubric, so they remember what kind of characteristics their arguments should have.

4. Argumentation session (20 minutes)

After students construct their arguments, they open the chatbot called “IndSciArg” in the given devices. There, students chat for their existing arguments and get feedback or guidance individually. During their conversation students are given suggested prompts as well. Besides, they are requested to form at least two questions not placed in the suggested prompts. When students complete the conversation, they determine the chatbot responses that led to revision in their arguments.

5. Creation of a written investigation report (35 minutes)

Students generate their final version of the scientific arguments utilizing the scientific argument template. After that, they write an explanation for each chatbot response they utilized their arguments by stating reasons for selection, including response quality. Finally, students answer the perception questions about the learning module experience.

3. Topic: Results of the greenhouse effect on sea level rise

1. Identification of the task (10 minutes)

Students are reminded about the previous topic. Students are expected to correlate melting ice and sea level rise. The researcher directs guiding questions to attract students’ attention to the issue. Then, students open the simulation, which shows the amount of incoming water to city centers when the sea level rise reaches specific values in different scenarios. The researcher explains the simulation tools.

2. Generation of data (30 minutes)

Students interact with the sea level rise simulation individually and discuss their observations with their classmates. Then, students are given a text about the effects of sea level rise on Earth. After they read the text, they answer the given questions related to the text individually. Then, students share their answers orally with their classmates.

3. Production of a tentative argument (25 minutes)

Students construct their scientific arguments individually based on data they obtained from a simulation and text using the scientific argument template. While students form scientific arguments, the researcher facilitates an evaluation rubric, thus they keep in mind what kind of features their arguments should have.

4. Argumentation session (20 minutes)

After students form their scientific arguments, they open the chatbot called IndSciArg on the given devices. The researcher provides the suggested prompts that students can use during their conversation. There, students interacted with the chatbot to ask for feedback and guidance for their scientific arguments. During the conversation, the researcher asks students to structure at least two prompts generated by themselves rather than given prompts. When the conversation is over, students determine the chatbot responses they plan to use in their arguments.

5. Creation of a written investigation report (35 minutes)

Students revise their existing arguments utilizing the scientific argument template. Then, they explain the reasons to use chatbot responses in their arguments. Finally, students answer the perception questions about the learning module experience.

7. Simulations and hands-on materials

1. 1. Topic-time machine simulation
2. 1. Topic-time: machine simulation worksheet
3. 1. Topic-text and questions about the effect of temperature rise
4. Scientific argumentation template (see Appendix E)
5. Suggested prompts (see Appendix L)
6. 2 . Topic-melting ice simulation
7. 2 . Topic-text about the effect of melting ice
8. 3 . Topic-sea level rise risk map simulation
9. 3 . Topic-text about the effect of sea level rise
10. Perception questions about the learning module (See Appendix F)

1. Time machine simulation

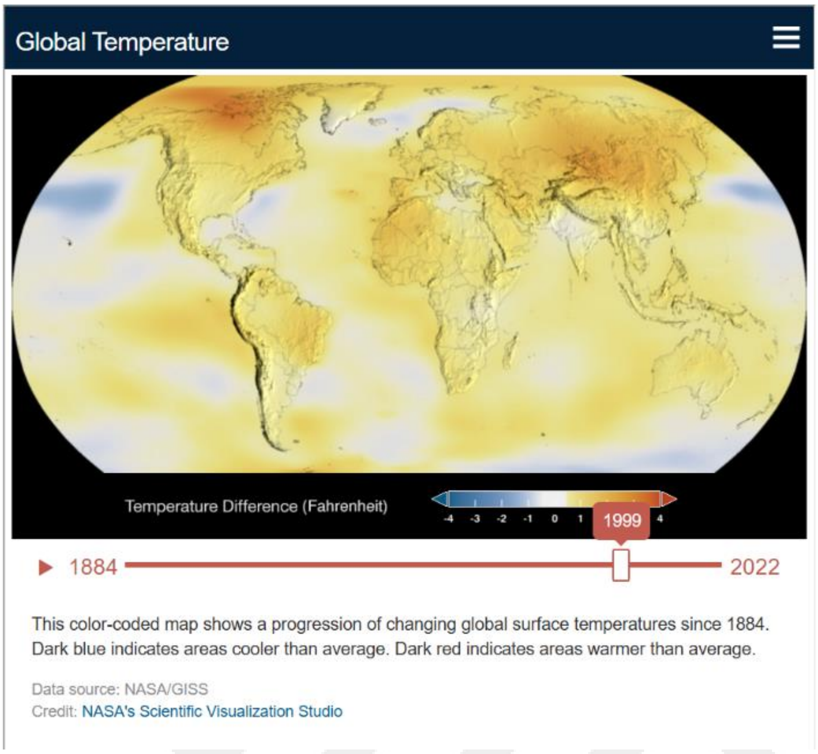


Figure K1 Screenshot from the time machine simulation

Source : NASA. (n.d.). *Interactives – Climate change: Vital Signs of the planet*. Retrieved December 10, 2024, from https://climate.nasa.gov/interactives/climate-time-machine/?intent=021

2. English version of time machine simulation worksheet

<u>Please fill in the table based on your observations.</u>

Name of Region	1884	1939	1976	2022
Europe				
Asia				
Africa				
South America				
North Pole				

The result I reached:

1. According to the obtained data, how did temperature values change all around the World?

2. What do you think the change in temperatures might have caused?

2. Turkish version of time machine simulation worksheet

<u>Bulgularınıza göre tabloyu doldurunuz.</u>

Bölge Adı	1884	1939	1976	2022
Avrupa				
Asya				
Afrika				
Güney Amerika				
Kuzey Kutbu				

Ulaştığım Sonuç:

1. Elde ettiğiniz verilere göre Dünya genelinde sıcaklık değerleri nasıl değişmiştir?

2. Sıcaklıklardaki değişim sizce nelere yol açmış olabilir ?

3. English version of the text and questions about the effect of temperature rise

The Effects of Global Warming on Temperature

Global temperature rise, largely caused by human activities, increases the amount of energy in the atmosphere, leading to the intensification of weather events. As a result, regions once known for their temperate climates are now experiencing record-breaking temperatures. This situation not only poses health risks but also puts energy and water resources at risk of depletion. Moreover, rising temperatures exacerbate drought conditions in many regions, leading to crop shortages, forest fires, and desertification.

According to NASA data, global temperatures have exceeded pre-industrial levels by 1°C. In Turkey, this increase has reached 1.5°C. This situation is amplifying economic and social challenges in our country. Scientists state that temperatures in Turkey could rise by up to 5°C, and precipitation in the southern and western regions of the country may decrease by as much as 30%.

In 2017, forest areas equivalent to 1.3 times the surface area of Yalova province were damaged by wildfires. That year, 2,411 forest fires broke out, and a total of 1,120 km² of forested land was burned. In cities such as Izmir, Antalya, Bursa, Tunceli, and many others, thousands of trees were reduced to ashes.

Scientists predict that if global temperatures rise by an additional 2°C, summers will become much hotter, and in regions like the Middle East and North Africa, thermometers could reach up to 50°C. Such temperatures could lead to an increase in forest fires and heatwave-related deaths. The impact of these extreme weather events extends beyond immediate physical damage, posing significant challenges to global food security, water resources, and economic stability.

Sources:

IPCC. (n.d.). IPCC 1.5°C raporu. Birbuçuk Derece. Retrieved October 8, 2024, from https://www.birbucukderece.com/bilimsel-kaynaklar/ipcc-1-5-c-raporu

Mentor, I. (2024, August 22). IELTS reading practice: The effects of global warming on weather extremes. IELTS.NET. https://www.ielts.net/the-effects-of-global-warming-on-weather-extremes/

Questions

1. What evidence is provided for the increase in temperatures?

2. Based on the information in the text, what might be the effects of the temperature increase?

3.Identify and underline the claim, reasoning, and evidence components in this text.

3. Turkish version of the text and questions about the effect of temperature rise

Küresel Isınmanın Sıcaklık Üzerine Etkileri

Büyük ölçüde insan faaliyetlerinin neden olduğu küresel sıcaklık artışı atmosferdeki enerji miktarını artırarak hava olaylarının şiddetlenmesine yol açmaktadır. Bu nedenle bir zamanlar ılıman iklimleriyle bilinen bölgeler artık rekor kıran sıcaklıklar yaşamaktadır. Bu durum sadece sağlık riskleri oluşturmakla kalmıyor, aynı zamanda enerji ve su kaynaklarını da tükenme riskiyle karşı karşıya bırakmaktadır. Dahası, artan sıcaklıklar birçok bölgede kuraklık koşullarını daha da kötüleştirerek ürün kıtlığına, orman yangınlarına ve çölleşmeye yol açmaktadır.

NASA verilerine göre Dünya genelinde sıcaklıklar, sanayi öncesi döneme göre 1°C'yi aşmıştır. Türkiye'de ise bu artış, 1,5°C olmuştur. Bu durum, ülkemizde ekonomik ve sosyal sorunları artırmaktadır. Bilim insanları, Türkiye'de sıcaklıkların 5°C'ye kadar yükselebileceğini ve ülkenin güney ve batı bölgelerinde yağışların %30'a kadar azalabileceğini söylemektedir.

2017 yılında, Yalova ilimizin yüzölçümünün 1,3 katı büyüklüğünde orman alanı yangınlarda zarar görmüştür. O yıl 2 bin 411 orman yangını çıkmış ve toplam 1120 km² ormanlık alan yanmıştır. İzmir, Antalya, Bursa, Tunceli ve birçok başka şehirde binlerce ağaç küle dönmüştür.

Bilim insanları, dünya sıcaklıklarının 2°C daha artması durumunda, yaz aylarının çok daha sıcak geçeceğini ve Ortadoğu, Kuzey Afrika gibi bölgelerde termometrelerin 50°C'ye kadar çıkabileceğini tahmin etmektedir. Bu sıcaklıklar, orman yangınlarının ve sıcak hava dalgalarına bağlı ölümlerin artmasına neden olabilir. Bu aşırı hava olaylarının etkisi, anlık fiziksel hasarın ötesine uzanmaktadır. Küresel gıda güvenliği, su kaynakları ve ekonomik istikrar için önemli zorluklar teşkil etmektedirler.

Kaynak:

IPCC. (n.d.). IPCC 1.5°C raporu. Birbuçuk Derece. Retrieved October 8, 2024, from https://www.birbucukderece.com/bilimsel-kaynaklar/ipcc-1-5-c-raporu

Mentor, I. (2024, August 22). IELTS reading practice: The effects of global warming on weather extremes. IELTS.NET. https://www.ielts.net/the-effects-of-global-warming-on-weather-extremes/

Sorular

1. Sıcaklık artışına dair kanıtlar nelerdir?

2. Metindeki bilgiler doğrultusunda sıcaklık artışının ne gibi etkileri olabilir?

3. Bu metinde geçen iddia, gerekçe ve kanıt bileşenlerini bularak altını çiziniz.

7. Melting ice simulation

Figure K2. Screenshot from melting ice simulation

NASA. (n.d.). Global ice viewer [Interactive tool]. Retrieved October 8, 2024, from https://climate.nasa.gov/interactives/global-ice-viewer/?intent=021#/1

8. English version of the text about the effect of melting ice

Melting Glaciers

The ice sheets in Greenland and Antarctica are massive land-based ice masses that store approximately 70% of the Earth's freshwater. However, global temperature increases have gradually melted sea ice over time, reducing the brightness of surfaces that reflect sunlight back into the atmosphere. As a result, more solar energy is absorbed by the surface, leading to rising ocean temperatures. This initiates a cycle of warming and melting. The increase in water temperatures delays ice growth during autumn and winter, making ice formation more difficult, and glaciers mix with ocean waters.

These changes in sea ice levels can disrupt normal ocean cycles, leading to shifts in global climate. For example, even a small increase in temperature can cause further warming over time, making polar regions the most vulnerable areas to climate change on Earth.

Consequently, as ice sheets shrink, the water they add to the oceans raises global sea levels. According to NASA data, Antarctica loses an average of 150 billion tons of ice mass annually, while Greenland loses approximately 270 billion tons per year, contributing to sea level rise. The melting of ice sheets in Greenland and Antarctica accounted for about one-third of the global sea level rise observed between 2006 and 2015. Thus, ice sheets are significant indicators not only due to their sensitivity to climate changes but also because of their role in global sea level rise.

The melting glaciers contribute to rising sea levels, increase coastal erosion, and lead to more frequent and intense coastal storms such as hurricanes and typhoons due to warming air and ocean temperatures. Today, the glacier melt observed in Antarctica and Greenland is altering the circulation of the Atlantic Ocean, which is thought to be linked to the collapse of fisheries in the Gulf of Maine and more destructive storms and hurricanes across the planet.

Even if we significantly reduce greenhouse gas emissions in the coming years, more than one-third of the world's remaining glaciers will melt before 2100. If all glaciers and ice sheets were to melt completely, global sea levels would rise by more than 60 meters.

Resources:

U.S. Environmental Protection Agency. (2025, January 15). *Climate change indicators: Ice sheets.*

NASA. (n.d.). *Ice melt.* NASA Sea Level Change. Retrieved October 8, 2024, from https://sealevel.nasa.gov/understanding-sea-level/global-sea-level/ice-melt/

NASA. (n.d.). *Ice sheets.* NASA Climate. Retrieved October 8, 2024, from https://climate.nasa.gov/vital-signs/ice-sheets/

NOAA. (n.d.). *Sea ice and climate.* NOAA Ocean Service. Retrieved October 8, 2024, from https://oceanservice.noaa.gov/facts/sea ice-climate.html

Please answer the following questions.

1. Based on the information in the text, what effects could the melting of glaciers have?

2. Find and underline the components of claim, evidence and reasoning in this text.

8. Turkish version of the text about the effect of melting ice

Buzulların erimesi

Grönland ve Antarktika'daki buz tabakaları, dünya yüzeyindeki tatlı suyun yaklaşık %70'ini depolayan devasa karasal buz kütleleridir. Ancak sıcaklıklardaki küresel artış, deniz buzunu zaman içinde yavaş yavaş erittiği için, güneş ışığını atmosfere geri yansıtacak olan yüzeylerin parlaklığı azalır. Bu nedenle yüzeyde daha fazla güneş enerjisi emilir ve okyanus sıcaklıkları yükselir. Bununla birlikte ısınma ve erime döngüsü başlar. Su sıcaklıklarındaki artış, sonbahar ve kış aylarında buzun büyümesini geciktirir ve buz oluşumu zorlaşır, buzullar okyanus sularına karışır. Deniz buzu miktarındaki bu değişiklikler normal okyanus döngüsünü bozarak küresel iklimde değişikliklere yol açabilir. Örneğin sıcaklıktaki küçük bir artış bile zaman içinde daha fazla ısınmaya yol açabilir, bu da kutup bölgelerini Dünya'daki iklim değişikliğine karşı en hassas alanlar haline getirir.

Sonuçta, buz tabakaları küçüldükçe, okyanusa ekledikleri su dünya çapında deniz seviyesini yükseltmektedir. NASA verilerine göre Antarktika yılda ortalama 150 milyar ton buz kütlesi kaybediyor ve Grönland yılda yaklaşık 270 milyar ton kaybederek deniz seviyesinin yükselmesine katkıda bulunuyor. Grönland ve Antarktika'daki buz tabakalarının erimesi, 2006-2015 yılları arasında gözlemlenen küresel deniz seviyesi artışının yaklaşık üçte birini oluşturmuştur. Dolayısıyla, buz tabakaları sadece iklimdeki değişikliklere olan duyarlılıkları nedeniyle değil, aynı zamanda küresel deniz seviyesi artışındaki rolleri nedeniyle de önemli göstergelerdir. Eriyen buzullar deniz seviyesinin yükselmesine katkıda bulunmakta birlikte kıyı erozyonunu artırmakta ve ısınan hava ve okyanus sıcaklıkları kasırga ve tayfun gibi kıyı fırtınalarının daha sık ve yoğun yaşanmasına neden olmaktadır. Bugün Antarktika ve Grönland'da tanık olduğumuz buzul erimesi, Atlantik Okyanusu'nun dolaşımını değiştiriyor ve Maine Körfezi'ndeki balıkçılığın çökmesi ve gezegenin her yerinde daha yıkıcı fırtınalar ve kasırgalarla bağlantılı olduğu düşünülüyor. Önümüzdeki yıllarda sera gazı emisyonlarını önemli ölçüde azaltsak bile, dünyanın kalan buzullarının üçte birinden fazlası 2100 yılından önce eriyecektir. Eğer tüm buzullar ve buz tabakaları erirse, küresel deniz seviyesi 60 metreden daha fazla yükselecektir.

Kaynak:

U.S. Environmental Protection Agency. (2025, January 15). *Climate change indicators: Ice sheets.*

NASA. (n.d.). *Ice melt.* NASA Sea Level Change. Retrieved October 8, 2024, from https://sealevel.nasa.gov/understanding-sea-level/global-sea-level/ice-melt/

NASA. (n.d.). *Ice sheets.* NASA Climate. Retrieved October 8, 2024, from https://climate.nasa.gov/vital-signs/ice-sheets/

NOAA. (n.d.). *Sea ice and climate.* NOAA Ocean Service. Retrieved October 8, 2024, from https://oceanservice.noaa.gov/facts/sea ice-climate.html

Lütfen aşağıdaki soruları cevaplayınız.

1. Metindeki bilgiler doğrultusunda buzulların erimesinin ne gibi etkileri olabilir?

2. Bu metinde geçen iddia, gerekçe ve kanıt bileşenlerini bularak altını çiziniz.

9. Sea level rise risk map simulation

Climate Central. (n.d.). *Sea level rise map* [Interactive map]. Retrieved October 8, 2024, from https://coastal.climatecentral.org/map/9/35.2506/37.1851/?theme=water_level&map_type=water_level_above_mhhw&basemap=roadmap&contiguous=true&elevation_model=best_available&refresh=true&water_level=1.0&water_unit=m

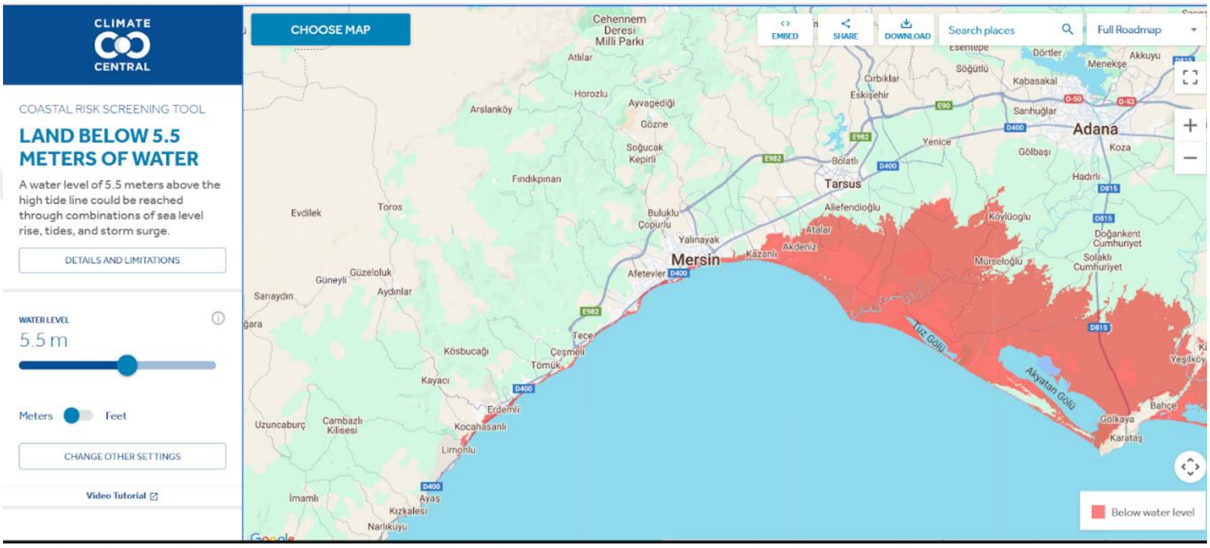


Figure K3. Screenshot from the sea level risk map

10. English version of the text and questions about the effect of sea level rise

The consequences of sea level rise

Sea level rise, as the name suggests, refers to an increase in the total volume of ocean water. This phenomenon is a natural consequence of our planet's warming. When ice is exposed to heat, it melts; as land-based ice melts and flows into the ocean as water, sea levels rise. This increase is caused not only by melting glaciers and polar ice sheets but also by the natural expansion of water as it warms, meaning that sea levels rise when oceans heat up as well.

Since 1850, sea levels have risen by 20 cm globally, driven by global warming caused by greenhouse gas emissions from the burning of coal, oil, and gas. For thousands of years, sea levels remained relatively stable, and human communities settled along the Earth’s coastlines. However, over the past 20 years, sea levels have risen by more than 5 cm. According to data, the rate of this rise has accelerated in recent years.

The impacts of sea level rise on communities are already being observed in many regions. In Miami, streets in the city center regularly experience flooding during high tides. The South Pacific island nation of Kiribati saw a sea level rise of 2.6 mm between 1992 and 2010. Considering that the land is only 2 meters above sea level, this increase poses significant risks. Some villages have already had to relocate to escape rising tides. Additionally, China’s Yellow River Delta is being inundated due to sudden sea level rises.

Rising seas are causing saltwater to intrude into inland freshwater sources. This poses a risk of contaminating drinking water in coastal areas, including major cities like Miami, New Orleans, and Shanghai. Furthermore, as more saltwater enters freshwater sources and soil, agricultural lands are becoming contaminated. This issue is evident in Bangladesh, where vast agricultural lands are affected, forcing people to abandon their lands. In addition to contaminating freshwater sources, floods can lead to waterborne diseases that threaten human health, as well as cause stress and mental health issues. Moreover, beaches, resorts, and coral reefs—key tourist attractions and significant sources of income, particularly for many developing small island states—may suffer damage.

Due to sea level rise, it is predicted that entire populations of many countries will be forced to relocate, potentially leading to refugee crises. It is estimated that approximately 900 million people—about one in every ten people on Earth—live in low-lying coastal areas. People living in densely populated coastal regions of countries such as Bangladesh, China, India, the Netherlands, and Pakistan are at risk and expected to face flooding disasters.

As Pacific Island nations become inundated, hundreds of thousands of people will be displaced. Even now, more than 50,000 people in the Pacific are forced to leave their homes each year. Some of these countries have already developed relocation plans. Low-lying small islands face the most critical threats. Sea level rise and other climate impacts are already forcing people in Pacific Ocean nations such as Fiji, Vanuatu, and the Solomon Islands to relocate. It is foreseen that the expected rise in sea levels will force people to abandon their homes, move to higher ground if available, or ultimately migrate.

Resources:

Siegert, M. J. (2017). Why should we worry about sea level change? *Frontiers for Young Minds*, *5*. https://doi.org/10.3389/frym.2017.00041

NRDC. (n.d.). *Sea level rise 101.* Retrieved October 8, 2024, from https://www.nrdc.org/stories/sea-level-rise-101

NASA. (2014, August 26). *Warming seas and melting ice sheets.* NASA Climate. https://climate.nasa.gov/news/2328/warming-seas-and-melting ice-sheets/

UN News. (2024, August 26). *What is sea level rise and why does it matter to our future?*

Please answer the questions given below.

1. Based on the information in the text, what is the cause of sea level rise?

2. Based on the information in the text, what effects has sea level rise had on people?

3. What do experts predict the future rise in sea levels might lead to?

4. Identify and underline the claim, reasoning, and evidence components in this text.

10. Turkish version of the text and questions about the effect of sea level rise

Deniz seviyesinin yükselmesinin sonuçları

Deniz seviyesinin yükselmesi, adından da anlaşılacağı üzere, okyanus suyunun toplam hacminin artmasıdır. Bu durum, gezegenimizin ısınmasının doğal bir sonucudur. Buz ısıya maruz kaldığında erir; karadaki buzlar eriyip su olarak okyanusa akınca da deniz seviyesi yükselir. Bu artış, eriyen buzullar ve kutuplardaki buz tabakalarının yanı sıra suyun ısındıkça doğal olarak genleşmesinden de kaynaklanır dolayısıyla okyanuslar ısındığında da deniz seviyesi yükselir.

Deniz seviyesindeki yükselme 1850 yılından bu yana dünya genelinde 20 cm daha yüksektir bunun nedeni kömür, petrol ve gazın yakılmasından kaynaklanan sera gazı emisyonlarının neden olduğu küresel ısınmadır. Binlerce yıldır deniz seviyesi nispeten sabit kalmıştır ve insan toplulukları Dünya'nın kıyı şeridine yerleşmiştir. Ancak son 20 yılda deniz seviyesi 5 santimetreden fazla yükselmiştir. Verilere göre son yıllarda bu yükseliş hızını arttırmıştır.

Birçok bölgede deniz seviyesinin yükselmesinin toplumlar üzerindeki etkileri şimdiden görülmeye başlamıştır. Miami'de gelgit sırasında şehir merkezindeki caddelerde düzenli olarak sel baskınları yaşanıyor. Güney Pasifik ada ülkesi Kiribati'de 1992 ve 2010 yılları arasında deniz seviyesinde 2,6 milimetrelik bir yükselme gördü. Toprağın deniz seviyesinden sadece 2 metre yüksekte olduğu düşünüldüğünde bu artışın ciddi riskler taşıdığı söylenebilir. Bazı köyler yükselen gelgitlerden kaçmak için şimdiden yer değiştirmek zorunda kalmıştır. Buna ilaveten, Çin'in Sarı Nehir deltası deniz seviyesinin ani yükselmesi nedeniyle sular altında kalmaktadır.

Yükselen denizler, iç kesimlerdeki tatlı su kaynaklarına tuzlu su girmesine neden olmaktadır. Bu durum Miami, New Orleans ve Şangay gibi büyük şehirler de dahil olmak üzere kıyı bölgelerindeki içme suyunu kirletme riski taşımaktadır. Ayrıca, tatlı su kaynaklarına ve toprağa daha fazla tuz girdiği için tarım arazileri de kirlenmektedir. Bu durum, geniş tarım arazilerinin olduğu ve insanların topraklarından ayrılmak zorunda kaldığı Bangladeş'te görülmektedir. Taşkınların tatlı su kaynaklarını kirletmesinin yanısıra insanların sağlığını tehdit eden su kaynaklı hastalıklara neden olup stres ve ruh sağlığı sorunlarına yol açabilir. Aynı zamanda, özellikle gelişmekte olan birçok küçük ada devletinde önemli bir gelir kaynağı olan plajlar, tatil köyleri ve mercan resifleri gibi diğer turistik yerler zarar görebilir. Deniz seviyesinin yükselmesi nedeniyle, birçok ülkenin nüfusunun tamamının yer değiştirmek zorunda kalacağı ve potansiyel mülteci krizlerinin oluşacağı öngörülmektedir. Yaklaşık 900 milyon insanın, yani dünya üzerindeki her 10 kişiden birinin denize yakın yerlerde yaşadığı tahmin edilmektedir. Bangladeş, Çin, Hindistan, Hollanda ve Pakistan gibi yoğun nüfuslu ülkelerin kıyı bölgelerinde yaşayan insanlar risk altında olacağı ve sel felaketlerine maruz kalmaları beklenmektedir.

Pasifik ada ülkeleri sular altında kaldıkça, yüz binlerce kişi yerinden olacaktır. Şu anda bile Pasifik'te her yıl 50.000'den fazla insan evlerinden ayrılmak zorunda kalmaktadır. Bu ülkelerden bazıları şimdiden yeniden yerleştirme planları yapmış durumdadır. Alçakta kalan küçük adalar en kritik tehditlerle karşı karşıyadır. Deniz seviyesinin yükselmesi ve diğer iklim etkileri, Fiji, Vanuatu ve Solomon adaları gibi

Pasifik Okyanusu ülkelerindeki insanları şimdiden yer değiştirmeye zorlamaktadır. Deniz seviyesinde beklenen bu yükselmenin insanları evlerini terk etmeye, varsa daha yüksek yerlere taşınmaya veya nihayetinde göç etmeye zorlayacağı ön görülmektedir.

Kaynak:

Siegert, M. J. (2017). Why should we worry about sea level change? *Frontiers for Young Minds, 5*. https://doi.org/10.3389/frym.2017.00041

NRDC. (n.d.). *Sea level rise 101.* Retrieved October 8, 2024, from https://www.nrdc.org/stories/sea-level-rise-101

NASA. (2014, August 26). *Warming seas and melting ice sheets.* NASA Climate. https://climate.nasa.gov/news/2328/warming-seas-and-melting ice-sheets/

UN News. (2024, August 26). *What is sea level rise and why does it matter to our future?*

Lütfen aşağıda verilen soruları cevaplayınız.

1. Metindeki bilgiler doğrultusunda deniz seviyesindeki yükselmenin nedeni nedir?

2. Metindeki bilgiler doğrultusunda deniz seviyesindeki yükselmenin insanlar üzerinde ne gibi etkileri olmuştur?

3. Uzmanlar gelecekte deniz seviyesindeki artışın nelere yol açabileceğini öngörmektedir?

4. Bu metinde geçen iddia, gerekçe ve kanıt bileşenlerini bularak altını çiziniz.

APPENDIX L

ENGLISH AND TURKISH VERSIONS OF GUIDING QUESTIONS

Questions I can ask the chatbot:

1. Which components of my argument need improvement?
2. How can I make my claim better?
3. What kind of evidence and reasoning can I add to support my claim?
4. How can I make my argument stronger?
5. How can I express my argument in a more scientific way?
6. Is the evidence I provided sufficient?
7. What other inferences can be made from all this data?

While evaluating the chatbot's response:

1. Which parts do you think are strong, and why are they strong?
2. For what reason did you choose to use this particular response?
3. What kind of features did the response have?

Chatbot'a sorabileceğim sorular:

1. Argümanımın hangi bileşenlerinin geliştirilmesi gerekiyor?
2. İddiamı nasıl daha iyi hale getirebilirim?
3. İddiamı desteklemek için ne tür kanıtlar ve gerekçeler ekleyebilirim?
4. Argümanımı nasıl daha güçlü hale getirebilirim?
5. Argümanımı daha bilimsel bir şekilde nasıl ifade edebilirim?
6. Sunduğum kanıtlar yeterli mi?
7. Tüm bu verilerden başka hangi çıkarımlar yapılabilir?

Chatbot'un yanıtını değerlendirirken:

1. Hangi bölümlerin güçlü olduğunu düşünüyorsun ve neden güçlüler?
2. Bu yanıtı kullanmayı neden seçtin?
3. Yanıtın ne tür özelliklere sahipti?

APPENDIX M

ENGLISH AND TURKISH VERSIONS OF THE EXAMPLE OF THE PARTICIPANT INTERFACE

The Participant: Hello

The Chatbot**:** Hello! I’d be happy to help with the arguments written by students. I will provide feedback by evaluating the components of claim, evidence, and reasoning. You can share your argument when you're ready!

The Participant: My claim is: "In many regions, the effects of sea level rise on communities have already started to become visible." How can I make this claim stronger?

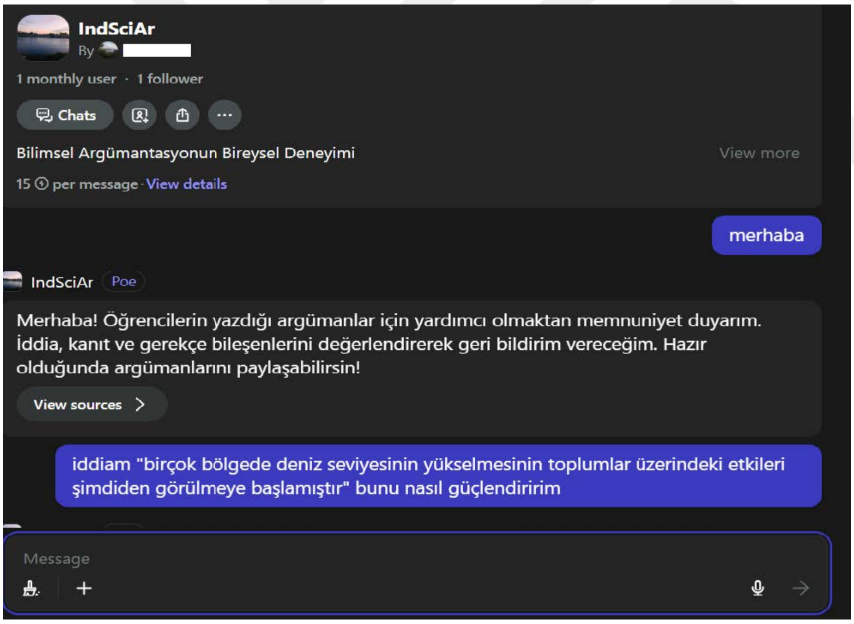


Figure M1 Turkish version of screenshot from the participant interface

REFERENCES

Abdelghani, R., Murayama, K., Kidd, C., Sauzéon, H., & Oudeyer, P.-Y. (2025). *Investigating middle school students' question-asking and answer-evaluation skills when using ChatGPT for science investigation* [Preprint]. arXiv. https://arxiv.org/abs/2505.01106

Bächtold, M., Pallarès, G., De Checchi, K., & Munier, V. (2022). Combining debates and reflective activities to develop students' argumentation on socioscientific issues. *Journal of Research in Science Teaching*, *60*(4), 761–806. https://doi.org/10.1002/tea.21816

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, *94*(5), 765–793. https://doi.org/10.1002/sce.20402

Berland, L. K., & Hammer, D. (2011). Framing for scientific argumentation. *Journal of Research in Science Teaching*, *49*(1), 68–94. https://doi.org/10.1002/tea.20446

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brockriede, W. (1975). Where is argument? *the Journal of the American Forensic Association*, *11*(4), 179–182. https://doi.org/10.1080/00028533.1975.11951059

Chang, J., Park, J., & Park, J. (2023). Using an artificial intelligence chatbot in scientific inquiry: Focusing on a Guided-Inquiry activity using InquiryBot. *Asia-Pacific Science Education*, *9*(1), 44–74. https://doi.org/10.1163/23641177-bja10062

Clarke, V., & Braun, V. (2016). Thematic analysis. *The Journal of Positive Psychology*, *12*(3), 297–298. https://doi.org/10.1080/17439760.2016.1262613

Climate Central. (n.d.). *Sea level rise map* [Interactive map]. Retrieved October 8, 2024, from https://coastal.climatecentral.org/map/9/35.2506/37.1851/?theme=water_level&map_type=water_level_above_mhhw&basemap=roadmap&contiguous=true&elevation_model=best_available&refresh=true&water_level=1.0&water_unit=m

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers and Education*, *210*, 104967. https://doi.org/10.1016/j.compedu.2023.104967

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287–312. https://doi.org/10.1002/(sici)1098-237x(200005)84:3

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education*, *88*(6), 915–933. https://doi.org/10.1002/sce.20012

Faize, F. A., Husain, W., & Nisar, F. (2017). a critical review of scientific argumentation in science education. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(1). https://doi.org/10.12973/ejmste/80353

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 1–15. https://doi.org/10.1080/14703297.2023.2198446

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, *5*(1), 80–92. https://doi.org/10.1177/160940690600500107

Frey, B. B., Ellis, J. D., Bulgren, J. A., Hare, J. C., & Ault, M. (2015). Development of a test of scientific argumentation. *The Electronic Journal of Science Education*, *19*(4), 1–18. https://files.eric.ed.gov/fulltext/EJ1188271.pdf

Gautier, C. (2012). A new type of debate for global warming and scientific literacy. *Cultural Studies of Science Education*, *7*(3), 683–691. https://doi.org/10.1007/s11422-012-9417-z

Gay, L. R., Mills, G. E., & Airasian, P. W. (2011). *Educational research: Competencies for analysis and applications* (10th ed.). Pearson Education.

Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2023). Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. *Computers and Education*, *203*, 104862. https://doi.org/10.1016/j.compedu.2023.104862

Haudek, K.C., Zhai, X. Examining the effect of assessment construct characteristics on machine learning scoring of scientific argumentation. *Int J Artif Intell Educ* (2023). https://doi.org/10.1007/s40593-023-00385-8

Heeg, D. M., & Avraamidou, L. (2024). Young children's understanding of AI. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-13169-x

Henderson, J. B., McNeill, K. L., González- Howard, M., Close, K., & Evans, M. J. (2017). Key challenges and future directions for educational research on scientific argumentation. *Journal of Research in Science Teaching*, *55*(1), 5–18. https://doi.org/10.1002/tea.21412

Henderson, J. B., & Aguilera, E. (2020). Utilizing technology to support scientific argumentation in active learning classrooms. In J. J. Mintzes & E. M. Walter (Eds.), *Active learning in college science* (pp. 587–602). Springer. https://doi.org/10.1007/978-3-030-33600-4_36

Hendratmoko, A. F., Madlazim, M., Widodo, W., Suyono, S., & Supardi, Z. a. I. (2023). Inquiry and debate in science learning: Potential strategy for improving students’ scientific Argumentation skills. *International Journal of Education in Mathematics, Science and Technology*, *12*(1), 114–138. https://doi.org/10.46328/ijemst.3152

Ho, H.-Y., Chang, T.-L., Lee, T.-N., Chou, C.-C., Hsiao, S.-H., Chen, Y.-H., & Lu, Y.-L. (2019). Above- and below-average students think differently: Their scientific argumentation patterns. *Thinking Skills and Creativity, 34*, 100607. https://doi.org/10.1016/j.tsc.2019.100607

IPCC. (2023). *Climate change 2023: Synthesis report.* H. Lee & J. Romero (Eds.), Intergovernmental Panel on Climate Change. https://www.ipcc.ch/report/ar6/syr/

IPCC. (n.d.). *IPCC 1.5°C raporu.* Birbuçuk Derece. Retrieved October 8, 2024, from https://www.birbucukderece.com/bilimsel-kaynaklar/ipcc-1-5-c-raporu

Jay, A. K., Crimmins, A. R., Avery, C. W., Dahl, T. A., Dodder, R. S., Hamlington, B. D., Lustig, A., Marvel, K., Méndez-Lazaro, P. A., Osler, M. S., Terando, A., Weeks, E. S., & Zycherman, A. (2023). Overview. In A. R. Crimmins (Ed.), *Fifth national climate assessment.* U.S. Global Change Research Program. https://doi.org/10.7930/nca5.2023.ch1

Jeffry, L., Ong, M. Y., Nomanbhay, S., Mofijur, M., Mubashir, M., & Show, P. L. (2021). Greenhouse gases utilization: A review. *Fuel, 301*, 121017. https://doi.org/10.1016/j.fuel.2021.121017

Jiménez-Aleixandre, M. P., & Erduran, S. (2007). Argumentation in science education: An overview. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 3–27). Springer. https://doi.org/10.1007/978-1-4020-6670-2_1

Jin, Q., & Kim, M. (2021). Supporting elementary students’ scientific argumentation with argument-focused metacognitive scaffolds (AMS). *International Journal of Science Education*, *43*(12), 1984–2006. https://doi.org/10.1080/09500693.2021.1947542

Kaçar, S., & Balım, A. G. (2018). An activity related to the use of argumentation-driven inquiry methods in electricity energy topic. *Journal of Inquiry Based Activities*, *8*(2), 127–149. Retrieved from https://ated.info.tr/index.php/ated/article/view/30

Kızkapan, O., & Nacaroğlu, O. (2020). Gifted students' views and argument structures on a socio-scientific issue: Cloning. *Acta Didactica Napocensia*, *14*(2), 217–224. https://doi.org/10.24193/adn.14.2.26

Kim, D., Majdara, A., & Olson, W. (2024). A pilot study inquiring into the impact of ChatGPT on lab report writing in introductory engineering labs. *International Journal of Technology in Education*, *7*(2), 259–289. https://doi.org/10.46328/ijte.691

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review, 62*(2), 155–178. https://doi.org/10.17763/haer.62.2.9r424r0113t670l1

Kuki, A. D., Agustini, R., & Azizah, U. (2023). Analysis of effectiveness argument-driven inquiry to improve students' argumentation skill and conceptual understanding. *IJORER*, *4*(3), 329–342. https://doi.org/10.46245/ijorer.v4i3.316

Lee, H., Gweon, G., Lord, T., Paessel, N., Pallant, A., & Pryputniewicz, S. (2021). Machine learning-enabled automated feedback: Supporting students' revision of scientific arguments based on data drawn from simulation. *Journal of Science Education and Technology*, *30*(2), 168–192. https://doi.org/10.1007/s10956-020-09889-7

Lee, J. H., Shin, D., & Hwang, Y. (2024). Investigating the capabilities of large language model-based task-oriented dialogue chatbots and the impact of "suggest replies" function from a learner's perspective. *System*, *127*, 103538. https://doi.org/10.1016/j.system.2024.103538

Lieskovský, J., & Sunyík, V. (2022). How to enhance scientific literacy? Review of interventions focused on improving high school students' scientific reasoning skills and attitudes toward science. *Československá Psychologie*, *66*(1), 30-45. https://doi.org/10.51561/cspsych.66.1.30

Lin, S., & Mintzes, J. J. (2010). Learning argumentation skills through instruction in socioscientific issues: The effect of ability level. *International Journal of Science and Mathematical Education/International Journal of Science and Mathematics Education*, *8*(6), 993–1017. https://doi.org/10.1007/s10763-010-9215-6

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Thousand Oaks, CA: Sage.

Lo, C. K., Hew, K. F., & Jong, M. S. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers and Education/Computers & Education*, *219*, 105100. https://doi.org/10.1016/j.compedu.2024.105100

Lu, X., Sales, A., & Heffernan, N. T. (2021). Immediate versus delayed feedback on learning: Do people's instincts really conflict with reality? *Journal of Higher Education Theory and Practice, 21*(16). https://doi.org/10.33423/jhetp.v21i16.4925

Martineau, K. (2024, September 1). *What is generative AI?* IBM Research. https://research.ibm.com/blog/what-is-generative-AI

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*, *15*(2), 153–191. https://doi.org/10.1207/s15327809jls1502_1

McNeill, K. L., & Krajcik, J. S. (2011). *Supporting grade 5–8 students in constructing explanations in science: The claim, evidence, and reasoning framework for talk and writing*. Pearson Allyn & Bacon.

Mentor, I. (2024, August 22*). IELTS reading practice: The effects of global warming on weather extremes*. Retrieved August 22, 2024, from https://www.ielts.net/the-effects-of-global-warming-on-weather-extremes/

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation* (4th ed.). John Wiley & Sons.

Ministry of National Education. (2018). *Fen bilimleri dersi öğretim programı* [Science curriculum]. Retrieved September 8, 2024, from https://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=325

Ministry of National Education. (2024). *Fen bilimleri dersi öğretim programı* [Science curriculum]. Retrieved September 8, 2024, from https://mufredat.meb.gov.tr/

Mitchell, J. F. B. (1989). The "Greenhouse" effect and climate change. *Reviews of Geophysics*, *27*(1), 115–139. https://doi.org/10.1029/rg027i001p00115

NASA. (n.d.). *Interactives – Climate change: Vital Signs of the planet*. Retrieved December 10, 2024, from https://climate.nasa.gov/interactives/climate-time-machine/?intent=021

NASA. (n.d.). *Global ice viewer* [Interactive tool]. Retrieved October 8, 2024, from https://climate.nasa.gov/interactives/global-ice-viewer/

NASA. (n.d.). *Ice melt.* NASA Sea Level Change. Retrieved October 12, 2024, from https://sealevel.nasa.gov/understanding-sea-level/global-sea-level/ice-melt/

NASA. (n.d.). *Ice sheets*. NASA Climate. Retrieved October 12, 2024, from https://climate.nasa.gov/vital-signs/ice-sheets/

National Research Council, 2003, *Protecting participants and facilitating social and behavioral sciences research*. National Academies Press, Science Education, Washington.

Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13454

NOAA. (n.d.). Sea ice and climate. NOAA Ocean Service. Retrieved October 14, 2024, from https://oceanservice.noaa.gov/facts/sea ice-climate.html

NRDC. (n.d.). Sea level rise 101. Retrieved October 14, 2024, from https://www.nrdc.org/stories/sea-level-rise-101

Okulu, H. Z., & Muslu, N. (2024). Designing a course for pre-service science teachers using ChatGPT: What ChatGPT brings to the table. *Interactive Learning Environments*, 1–18. https://doi.org/10.1080/10494820.2024.2322462

OpenAI. (2022, November 30). *Introducing ChatGPT*. https://openai.com/blog/chatgpt

Pasca, M. G., & Arcese, G. (2024). ChatGPT between opportunities and challenges: an empirical study in Italy. *the TQM Journal*. https://doi.org/10.1108/tqm-08-2023-0268

Qun, Y. (2025). Investigating the impact of immediate vs. delayed feedback timing on motivation and language learning outcomes in online education: Perspectives from Feedback Intervention Theory. *Learning and Motivation, 90*, 102132. https://doi.org/10.1016/j.lmot.2025.102132

Rasmussen, M. V. C. (2015, August 28). Warming seas and melting ice sheets. *Climate Change: Vital Signs of the Planet*. https://climate.nasa.gov/news/2328/warming-seas-and-melting-ice-sheets/

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. Journal of Research in Science Teaching, 41, 513–536.

Sampson, V., Grooms, J., & Walker, J. P. (2010). Argument- Driven Inquiry as a way to help students learn how to participate in scientific argumentation and craft written arguments: An exploratory study. *Science Education*, *95*(2), 217–257. https://doi.org/10.1002/sce.20421

Sampson, V., Enderle, P., Grooms, J., & Witte, S. (2013). Writing to learn by learning to write during the school science laboratory: Helping middle and high school students develop argumentative writing skills as they learn core ideas. *Science Education*, *97*(5), 643–670. https://doi.org/10.1002/sce.21069

Sampson, V., & Schleigh, S. (2013). *Scientific argumentation in biology: 30 Classroom Activities*. NSTA Press.

Siegert, M. J. (2017). Why should we worry about sea level change? *Frontiers for Young Minds*, *5*. https://doi.org/10.3389/frym.2017.00041

Simon, S., Erduran, S., & Osborne, J. (2006). Learning to Teach Argumentation: Research and development in the science classroom. *International Journal of Science Education*, *28*(2–3), 235–260. https://doi.org/10.1080/09500690500336957

Songsil, W., Pongsophon, P., Boonsoong, B., & Clarke, A. (2019). Developing scientific argumentation strategies using revised argument-driven inquiry (rADI) in science classrooms in Thailand. *Asia-Pacific Science Education, 5*(1). https://doi.org/10.1186/s41029-019-0035-x

Su, Y., Lin, Y. G., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, *57*, 100752. https://doi.org/10.1016/j.asw.2023.100752

Topal, A. D., Eren, C. D., & Geçer, A. K. (2021). Chatbot application in a 5th grade science

course. *Education and Information Technologies*, *26*(5), 6241–6265. https://doi.org/10.1007/s10639-021-10627-8

Topal, A. D., Gökçe, A. T., Eren, C. D., & Geçer, A. K. (2024). Artificial intelligence literacy scale: A study of reliability and validity in Turkish university students. *Journal of Learning and Teaching in Digital Age*. https://doi.org/10.53850/joltida.1440845

Toulmin, S., Rieke, R., & Janik, A. (1984). *An introduction to reasoning.* Prentice Hall.

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers and Education/Computers & Education*, *215*, 105031. https://doi.org/10.1016/j.compedu.2024.105031

Urhan, S., Gençaslan, O., & Dost, Ş. (2024). An argumentation experience regarding concepts of calculus with ChatGPT. *Interactive Learning Environments*, 1–26. https://doi.org/10.1080/10494820.2024.2308093

U.S. Environmental Protection Agency. (2025, January 15). *Climate change indicators: Ice sheets.*

Walker, J. P., & Lower, M. A. (2019). Facilitating argumentation in the laboratory: The challenges of claim change and justification by theory. *Journal of Chemical Education*, *96*(3), 435–444. https://doi.org/10.1021/acs.jchemed.8b00745

Wang, L., Chen, X., Wang, C., Xu, L., Shadiev, R., & Li, Y. (2024). ChatGPT's capabilities in providing feedback on undergraduate students' argumentation: a case study. *Thinking Skills and Creativity*, *51*, 101440. https://doi.org/10.1016/j.tsc.2023.101440

Wen, Tsung, Yang., Yu, Ren, Lin., Hsiao-Ching, She., Kai, Yi, Huang. (2015). The effects of prior-knowledge and online learning approaches on students' inquiry and argumentation abilities. *International Journal of Science Education, 37*(10):1564-1589. doi: 10.1080/09500693.2015.1045957

Wu, Q., & Xu, A. (2025). Poe or Gemini for fostering writing skills in Japanese upper- intermediate learners: Uncovering the consequences on positive emotions, boredom to write, academic self- efficacy and writing development. *British Educational Research Journal*. https://doi.org/10.1002/berj.4119

Yang, F. J., Su, C. Y., Xu, W. W., & Hu, Y. (2022). Effects of developing prompt scaffolding to support collaborative scientific argumentation in simulation-based physics learning. *Interactive Learning Environments, 31*(10), 6526–6541. https://doi.org/10.1080/10494820.2022.2041673

Zhai, X., Haudek, K. C., & Ma, W. (2022). Assessing argumentation using machine learning and cognitive diagnostic modeling. *Research in Science Education, 53*(2), 405–424. https://doi.org/10.1007/s11165-022-10062-w

Zhang, P., & Tur, G. (2023). A systematic review of ChatGPT use in K- 12 education. *European Journal of Education*. https://doi.org/10.1111/ejed.12599

Zhou, D. (2023). “Learn to Argue” and “Argue to Learn”: meta-analysis of effective instructional design for online scientific argumentation activities. *Interactive Learning Environments*, 1–24. https://doi.org/10.1080/10494820.2023.2205904

Zhu, M., Lee, H. S., Wang, T., Liu, O. L., Belur, V., & Pallant, A. (2017). Investigating the impact of automated feedback on students’ scientific argumentation. *International Journal of Science Education, 39*(12), 1648–1668. https://doi.org/10.1080/09500693.2017.1347303

Zhu, M., Ou, L., Liu, H.-S., & Lee, H.-S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education, 143*, 103668. https://doi.org/10.1016/j.compedu.2019.103668