



**AI-GENERATED CONTENT
IMPLEMENTATION IN LANGUAGE
CLASSROOMS THROUGH INTERACTIVE
STORYTELLING**

Yüksek Lisans Yeterlik Tezi

Berkay KURTOĞLU

Eskişehir 2025

AI-GENERATED CONTENT IMPLEMENTATION IN LANGUAGE
CLASSROOMS THROUGH INTERACTIVE STORYTELLING
DİL SINIFLARINDA ETKİLEŞİMLİ HİKAYE ANLATIMI YOLUYLA YAPAY
ZEKA DESTEKLİ İÇERİK UYGULAMASI

Berkay KURTOĞLU

YÜKSEK LİSANS YETERLİK TEZİ

Yabancı Diller Eğitimi Anabilim Dalı

İngilizce Eğitimi Programı (Tezli YL) Programı

Danışman: Prof. Dr. Handan YAVUZ

Anadolu

Üniversitesi

Lisansüstü Eğitim Enstitüsü

Haziran 2025

Eskişehir

JÜRİ VE ENSTİTÜ ONAYI

Berkay KURTOĞLU'nun "Ai-Generated Content Implementation In Language Classrooms Through Interactive Storytelling" başlıklı tezi/....../20... tarihinde aşağıdaki jüri tarafından değerlendirilerek "Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca, Ana Bilim/Ana Sanat dalındaProgramında Yüksek Lisans/Doktora/Sanatta Yeterlik tezi olarak kabul edilmiştir.

		<u>Ünvanı Adı Soyadı</u>	<u>İmza</u>
Üye (Danışman)	:		
Üye	:		
Üye	:		
Üye	:		
Üye	:		

.....

Lisansüstü Eğitim Enstitüsü Müdürü

ÖN SÖZ /TEŞEKKÜR

Prior to enrolling in foreign language M.A. program, I had not realized how arduous writing an academic paper would be. Ostensibly, a thesis is no more dissimilar than any other article one could come across on the internet. But, I now understand that the meticulous work that goes into writing scientifically sets this medium apart from other kinds of written text. Thankfully, I had the backing of a great number of people who helped and encouraged me into completing my thesis. Without their assistance and guidance, this paper would have never been realized.

My most heartfelt gratitude goes to my supervisor Prof. Dr. Handan YAVUZ for her contributions to my very first academic paper from the beginning to the end. Her insights on language teaching and meticulous care in reading and re-reading the drafts allowed me to produce this paper in its highest potential. Additionally, her invaluable feedbacks and thoughts on both academic and personal matters helped me find my way along the M.A. program and my life. I also want to thank the jury members, Prof. Dr. Ümit Deniz TURAN and Asst. Prof. Işıl YALÇIN, for their inputs and corrections.

I would like to extend my thanks to İbrahim AKKAYA and Aslı AYDEMİR for their continuous support throughout the process. They graciously offered their feedback, provided me with the data I needed. I cannot thank them enough for their efforts.

I am indebted to thank my students as well, who kindly assisted me during our classes together and participated in the activity. Thanks to our mutual experiences, conversations and discussions, I was reminded why I wanted to become a teacher once again.

I would like to thank my dear family members for their help, advice and support. They helped me survive my M.A. program by helping me and cheering me on and providing their unconditional love whenever I needed them the most.

I owe a special thanks to TÜBİTAK for their financial assistance with their scholarship program 2210-A. Throughout my M.A. education, I've been financially supported by them, which allowed me to focus on my studies and my paper. I am grateful for their contributions to my thesis.

ABSTRACT

AI-GENERATED CONTENT IMPLEMENTATION IN LANGUAGE CLASSROOMS THROUGH INTERACTIVE STORYTELLING

Berkay KURTOĞLU

Department of Foreign Language
Education

Programme in English Language Teaching

Graduate School of Anadolu University, June 2025

Supervisor: Prof. Dr., Handan YAVUZ

As artificial intelligence (AI) continues to develop, researchers are increasingly interested in its potential applications in education, particularly in the realm of language learning. A literature review on AI usage revealed that research on this topic generally used AI as a supplementary tool to explicitly revise or correct errors. In addition, studies that use AI as a student-led tool of real-time lesson material creation remain sparse. Therefore, this study design explores a novel approach that leverages AI content generators, specifically those that can create customized text based on user-provided keywords as input. This study aims to investigate whether this technology can lead to improvements in students' L2 reading comprehension. To this end, this study is designed to implement a quasi-experimental design by dividing participants into two groups, with one group engaging with the AI content generator, and the other receiving traditional language instruction. Both groups took pre- and post-tests to measure any changes in their L2 reading comprehension skills. Additionally, participants' perceptions regarding the AI content generator are examined through semi-structured interviews. The results revealed that AI storytelling activity did not significantly impact the reading comprehension of the participants. However, shareholders believe it affects student motivation and engagement.

Keywords: Artificial Intelligence, Reading Comprehension, Computer-assisted Language Learning, AI-based Procedural Content Generation

10/06/2025

ÖZET

DİL SINIFLARINDA ETKİLEŞİMLİ HİKAYE ANLATIMI YOLUYLA YAPAY ZEKA DESTEKLİ İÇERİK UYGULAMASI

Berkay KURTOĞLU

Yabancı Diller Eğitimi Anabilim Dalı

İngiliz Dili Eğitimi Tezli Yüksek Lisans

Anadolu Üniversitesi Lisansüstü Eğitim Enstitüsü, Haziran 2025

Danışman: Prof. Dr., Handan YAVUZ

Yapay zekanın (YZ) gelişmesiyle birlikte, araştırmacılar YZ'nin eğitimdeki, özellikle de dil öğrenimi alanındaki potansiyel uygulamalarına giderek daha fazla ilgi duymaktadır. YZ kullanımı üzerine yapılan literatür taraması, bu konuda yapılan araştırmaların genellikle YZ'yi hataları açıkça gözden geçirmek veya düzeltmek için tamamlayıcı bir araç olarak kullandığını ortaya koymuştur. YZ'yi öğrenci liderliğindeki gerçek zamanlı ders materyali oluşturma aracı olarak kullanan çalışmaların sayısı ise henüz azdır. Bu nedenle, bu çalışmada, YZ içerik oluşturunucularından, özellikle de kullanıcı tarafından sağlanan anahtar kelimelere dayalı olarak özelleştirilmiş metinler oluşturabilenleri kullanan yeni bir yaklaşımın öğrencilerin ikinci dil (L2) okuma-anlama becerileri üzerindeki etkiyi araştırmaktadır. Bu amaca ulaşabilmek için yarı deneysel bir çalışma tasarlanmış ve katılımcılar iki gruba ayrılmıştır. Bir gruba YZ içerik oluşturunucuyla etkileşime girmesine olanak sağlanmış, diğer gruba ise geleneksel dil eğitimi ile devam edilmiştir. Her iki grup da L2 okuma- anlama becerilerindeki değişiklikleri ölçmek için ön ve son testler almıştır. Ayrıca, katılımcıların YZ'e ilişkin algıları yarı yapılandırılmış mülakatlar yoluyla incelenmiştir. Sonuçlar, yapay zekâ destekli hikâye anlatımı etkinliğinin katılımcıların okuma-anlama becerisi üzerinde anlamlı bir etkisi olmadığını ancak, paydaşların bu etkinliğin öğrencilerin motivasyonu ve derse katılımı üzerinde etkili olduğuna düşündüğünü ortaya koymuştur.

Anahtar Kelimeler: Yapay Zeka, Okuma-Anlama, Bilgisayar Destekli Dil Öğrenimi, Yapay Zeka Tabanlı Prosedürel İçerik Üretimi

10/06/2025

STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES

I hereby truthfully declare that this thesis/dissertation is an original work prepared by me; that I have behaved in accordance with the scientific ethical principles and rules throughout the stages of preparation, data collection, analysis and presentation of my work; that I have cited the sources of all the data and information that could be obtained within the scope of this study, and included these sources in the references section; and that this study has been scanned for plagiarism with “Scientific Plagiarism Detection Program” used by Anadolu University, and that “it does not have any plagiarism” whatsoever. I also declare that, if a case contrary to my declaration is detected in my work at any time, I hereby express my consent to all the ethical and legal consequences that are involved.

Berkay KURTOĞLU

ETİK İLKELERE VE KURALLARA UYGUNLUK BEYANI

Bu tez çalışmasının tarafımdan özgün olarak hazırlandığını; çalışmamın hazırlanması, veri toplama, analiz ve sunumu aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilebilecek tüm veri ve bilgilerin kaynaklarını gösterdiğimi ve bu kaynakları referanslar bölümüne eklediğimi; ve bu çalışmanın Anadolu Üniversitesi tarafından kullanılan “Bilimsel İntihal Tespit Programı” ile intihal taramasından geçirildiğini ve "hiçbir intihal içermediğini" doğru bir şekilde beyan ederim. Ayrıca, çalışmamda beyanıma aykırı bir durumun herhangi bir zamanda tespit edilmesi halinde, ilgili tüm etik ve hukuki sonuçlara rıza gösterdiğimi bildiririm.

Berkay KURTOĞLU

10/06/2025

ETİK KURUL BELGESİ BEYANNAMESİ

Evrak Kayıt Tarihi: 15.10.2024 Protokol No: 799711

Tarih: 31.10.2024



ANADOLU ÜNİVERSİTESİ
SOSYAL VE BEŞERÎ BİLİMLER BİLİMSEL ARAŞTIRMA VE YAYIN ETİĞİ KURULU
KARAR BELGESİ

ÇALIŞMANIN TÜRÜ:	Yüksek Lisans Tez Çalışması
KONU:	Eğitim Bilimleri
BAŞLIK:	AI-Generated Content Implementation In Language Classrooms Through Interactive Storytelling
PROJE/TEZ YÜRÜTÜCÜSÜ:	Prof. Dr. Handan YAVUZ
TEZ YAZARI:	Berkay KURTOĞLU
ALT KOMİSYON GÖRÜŞÜ:	-
KARAR:	Olumlu
Prof. Dr. Saim ÖNCE (Başkan-İkt. ve İdari Bil. Fak.)	
Prof. Dr. M. Erkan ÜYÜMEZ (Başkan Yardımcısı -İkt. ve İdari Bil. Fak.)	Prof. Dr. Fatime GÜNEŞ (Edebiyat Fak.)
Prof. Dr. Yıldız UZUNER (Eğitim Fak.)	Prof. Dr. Erkan YÜKSEL (İletişim Bil. Fak.)
Prof. Dr. Handan DEVECİ (Eğitim Fak.)	Prof. Dr. Kamil ÇEKEROL (Açıköğretim Fak.)

10/06/2025

DECLARATION OF GENERATIVE ARTIFICIAL INTELLIGENCE USAGE

I declare that I have not used generative artificial intelligence programs at any stage of the preparation, execution, data analysis, etc., of this study.

I also declare that, if a case contrary to my declaration is detected in my work at any time, I hereby express my consent to all the ethical and legal consequences that are involved.



Berkay KURTOĞLU

10/06/2025

ÜRETKEN YAPAY ZEKA KULLANIM BEYANI

Bu çalışmanın hazırlanması, yürütülmesi, veri analizi vb. aşamalarının hiçbirinde üretken yapay zeka programlarını kullanmadığımı beyan ederim.

Ayrıca, çalışmamda beyanıma aykırı bir durumun herhangi bir zamanda tespit edilmesi halinde, ilgili tüm etik ve hukuki sonuçlara rıza gösterdiğimi bildiririm.

Berkay KURTOĞLU



TABLE OF CONTENTS

JÜRİ VE ENSTİTÜ ONAYI.....	iii
ÖN SÖZ /TEŞEKKÜR.....	iv
ABSTRACT.....	iii
ÖZET	iv
STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES	v
ETİK İLKELERE VE KURALLARA UYGUNLUK BEYANI	vi
ETİK KURUL BELGESİ BEYANNAMESİ	vii
DECLARATION OF GENERATIVE ARTIFICIAL INTELLIGENCE USAGE.....	viii
ÜRETKEN YAPAY ZEKA KULLANIM BEYANI	ix
TABLE OF CONTENTS	x
LIST OF TABLES	xii
LIST OF GRAPHS	xiii
LIST OF FIGURES	xiv
ABBREVIATIONS.....	xv
1 INTRODUCTION	1
2 LITERATURE REVIEW	4
2.1 Reading Comprehension	4
2.1.1 Linguistic Knowledge	4
2.1.2 Orthographic Knowledge	4
2.1.3 General Knowledge.....	5
2.2 Artificial Intelligence and L2 Teaching	10
2.2.1 AI Dungeon.....	13
3 METHODOLOGY	17
3.1 Research Design.....	17
3.2 Participants.....	18
3.3 Intervention	19

3.3.1 Control Group	19
3.3.2 Experimental Group	19
3.4 Data Collection	22
3.4.1 English Assessment Tests	22
3.4.2 Semi-structured Interviews	23
3.5 Data Analysis	24
3.5.1 Pre-test and Post-test	24
3.5.2 Interviews	24
4 RESULTS	25
4.1 Q1: Do AI-based interactive storytelling tools improve general reading comprehension?	25
4.2 Q2: What are the perceptions of the stakeholders (i.e., the English teacher, headmaster of the school and students in the experimental group) regarding AI-based reading activities? ..	27
4.2.1 Effects on Student Motivation	28
4.2.2 Impressions on AI and AI-based Reading Activity	31
5 DISCUSSION AND CONCLUSION	35
REFERENCES.....	40
APPENDIX.....	44
A1: Interview Questions and Translations.....	44

LIST OF TABLES

	<u>Page</u>
Table 3.1 English Assessment Test Parts.....	23
Table 4.1 Mean Scores and Standard Deviations for the Control and Experimental Groups	26



LIST OF GRAPHS

	<u>Page</u>
Graph 3.1 Action Flow.....	18



LIST OF FIGURES

	<u>Page</u>
Figure 2.1 Text Generation Settings Menu and Image Generation Menu	14
Figure 2.2 User Interface and Side Bar Menu.....	14
Figure 2.3 AI Responding to Student Prompt with an Image.....	15
Figure 4.1 Distribution of the Pretest Scores in Percentages	25
Figure 4.2 Pre-test to Post-test Score Changes for Each Student in the Experimental and Control Groups	27

ABBREVIATIONS

AI	: Artificial Intelligence
CUI	: Conversational User Interfaces
CACD	: Computer Assisted Classroom Discussion
EFL	: English as a Foreign Language
ELT	: English Language Teaching
ESL	: English as a Second Language
LLM	: Large Language Model
NLP	: Natural Language Processing

1 INTRODUCTION

Storytelling has long been recognized as a valuable tool for facilitating the development of young learners, particularly their language skills. By engaging with captivating narratives, students are presented with opportunities to ask questions and formulate responses in clear, concise sentences. However, traditional storytelling methods often rely on readily available resources such as fairy tales, picture books, and basic classroom materials. This conventional approach, while effective, may not fully utilize the potential of fostering a truly interactive and creative learning environment. The emergence of the digital age has introduced a wealth of technological advancements that can significantly enhance the storytelling experience in the classroom. Among these advancements are mobile applications, readily accessible on devices such as tablets or smartphones. When integrated into the curriculum, these interactive applications can transform storytelling from a passive experience into a dynamic and engaging one. This shift can unlock a new dimension of learning by encouraging active participation from students. Through the use of technology-based storytelling tools, educators can create an immersive learning atmosphere. Interactive elements within these applications can prompt students to actively engage with the narrative, allowing them to make choices, solve problems, and influence the story's direction. This allows deeper comprehension of the material and provides a platform for students to practice critical thinking and communication skills in a stimulating and interactive way. Although effective, these advancements are rendered outdated compared to the product of cutting-edge technology; artificial intelligence (AI).

Contemporary education is undergoing a major transformation with the rise of AI. This powerful technology can revolutionize teaching methods, tailor learning to the individual needs of the students, and streamline teachers' duties. English language teaching (ELT) exemplifies the potential of such technology. AI-powered tools and technologies are being utilized to create more immersive and engaging learning experiences for students at all grade levels, to personalize instruction to individual student needs, and to provide prompt feedback (Chen *et al.*, 2021). Advancements in AI, particularly in natural language processing (NLP), have enabled the creation of AI-powered conversational user interfaces (CUI). CUIs are situated under the umbrella term "user interface", which means methods of converting a user's stimuli to a language a computer can understand. Chatbots are considered one of the two main interface types; one being text-based, the category

they fall under, and the other type being a voice-based interface. These fine-tuned chatbots can dynamically generate discourses based on user input, creating personalized and engaging dialogues and even provide customer support (Landay *et al*, 2019).

In addition to AI tools that are designed specifically to be integrated into language education and provide explicit feedback, such as chatbots, there are also AI-based applications developed with the intent of recreation. The continuously developing field of large language models (LLM), a machine learning model with self-supervised learning capabilities, has led to the development of various AI-powered generative language processors. These language processors, generally referred to as Generative Pre-trained Transformers (GPT), use digital neural network structures called transformers, trained to learn from massive amounts of text, predict the next word in a sequence and create coherent and cohesive text for different purposes (Yenduri *et al*, 2024).

Such chatbots can be trained specifically for use in narrative applications. LLMs have advanced to a stage in creative writing by which stories can be created spanning more than dozens of pages in just a few seconds with a few simple prompts and minor edits (Gunser *et al*, 2021). Although generative AI could be too verbose and repetitive at times, AI-based interactive storytelling certainly has the potential to significantly improve second language acquisition, particularly in children. Considering the level of sophistication of storytelling LLMs over chatbots, it could be promising to leverage that fact and explore methods of facilitation of alternative AI tools such as those in language classrooms and elucidate their advantages, limitations and effects. Increased level of student autonomy and adaptability that can be implemented through such tools could provide instructors with unprecedented potential to tailor the lesson content to best resonate with students' interests. Accordingly, further research on the topic that investigates the methods and relevant use of AI storytelling applications to enhance the language performance and motivation of students is necessary. Therefore, this study aims to provide answers to the following questions:

1. Do AI-based interactive storytelling tools improve general reading comprehension?
2. What are the perceptions of the stakeholders (i.e., the English teacher, headmaster of the school and students in the experimental group) regarding AI-based reading activities?

Time Limitation

This experiment lasted 10 weeks including the introduction session and time allocated for the experiment for each week was limited to two hours due to reasons that were beyond the control of the researcher.

Subjects' Language Competence

It is assumed that experimental and control groups would display comparable levels of performance before the commencement of the study. The institution where this study took place assigned students according to their language proficiency at the beginning of the semester. A proficiency exam is administered before the beginning of the semester and based on their scores, students are placed in classes. Naturally, this creates a not insignificant mean score difference between a higher-ranking group compared to a lower-ranking one. As the researcher did not have a say in the administrative practices of the institution, it was not possible to redistribute students based on their pre-test scores. Thus, students in both experimental and control groups are assumed to have similar proficiency levels.

2 LITERATURE REVIEW

2.1 Reading Comprehension

Understanding a written text is a surprisingly intricate mental activity. Even a seemingly deceptively simple sentence might require a lot of work on the reader's part. First, to make sense of a written text, readers have to process the words visually, recognizing the letters (decoding) and understanding their sounds (phonology) and written forms (orthography). They then need to connect these words using the rules of grammar (syntax) to grasp the overall meaning of the sentence (Perfetti & Stafura, 2014). Researchers suggest that a reader's knowledge toolkit consists of three main components: linguistic knowledge, orthographic knowledge and general (background) knowledge (Perfetti & Stafura, 2014; Tighe & Schatschneider, 2016; Kendeou *et al.*, 2016,).

2.1.1 Linguistic Knowledge

Linguistic knowledge is defined as foundational knowledge that enables a reader to understand the mechanics of a language such as syntax, phonology, morphology, and semantic knowledge. These methodical aspects of the language encompass the cognitive frameworks necessary for smooth and effective communication. Syntax, which pertains to the functions and combinations of smaller units of a language (morphemes) and arrangements of sentence structure, used in conjunction with morphology, the internal structure of words, and semantic knowledge, the meanings we understand from words and expressions within writing systems, to create an internalized system.

2.1.2 Orthographic Knowledge

Orthographic knowledge typically refers to the understanding of how words are spelled or an awareness of the particular patterns of letters utilized in writing systems (Apel, 2011). Understanding how words are spelled, or orthographic knowledge, plays a crucial role in our ability to quickly recognize and read words. This knowledge is not innate. It is widely accepted that we develop it through repeated exposure to written language and formal or informal instruction. Formal education is a type of structured and explicit teaching method which encompasses preschool, primary, secondary, and higher education, while informal education includes all educational activities, either organized or organic, that occur alongside or outside of formal education (Güneş & Tanju, 2020). Essentially, through these repeated exposures, we learn the patterns and regularities of

spelling through reading and seeing words in written material over and over again. But the underlying principles are not yet clear (Apel, 2011). What is clear is that non-phonological skills, such as orthographic knowledge, have a substantial impact on the proper acquisition of reading comprehension, as reading skill by itself is insufficient in acquiring orthographic knowledge (Fletcher-Flinn & Thompson, 2004; Proctor *et al.*, 2005).

2.1.3 General Knowledge

The first step of reading comprehension is correctly identifying and decoding words. Enabling one to perform these actions requires morphological understanding, familiarity with syntax rules, and a solid vocabulary foundation. In addition to these requirements, there is an often-neglected factor which influences our ability to grasp the topic of a text. This foundation is called general knowledge. In written texts, it is often assumed that the reader has a certain predisposition and a certain level of familiarity with the subject matter and the terms related to it. It is necessary to make such assumptions as they allow writers to avoid tedium and improve the fluency of the text.

General knowledge plays a crucial role in reading comprehension and growth across various educational levels and language backgrounds (Kozminsky & Kozminsky, 2001). Having background knowledge allows students to improve their comprehension and facilitates new learning since adding knowledge to pre-existing schemas is much easier than creating and organizing new ones outright (Piaget, 1923). In second language reading specifically, both general reading knowledge (metacognitive awareness) and language-specific knowledge (vocabulary and syntax) significantly predict comprehension, with the latter factor being the stronger predictor (Guo & Roehrig, 2011). Therefore, one must adhere to technical aspects for accuracy while refraining from overlooking the fact that the importance of knowledge in reading extends to topic, domain, general, and cultural knowledge, all of which are closely related to reading skill and language development.

This language system, constituted by the aforementioned three types of knowledge, allows us to decode and comprehend sentences of infinite combinations and variety. The conjoined application of these three components is essential for comprehension. Linguistic and orthographic knowledge, which are acquired through explicit instruction, is applied strategically with general knowledge, familiarity of language used in academics

that is acquired through curriculum, literature or the internet, and the meaning of the words within sentences and language structures between them are interpreted (Kendeou *et al.*, 2016). These three components are not exclusive of each other. For example, vocabulary is shown to be a strong, for adolescents, if not the strongest, predictor of reading comprehension and a component of the umbrella term that is the reading comprehension abilities (Cromley *et al.*, 2010). While vocabulary knowledge might entail connections made in our lexicon between meanings and words, it can also involve spelling, pronunciation, morphemes, and syntax as well. The interconnected influence of multiple cognitive and linguistic processes on reading comprehension can be seen throughout the literature.

Although comprehension on a sentence-by-sentence basis is not sufficient for a complete comprehension of a text, it is the beginning. The concept of linguistic knowledge involves not only accurate usage but also an understanding of how it is adapted to achieve certain goals, interact socially or convey a specific sentiment. True comprehension lies within the integration of meaning across sentences, contextual knowledge, and establishing cohesion in the process. Linking the ideas across sentences requires one to use their existing knowledge of the world to fill in any information gaps and make educated guesses (inferences) about what the author might be implying. Additionally, the text's overall structure, theme and author's purpose of writing must be considered to properly decipher the message being conveyed. The successful completion of these tasks results in the formation of a mental representation, termed the situation model, that reflects the overall meaning of the text in question, constituting the concept of general reading comprehension. Many factors influence this complex process. Reader characteristics, the properties of the text itself, and the specific demands of the reading task all play a role. While sufficient reading comprehension is crucial for foundational learning, it is also necessary for deeper forms of learning and the development of essential 21st-century skills (Graeser, 2015).

Until recently, the single most influential factor in reading comprehension was considered to be the reader's language proficiency (Kendeou *et al.*, 2016). This proficiency is constantly tapped upon throughout the reading process, with different aspects being utilized at various stages. Most of the time, isolated one-off-never-seen-again questions that could be answered with a glance at the text are given as a measurement of

comprehension. As such, instructional applications often fail to address the reality that most texts encountered by students are of mixed or multiple structure varieties (Shohamy, 1984). Language instruction failing to cover complex/multiple texts with diverse organization and approaches, such as digitalization, may be insufficient in fostering effective information management, reading and learning (Payne & Reader, 2006; Heyns & Van Zaanen, 2022). While word recognition, fluency and syntactic understanding are essential, the facilitative impact of orthography and topic knowledge on reading comprehension must also be recognized. Perfetti and Stafura's (2014) three key knowledge sources that contribute to reading comprehension (linguistic, orthographic and general knowledge) have been demonstrated to have an impact on various aspects of reading, with the general knowledge also relating to academic language that includes academic knowledge and vocabulary. Despite the established importance of general knowledge in reading, incorporating both text-specific knowledge and general knowledge into reading instruction has only recently gained prevalence (Cervetti & Hiebert, 2015). Promising results have been achieved with instructional programs that teach students to recognize different text structures, such as cause-and-effect, comparison, and problem-solution structures requiring practical, context-specific information. However, this work has primarily focused on elementary school students and specific informational text structures.

The positive impact of content knowledge in subjects like social studies on reading comprehension (Vaughn *et al.*, 2013) highlights the need to explore methods for integrating content knowledge into effective reading instruction. What we need to ensure is that the test items that we design capture students' reading comprehension ability. For more challenging reading skills, such as reading to integrate information or summarize, it is important that students are provided with opportunities to demonstrate understanding. In EFL (English as a Foreign Language) or ESL (English as a Second Language) contexts, reading comprehension is usually measured through objective test items such as gap-fills, true/false, and multiple-choice questions (Royer *et al.*, 1981, Fletcher, 2006). Such test items are highly structured and require specific responses. Those items cannot test general knowledge due to their particular nature being incompatible with traditional assessment methods. Initiatives to pursue innovations for the integration and assessment of multiple sources of knowledge pertaining to reading comprehension through a variety of media, such as technology, may be needed.

A potential candidate for that medium may be storytelling. Storytelling has long been recognized as a fundamental human activity, essential for communication and understanding the world around us. In the context of second language acquisition (SLA), storytelling is increasingly acknowledged as a valuable educational tool that can enhance language learning across various age groups and settings. Not only does it enhance language skills and comprehension but also supports cognitive and emotional development (Speaker *et al.*, 2004).

Friedberg (1994) explains storytelling as a developmentally sensitive tool. He argues that storytelling caters to the specific cognitive and linguistic capabilities of children at different stages of growth. Additionally, his research suggests that storytelling fosters a more ordered sense of the world by eliciting thoughts and experiences that are then organized within a narrative structure. Further emphasis is made on the dual benefit of storytelling for language development. According to the literature, it promotes expressive language development in speech, as children actively participate in the storytelling process (Speaker, 2004). This active engagement strengthens the fluency of verbal expression through the creative-predictive process inherent in storytelling. Research also highlights the positive impact on receptive language development in both reading and listening skills. The creative and predictive processes engaged during active storytelling strengthen a child's grasp of the intricate aspects of language structure, yet the true value lies in the holistic application of these techniques to refine overall language abilities (Cherry-Cruz, 2001). A recurring theme of research is that verbal expression acquisition is facilitated if the learner is actively participating in the storytelling process (Friedberg, 1994). This process strengthens a child's grasp of the intricate details of language structure. An amalgamation of active participation, holistic application of techniques, creative prediction, and engagement with language structure results in optimized refinement of a child's overall language aptitude. However, usually, a story is told by a narrator and the audience passively listens. Nonetheless, storytelling proves to be a versatile tool that can be adapted to digital formats, inclusive classrooms, and various cultural contexts.

Throughout history, storytelling has served as a valuable pedagogical tool, functioning as a proven instructional strategy for fostering language development in students. By encouraging the use of imagination and creativity, storytelling demonstrably enhances

reading, writing, and verbal proficiency (Speaker, 2004). The rich repository of context and variety found within different stories could help establish the knowledge base necessary for general comprehension as well, since it provides a natural context for language exposure, helping learners remember vocabulary, grasp grammatical structures, and develop habits for using language fluently. Not only that, it also promotes students to make connections with their pre-existing schemas and experiences to figure out the plot elements, implications, hidden details and motives of the characters in the story (Santi, 2023). The method is especially promising as older children are more likely to have developed the abilities necessary, such as creativity and language skills, to contribute to the story. However, in the contemporary technological landscape, language educators face significant challenges in motivating students towards reading plain, written text due to their prevalent technological engagement. As digital natives, students today are deeply immersed in the digital world. Thus, educators must strive to integrate contemporary technological tools into their pedagogical practices to effectively enhance students' language skills, potentially replacing traditional methods (Morgan, 2014).

Digitizing the stories and leveraging computers and the internet for their distribution on online platforms with the purpose of language education has been extensively studied. Nassim's (2018) study investigating Digital Storytelling as a pedagogical tool for English language learners demonstrates its significant benefits. Students found the process engaging due to the blend of technical and artistic elements. Participants reported increased confidence in their communication skills and enhanced collaborative abilities through story-building and project completion. The inclusion of this implementation as a graded project within their portfolio, where they collected every product pertaining to language class to track their progress, further motivated students to take this experiment seriously. Although the method provides arguments for the benefits of student involvement, they were incentivized to partake knowing that they would be graded for their performance. Nevertheless, the results of this study highlight the need for integration of technical skills into curricula, allowing students to explore various means of combining language skills with technology. Additionally, providing students with measurable and clear goals fosters motivation and facilitates language proficiency development. Moreover, emphasis on the engagement of higher-order processes and thought stimulation will not only result in a more substantial learning experience but also lead students to question and identify language lessons' objectives' authenticity (Kolling &

Shumway-Pitt, 2019). Digital Storytelling's alignment to authenticity can be demonstrated with constructivist learning theories, where students actively construct knowledge through experience, practice and collaboration, ultimately producing a product like a socially relevant piece of literature.

Another study that advocates the usage of computers in language education is Chun (1994). The findings showed that using computers through a method called Computer Assisted Classroom Discussions (CACD) to facilitate classroom discussions can significantly boost first-year foreign language learners' ability to interact effectively as CACD allows students to both start and participate in various conversation types and take more control over the flow of the discussion (Chun, 1994). However, cutting-edge technology resulted in technology that supersedes computers, namely artificial intelligence-based procedural generation. Considering the possible benefits of both computer integration and storytelling in language education, pre-service teachers and practicing educators or instructors aspiring to be digital natives need ways to interact with contemporary technology and its integration into contemporary education (Teo *et al.*, 2016).

2.2 Artificial Intelligence and L2 Teaching

Artificial intelligence is a system that can simulate human intelligence and thought processes as defined by IBM. This broad definition encompasses various technologies like machine learning, computer vision, and Natural Language Processing (NLP). AI has been powering many of the services and products we have been using for years. However, recent developments led to breakthroughs that we believed could only be seen in fiction. But, machine learning, self-improving algorithms and systems of Natural Language Processing are now a reality. Many scientific fields of research have benefited from AI and they continue to implement it more and more. Educators should consider the possibilities of AI as well. Although AI's potential to facilitate learning by reducing the work needed on the teacher's part, optimizing instruction, enhancing efficiency and improving literacy, in general, has been shown (Chen *et al.*, 2021), teachers believing in its application to a classroom environment remains at 52.1% (Xue & Wang, 2022). Integrating AI in education has the potential to revolutionize the learning process by providing a more personalized, efficient, and engaging experience for all students.

Burtenshaw (2020) highlights the benefits of AI in creating interactive narrative systems that can enhance language play and literacy skills. There is research which (Nurzaman *et al.*, 2020) further supports this finding, demonstrating the positive impact of an android-based interactive storytelling model that he developed on young learners' speaking skills. The findings showed that this model, which allows for interactive engagement with the story, was effective in improving speaking skills. This suggests that technology-based tools can be valuable in language learning, promoting interaction and fostering creativity particularly for young learners. Ammanabrolu *et al.* (2021) emphasize how a language education, that is interactive and context-wise grounded by the usage of visuals and virtual environments, can be made more focused, engaging and active using a computer system. The way the computer was taught to achieve a task and make sense of a narrative represents how a student would build their understanding when reading or listening to a language relatively closely, by remembering information and connections of schemas established in their lives. Furthermore, the progress of the computer depended on it making logical inferences, similar to how we interpret actions, cues and implications in actual discourses. The similarity of the computer to human behavior lies in how it generates commonsense expectations rather than concrete relations that are based on rigid logic and rules. For instance, person “A” gets a doctor's appointment. One might assume that “A” is suffering from some kind of disease or is down with a sickness. Yet, it doesn't necessarily discredit the possibility of him planning on getting a checkup. The ability of a computer to associate multiple possibilities with a given situation allows it to generate stories that are consistent with both logic and common sense. Therefore, grammar and narration should go hand-in-hand and reflect plot and character development to craft what can be considered a decent story. Although text input is the first step in building a foundation, findings suggest that interactive narratives supported by structured narratives and visuals can be effective environments for training language learners to develop reasoning, understanding and communication skills. These studies collectively suggest that AI-based interactive storytelling has the potential to revolutionize language acquisition.

Social interaction holds particular significance in language learning because language's primary function is communication. Hence, effective learning is more likely when learners can actively engage with other people, in this case, the story's characters and their group members. Meaningful interaction, in particular, can help learners notice new words

and use them for communicative purposes, thus enhancing their second language acquisition. Technology-supported learning environments can provide vicarious social interaction through audiovisual input; this can assist students in paying particular attention to the second language structures to be learned and interacting with them through AI storytelling. Essentially, AI-powered storytelling can help learners experience a second language more directly and naturally, either through interaction with the technology itself or with others through meaningful negotiation.

One of the emerging implementations of AI that has been popular for the last decade is chatbots. AI chatbots are essentially smart programs that can have conversations with people in a natural way, making them useful in many areas of daily life. They are commonly found as virtual assistants helping with customer service tasks. In education, AI chatbots can act as intelligent tutors, providing learning materials, prompting discussions, and giving feedback to students. They can even support human teachers by answering student questions and offering guidance around the clock, something that would be difficult or expensive to achieve with human tutors alone. Mageira *et al.* (2022) conducted a study using chatbots to test the plausibility of using chatbots in language education. Initial trials in Greek schools, both public and private, have shown promising results. It is argued that this technology offers an engaging and effective way to learn languages through interactive information and communication technology (ICT) based learning.

The increasing prevalence of AI in language learning environments signifies its growing importance in the field of language education. This development necessitates a critical examination of the utilization of NLP models, particularly in English language instruction, as breakthroughs made in the field generally originate in nations where English is either the native language or the second official language. Such an analysis would provide valuable insights into the potential benefits, drawbacks, and challenges associated with AI integration. While existing research offers inconclusive findings, some studies suggest positive outcomes for learning modules that incorporate AI for language learning, similar to the Chatbot-Human Interaction Satisfaction Model (CHISM) (Belda-Medina & Calvo-Ferrer, 2022). However, Lin *et al.* (2017) highlight potential limitations, indicating that users with minimal experience in e-learning tools exhibit lower

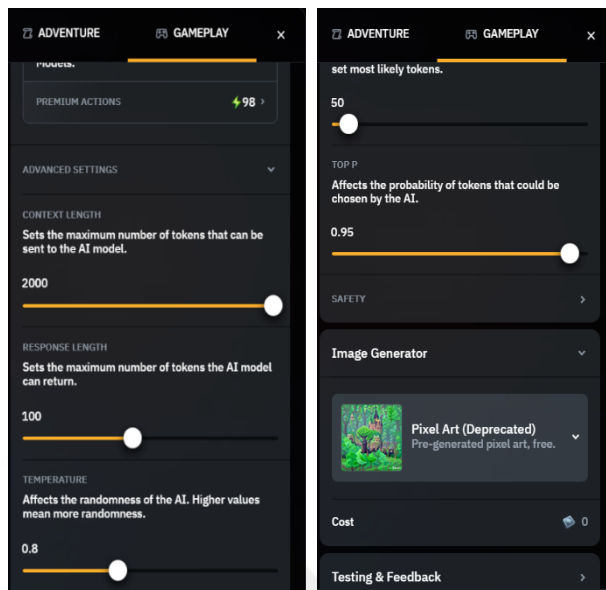
satisfaction and negative perceptions towards Intelligent Tutoring Systems (ITS) due to their divergence from traditional learning technologies.

2.2.1 AI Dungeon

Due to its unique perspective on established GPT models, practicality and user-friendly interface, this study has utilized a LLM platform known as AI Dungeon (<https://play.aidungeon.com/>). Developed by Nick Walton in 2019, it leveraged the Early GPT-2 language model to function as an automated text generation tool. Since then, its LLM have been upgraded with its newest and most advanced model, GPT-4 Turbo, with improved comprehension and higher quality generations. Considering the fact that the upkeep of an inferior LLM like GPT-3 requires an astronomical amount of funds, the availability of GPT-4 model made AI Dungeon an enticing choice among others. The program offers users a collection of pre-defined templates encompassing various genres, including, but not limited to, fantasy, mystery, apocalyptic, cyberpunk, and adventure narratives. Additionally, users can customize their stories further by selecting genre-specific options (e.g., character class in fantasy) and defining target keywords. As illustrated in Figure 2.1, users can also specify additional parameters such as desired reading difficulty level ("context length"), sentence structure ("response length ") and even utilize pixelated art generation tools that produce illustrations specific to the story being told, providing opportunity for creativity and options to its users. Finally, the program allows users to influence the overall tone of the generated text and change the appearance of the user interface.

Figure 2.1

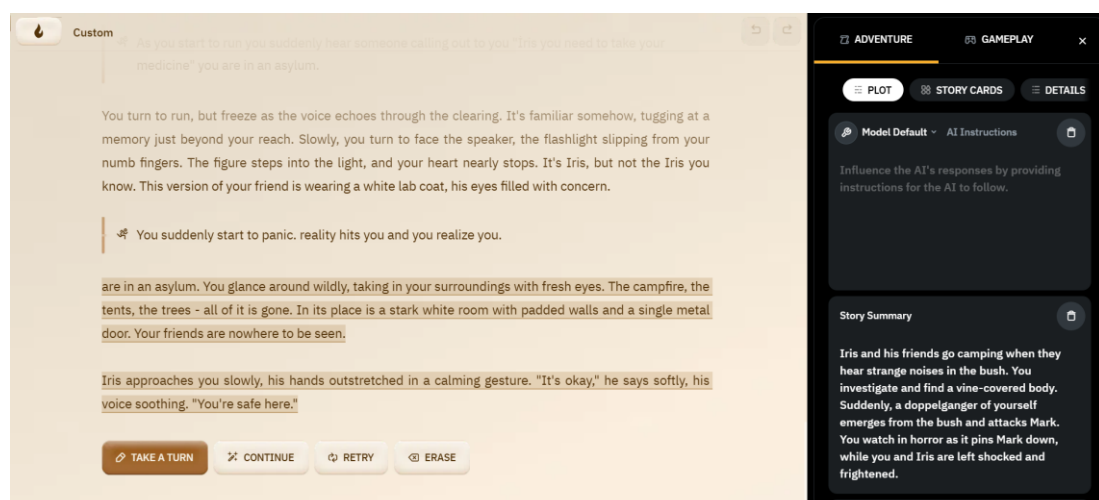
Text Generation Settings Menu (left) and Image Generation Menu (right)



To create a story, the user inputs the desired parameters regarding the type of content they desire for AI to generate. After the input is given, AI Dungeon will generate and narrate a story with a random storyline that is confined within the parameters set by the user. The AI reacts and adjusts itself based on decisions or context the user inputs in the game. Leaving the input blank lets the computer create more of the story. The user can also go back and make necessary changes to make the story flow better. User interface and a sidebar containing optional preferences such as plot components, story cards and details are depicted in Figure 2.2.

Figure 2.2

User Interface and Side Bar Menu

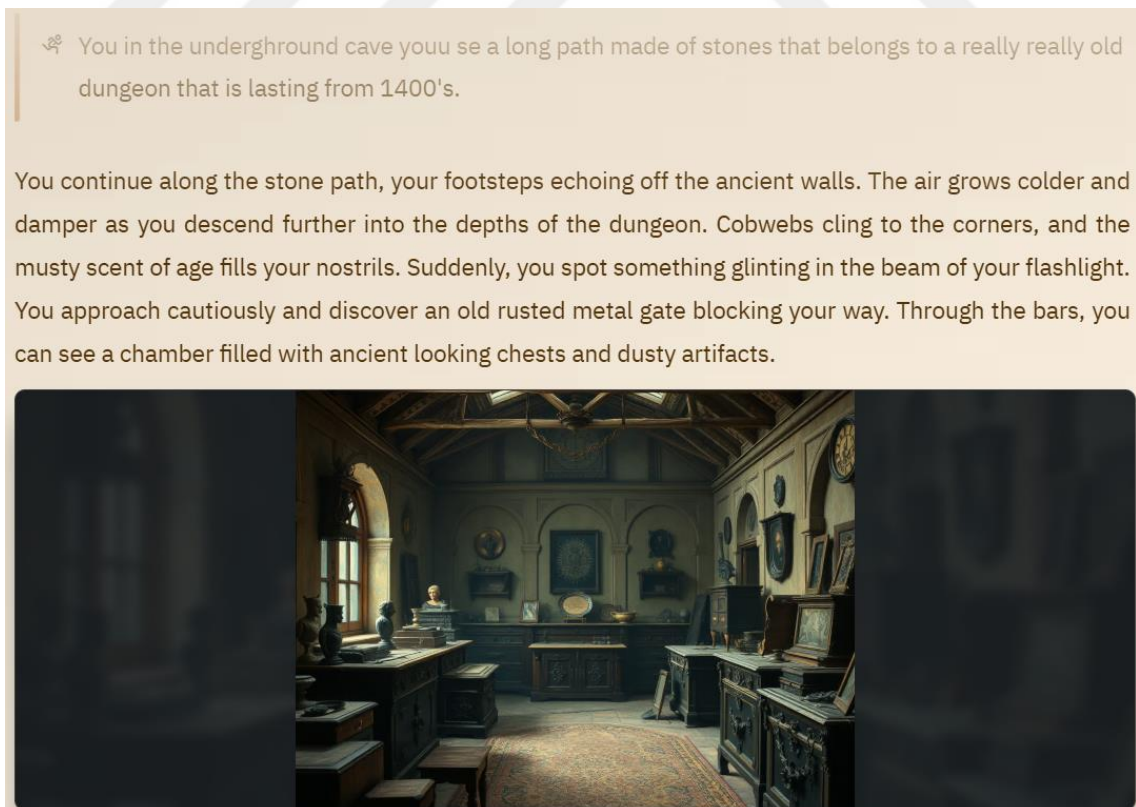


Students enter actions or words and the computer replies by creating one or more relevant paragraphs. This constructs the following part of the story. Every playthrough may vary from the others. The freedom offered by the tool allows students to shape the story, leading to a very personal, interactive, and often, one-of-a-kind tale. The game offers a space for creative thinking, where students use their minds to brainstorm. Instead of passively reading the story, they also join in its making. The narrative co-created by the AI and the students could be made even more immersive by utilizing the image-generation capabilities of the tool.

However, since AI Dungeon was not designed for educational purposes, it lacks direct error correction or grammar checking. As can be seen in Figure 2.3, this AI-based tool can compensate for minor grammar errors as long as the prompt itself does not deviate too much from the plot of the story or is not ungrammatical to the point of unintelligibility. Therefore, the utilization of implicit teaching methods or groupwork-induced scaffolding could be implemented to improve accuracy alongside comprehension.

Figure 2.3

AI Responding to Student Prompt with an Image



In addition, using advanced LLMs and creating detailed and sophisticated pictures, the user has to upgrade to one of the premium plans that the AI Dungeon has on offer. It is necessary to acquire tokens to utilize the aforementioned features. Tokens are in-app credits that can be bought for a certain amount of money which allows users to access features like a more up-to-date LLM or better image generators. Having advanced features would be beneficial as it significantly improves the responses of the AI and the pictures created by it.



3 METHODOLOGY

3.1 Research Design

This is a quasi-experimental study in which the effect of AI generated reading activities on reading skills of 7th graders was investigated. It is a mixed methods study because both quantitative and qualitative data were collected. While one 7th grade class (control group) continued to follow the school's curriculum (explained in detail below), the other 7th grade class (experimental group) received a 10-week intervention consisting of AI generated reading activities.

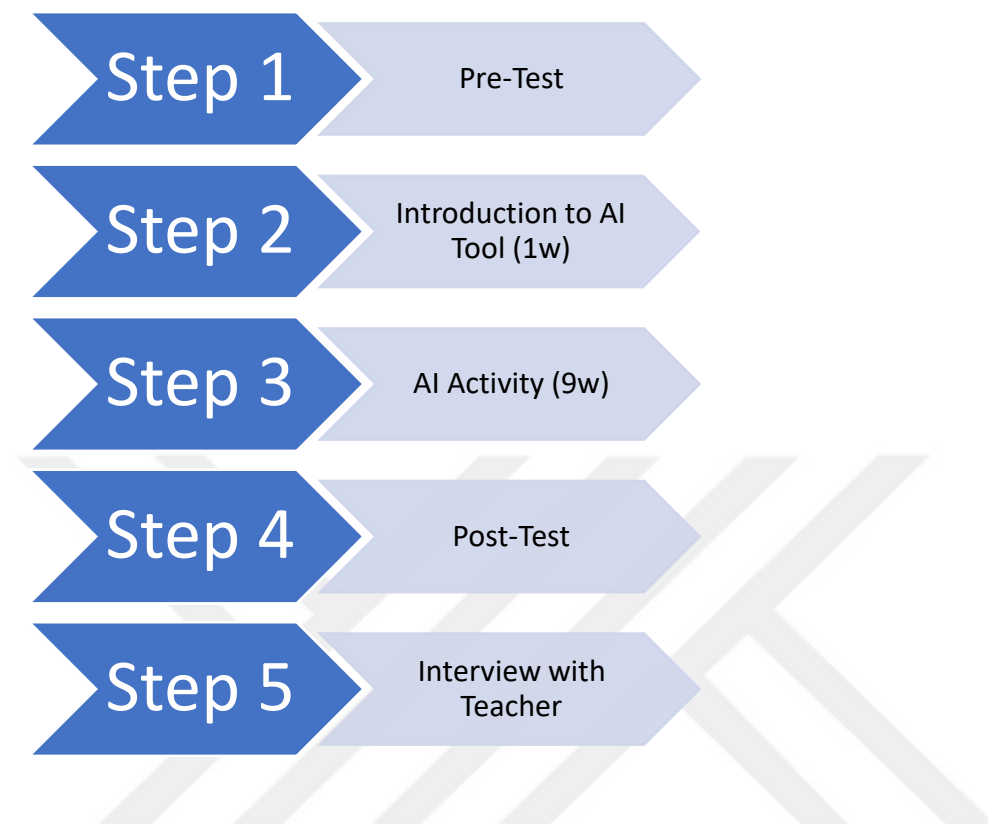
The author was given the role of an AI specialist by the school headmaster and the classroom teacher. He was assigned by the administration to conduct a part of the English lessons in cooperation with the classroom teacher and was responsible for only the AI activities. Outlines and general themes of the stories to be created were planned before the intervention with the classroom teacher to ensure the activities were appropriate for the students' age, proficiency level, cognitive development, and the learning objectives of the lessons. However, the classroom teacher mainly participated as an impartial observer and had no other influence on the implementation other than intervening in the lesson when she felt it was necessary. The researcher assumed the role of an assistant teacher, co-developed the lesson plan with the classroom teacher and directed the AI activity with minimal external intervention. Since the AI reading activity is almost entirely student-led, both the assistant and classroom teachers were present in the classroom in case the flow was somehow disrupted by a problem, such as technical difficulty or group arguments getting out of control.

Before the intervention, students had been grouped based on their English proficiency scores from highest to lowest by the school administration. The school administers English Assessment Tests twice per semester. The test that was administered before the intervention was taken as the pre-test. The post-test was administered one week after the completion of the intervention. Both pre- and post-test exams were co-prepared by the English teachers of the school and a testing company that the school administration employed. The exact nature of the exams is explained in the following sections.

An action flow consisting of the steps taken in this research is provided below.

Graph 3.1

Action Flow



3.2 Participants

Forty-eight 7th graders enrolled in a private school located in Eskişehir, Türkiye, were the participants of this study. There were five 7th-grade classes. Out of the five 7th-grade classes, two were chosen for this study as these two were taught by the same teacher, whereas the others had different teachers. Each class consisted of 24 students. One of the classes was randomly assigned as the control group, and the other as the experimental group. Students were between 12 and 13 years old ($M = 12.47$). All students had received approximately 6 hours of English lessons per week since the beginning of the 2nd grade. The majority of the students ($N = 35$) who participated in the study had enrolled in this private school when they began their K12 education. In the current semester, each class received 8 hours of English instruction per week. The English teachers reported that the students' proficiency levels were pre-intermediate to intermediate, which corresponds to B1 to B1+ levels (CEFR, 2001).

Both groups consisted of 24 students, the control group with 13 females and 11 males, and the experimental group with 14 females and 10 males.

3.3 Intervention

3.3.1 Control Group

The English instruction for the control group followed the school's curriculum. The curriculum focused on improving the four skills, reading, writing, listening, and speaking. For reading skills, students received traditional reading instruction, which included reading short passages, writing reviews or letters, reading informative reports, and filling in charts/graphs/forms to demonstrate their understanding (MEB, 2024). In addition to the national curriculum, students had extracurricular listening and watching materials.

Similar to the Ministry of National Education's (MONE) guidelines, the first hour of the lesson is usually allocated to listening and reading. At the start of the lesson, such student-engaging materials act as a primer to ready students to receive knowledge and active prior learning. Although flipped learning, videos, and online education platforms are incentivized, the exact nature of the material to be used as lesson starters is ultimately the teacher's prerogative. For the upcoming speaking activity in the second half, students are given time for preparation either individually or in groups. When the speaking activity is completed, the class usually ends with a writing activity and a feedback session.

3.3.2 Experimental Group

The experimental group received English instruction, as did the control group with the exception of allocating 2 hours per week for AI storytelling activity. The author was introduced to the classroom as a graduate student doing research on AI solutions to language education. During the 10-week period, the classroom teacher remained within the classroom as an impartial observer and supervised while the author (hereby referred to as the assistant teacher) conducted the AI study for two class hours. LLM module, image generator and extreme content filters were adjusted beforehand by the assistant teacher and the AI tool was ready for use in the class.

The first week was allocated to the introduction of the tool during which the participants were introduced to the technology that would be used (procedural content generation) for the reading activities. The classroom teacher was also given a tutorial on how to use the tool so that it would be accessible as an option to the control group should she desire to use it. Both the fine-tuned preset options and the resources necessary for full utilization

of the AI tool were made available to other English teachers working in that school as was done for the classroom teacher.

After the familiarization session, beginning with the second week and the following sessions, the class was divided into four groups with six students in each. The assistant teacher relied on the class teacher's assistance in grouping the students, as he was not familiar with the students and their level of achievement. The students were first asked to brainstorm and take notes on the three key elements, genre, characters, and events, they would use in their story. The genres students could choose from were horror, science fiction, comedy, or other genres suitable for lesson material. Once the groups finalized their selection of the genre, characters, and events, each group typed their keywords into the AI program as parameters. The assistant teacher reviewed the keywords, keeping in mind the participants' English proficiency level. This step served as a filter to prevent meaningless or unsuitable topics that the groups may have chosen. Otherwise, the story remained completely student-driven. Topic selection was crucial; thus, the assistant teacher had to personally ensure that the selected topic was compatible with the objectives of the learning outcomes. Members of the groups fervently discussed the details of the story, consequently, continuously interacting with each other. Such activities result in authentic task-based instruction, which theories of language acquisition, such as interaction theory, advocate. Furthermore, according to Long (1996), social interaction plays a crucial role in second language acquisition by enhancing meaning negotiation and, consequently, the comprehensibility of second language input. The group work aspect serves as an effective means to this end.

In the next stage, each group selected a spokesperson who reported the ideas their group brainstormed in the previous stage and the assistant teacher himself entered them into the AI tool as parameters. This spokesperson was rotated each time the turn came back to the group. Having groups come up to the smartboard to input their prompts which took time and caused disruptions, necessitated a more practical solution. Thus, a wireless keyboard and mouse were acquired to achieve a smoother tempo during the activity. From the second week onwards, the assistant teacher had a new responsibility: the movement of the wireless input devices between the groups and the distribution of them within the group. The decision of which student should be given the keyboard was based on who was the least engaged within the group. If not applicable, they were entrusted to the most

willing member of the group. Thus, the assistant teacher also monitored students' level of engagement during the activity.

As the tool kept the story going by reacting to students' inputs, each group took their turn to provide their response and directed the plot as they saw fit. Thus, through the inputs they provided, each group affected the plot and actions of the characters of the story. It was pivotal to keep the students involved and cognitively engaged throughout the procedure so that they would benefit from the positive effects of using natural language. Natural language is preferred as the authentic context provides them with input that is comprehensible, interesting, and linguistically challenging, leading to more effective language learning outcomes (Richards & Rodgers, 2014). Additionally, having students work within and between their designated groups fosters communicative development. As mentioned before, the students were divided into four groups with six students in each, and each group was given the same task of continuing the story by providing context-appropriate input to the AI tool. This required students to read the inputs from the other groups. Each member of the group shared their ideas with their groupmates, creating a discussion within the group. Also, each member of the group shared their ideas with their groupmates, creating a discussion within the group. Eventually, each group finalized its decision and entered it as the next input to the AI tool. Vygotsky (1978) emphasizes the role of the aforementioned social interactions and cultural context in learning. Having students work in groups provides them with the opportunity to work cooperatively to achieve the same objective, using authentic language supports them as they engage in the task, allowing them to possibly reach their zone of proximal development.

AI Dungeon, in its open scenario, emphasizes creativity and experimentation. Its user-friendly interface and error mitigation capabilities made it ideal for an implementation such as a storytelling activity to middle schoolers who might not possess the capability of AI input writing. Users can explore any kind of scenario as long as it complies with certain narrative filters established before the start of the game. Although this freedom encourages student-centered learning methods, role-playing in particular, the AI is not without its demerits. Occasionally, it produces irrelevant answers that deviate the plot from the learners' intended direction. In order to avoid such outcomes, users must write their prompts as grammatically accurate as possible and refrain from using phrases that could lead to misunderstandings. In situations where such a prompt was entered, the

“undo” function of the tool was used and it was explained to the students why their prompt caused the events to unfold undesirably.

AI Dungeon provides certain community features in which players can share their creations and followers can explore adventures crafted by others. However, in this study, preset stories from third parties were not used, as the intention was to allow students to interact naturally with the AI under the most natural conditions as possible.

3.4 Data Collection

To evaluate the effectiveness of the intervention, the tests administered by the school and several writing assignments given and graded by the English teacher were used as pre- and post-tests.

3.4.1 English Assessment Tests

The English Assessment Tests the school administers are developed by the English teachers in cooperation with a testing company. These tests, which are regarded as midterms and finals by the school, are designed as all-purpose exams testing three language skills, reading, writing and listening. Since listening is not a skill this study aimed to improve, the scores of the listening sections of the tests were not included in the data analysis. Writing part of the test, on the other hand, was included as the relationship between reading and writing is connected and mutual. Both language skills significantly enhance the other. One cannot be achieved without the existence of the other since both share fundamental key elements like vocabulary, grammar, and syntax. Proficiency in one implies the presence of the other, thus the scores of the writing parts were included in the analysis of the data.

The scores of the reading and writing parts administered before and after the implementation of AI-based practice were taken as pretest and posttest scores, respectively.

The test has five parts, four of which pertain to reading and one to writing as illustrated in Table 3.1. Part 1 is a vocabulary test in which students are asked to complete ten gaps within a text with the appropriate verb. In Part 2, students are required to complete sentences and order them in a text. Part 3 tests their ability to arrange given sentences in a cohesive order within a text. Lastly, Part 4 is reading comprehension with multiple-choice questions. Part 5 consists of a writing task, to be completed at home, in which

students are asked to write a brief paragraph about a theme they have discussed in one of the English lessons. Students are expected to write a paragraph with a minimum of 150 words.

The test has a total of 36 points. Each part of the test is scored differently, as seen in Table 3.1. The first part is 8 points, Parts 2 and 3 are 4 points, Part 4 is 5 points, and Part 5 is 15 points.

Table 3.1

Parts of the English Assessment Tests

	Content	Score
Part 1	Fill-in-the-Blanks	8
Part 2	Complete the Gaps	4
Part 3	Sentence Ordering	4
Part 4	Comprehension Questions	5
Part 5	Writing	15
Total		36

3.4.2 Semi-structured Interviews

Semi-structured interviews with the classroom teacher and the headmaster were conducted to determine students' perceptions of the intervention; whether they found the intervention beneficial, engaging and entertaining. The questions asked in the semi-structured interviews are given in the Appendix. The classroom teacher was also asked to comment on students' opinions about AI and whether there were any changes in their self-efficacy regarding their ability to comprehend general knowledge. This was based on the assumption that students' confidence in reading fluently and effectively may be related to this ability (Lee & Jonson-Reid, 2016).

The interviews were done individually in Turkish as people often feel better when they speak in their native language which result in an increased trust and connection between the interviewer and interviewee. Welch and Piekkari (2006) state that language is a way to build understanding and connection. Therefore, it is vital for that medium to be clear of any hindrance or obstruction to function. In order to avoid any misunderstandings and to gather richer and more detailed data, Turkish was chosen to be the language used in the interviews. Interview questions and their translations are provided in the Appendix.

The answers the classroom teacher and the headmaster gave to the interview questions were recorded by hand.

3.5 Data Analysis

3.5.1 Pre-test and Post-test

Scores of the English Assessment tests before and after the intervention were used to assess the effectiveness of the intervention, students' level of general reading comprehension, specifically linguistic and orthographic knowledge, and their language aptitude. The variables, which are pre- and post-test results of each student, were entered into Python for statistical analysis. As there was a difference in the pre-test scores of the experimental and control groups, ANCOVA was conducted with pre-test scores as a covariate to determine the differences in the performance of the students in the post-test. To determine the effect of the intervention, within-groups changes from pre- to post-tests, paired-samples t-tests were employed.

3.5.2 Interviews

The teacher's and the headmaster's comments were analyzed thematically to determine whether self-efficacy correlates with the results gathered. The data were analyzed under two themes; effects of the intervention on student motivation, and impressions of AI and AI-based reading activity. The thematic analysis approach was used to systematically interpret the data to identify recurring patterns. Another reason advocating the usage of thematic analysis is that this method is ideal for examining individual perceptions, values, and, in this case, their experiences with AI. This interview was prepared with the belief that it would be a valuable asset in gathering in-depth qualitative data. The classroom teacher's impressions on the implementation, her observations regarding student actions and relaying students' remarks to the researcher would help address as many key themes as possible. In addition, it is important to receive feedback from the actual users, students in this context, analyze their needs, issues and motivations. Combining the assistant teacher's findings, the headmaster's impressions and the classroom teacher's insights enables the triangulation of the data (Mashuri *et al.*, 2022).

4 RESULTS

In this section, the results of the data analysis are reported for each research question separately.

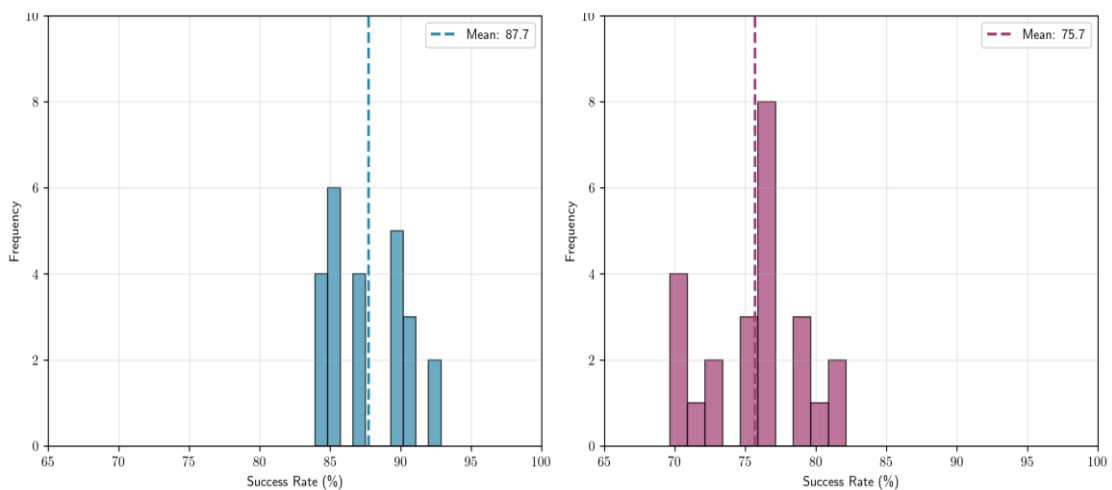
4.1 Q1: Do AI-based interactive storytelling tools improve general reading comprehension?

Following the 10-week AI storytelling intervention, students in both the experimental and control groups took the same post-test. This test was conducted a week after the conclusion of the experiment, and every student's value was included without any missing values. The post-test, although not the same test, was comparable in both content and difficulty to the pre-test. Differentiated tests that are similar to each other difficulty-wise were implemented as a reliability measure (Beauducel, 2017).

All analyses were conducted on percentage-based success scores to account for differences in the scoring scales of the pre- and post-tests. While the pre-test was scored out of 46, the post-test was scored out of 36, with the parts of the test being the same but with different scores. First, to determine the distribution of the pre-test, normality tests were conducted. The results indicated that both the control group ($W = 0.921, p = .062$) and the experimental group ($W = 0.924, p = .071$) had approximately normally distributed pre-test scores as seen in Figure 4.1. Homogeneity of variances was also supported with Levene's test ($F(1, 46) = 0.50, p = .484$).

Figure 4.1

Distribution of the Pretest Scores in Percentages (Left: Experimental Group, Right: Control Group)



As can be seen in Figure 4.2, there is a difference in the pre-test means of the experimental and control groups. A comparison of the pre-test scores showed that there is a significant difference between the two groups, with the experimental group scoring higher ($M = 87.72$, $SD = 2.80$) than the control group ($M = 75.67$, $SD = 3.72$), $t(46) = 12.68$, $p < .001$, Cohen's $d = 3.66$.

Due to the difference in the pre-test scores, ANCOVA was conducted with pre-test scores as a covariate to determine the differences in the performance of the students in the post-test. The result of the analysis showed that there was no significant group effect after controlling for pre-test performance, $F(1,45) = 2.65$, $p = .111$, partial $\eta^2 = .056$. The covariate (pre-test scores) was also not a significant predictor of post-test performance, $F(1, 45) = 2.61$, $p = .113$.

To further explore within-group changes, paired-sample t-tests were conducted. As seen in Table 4.1, for the experimental group, the mean scores of the post-test ($M = 88.22$, $SD = 6.60$) did not differ significantly from that of the pre-test ($M = 87.72$, $SD = 2.80$), $t(23) = 0.38$, $p = .710$, Cohen's $d = 0.08$. However, comparison of the control group's mean scores for the pre-test ($M = 75.67$, $SD = 3.72$) and the post-test ($M = 70.20$, $SD = 12.22$) indicated that their performance declined significantly, $t(23) = -2.25$, $p = .034$, Cohen's $d = -0.46$.

Table 4.1

Mean Scores and Standard Deviations for the Control and Experimental Groups

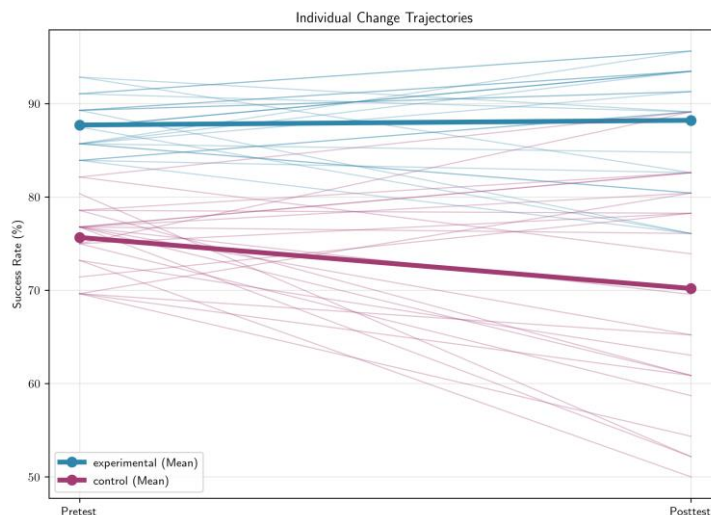
Group	N	Pre-test Score (%)		Post-test Score (%)	
		Mean	SD	Mean	SD
Control	24	75.67	3.72	70.20	12.22
Experimental	24	87.72	2.80	88.22	6.60

Based on the experimental group's mean scores on the pre- and post-tests, the AI-based intervention appears not to have affected the experimental group's performance. This outcome is indicated by the low " p " value that is below the .05 threshold. However, the significant decline in the control group's performance suggests that the AI intervention may have served a stabilizing role. In Figure 4.2, the change from the pre- to post-test

scores of each student is plotted. There is more variation in the performances of the students in the control group compared to those of the experimental group.

Figure 4.2

Pre-test to Post-test Score Changes for Each Student in the Experimental and Control Groups



Cohen's d for pre- to post-test change within each group shows how strong the effect of time or intervention was. When the Cohen's d values for the two groups are compared, the experimental group showed a negligible effect (Cohen's $d = 0.08$) while the control group showed a medium effect (Cohen's $d = -0.46$). Thus, based on the results of the analyses, it can be concluded that although the intervention did not enhance performance significantly, it prevented a decline in the experimental group's performance and had a positive effect on maintaining academic performance.

4.2 Q2: What are the perceptions of the stakeholders (i.e., the English teacher, headmaster of the school and students in the experimental group) regarding AI-based reading activities?

Following the completion of the experimental stage of the research, the classroom teacher and headmaster were asked to share their opinions on the AI tool, its generative capabilities, method of implementation and students' impressions of it.

The answers the classroom teacher and the headmaster gave are discussed under two main themes: their perceptions of the effect of AI-based activities on student motivation and their general impressions of AI and AI-based reading activities.

4.2.1 Effects on Student Motivation

Although the usage of technology in the form of conversational chatbots and web applications to learn a language is not a recent trend, incorporating AI in lesson plans is a relatively new prospect. Willing to brave that risk, the headmaster was, in fact, happy to accept a researcher working on an AI-based English practice. Choosing one of the most experienced English teachers in the school, one of her classes was designated as “project class” and had its class hours reallocated to accommodate this research. The headmaster shared his thoughts on AI activity and his reasoning behind his decision by stating:

“In this day and age, technology has become a must for education. It’s becoming increasingly difficult to maintain student attention using traditional methods. As such, we’d like to support initiatives and, hopefully, see them work out well. So, I entrusted one of our most esteemed teachers to help with your research.”

As the classroom teacher was the one who saw the practice firsthand, her thoughts and opinions about the efficacy of the intervention were very valuable. While she considered the practice a success overall, she did not always take this stance. According to her, students’ viewpoints on the AI practice changed over time. Initially, when she mentioned that an assistant teacher was coming to conduct a reading activity using AI, students saw this as an excuse to skip or not listen to English lessons. If there was a negative attitude of the students towards the assistant teacher and the activity could have resulted in the assistant teacher not being able to capture the students’ attention, thus hindering the research due to a lack of participation. In the words of the teacher:

“I wasn’t sure whether you could handle the classroom. They could be rowdy or wayward sometimes, given that most of them are already proficient in English.”

The classroom teacher made it clear that the introductory lesson was the make-or-break of this practice. The impressions it would leave on the students would determine their attitudes towards the assistant teacher, the intervention, and AI. In fact, an underwhelming introduction to AI-based reading activity could have jeopardized their perspectives on English as a whole. Fortunately, events did not develop that way. The classroom teacher noted that a warm greeting and a sincere dialogue exchanged with the students helped break the ice. She stated that the students were very comfortable and motivated to partake in the practice. She added that the actual issue was not students lacking motivation but rather being overzealous. They would become engrossed in their group discussions about what steps they should take next or what picture they should have AI draw. The teacher

noted that initially, students were hesitant because they were not sure about what they were expected to do when their turn came. However, once they familiarized themselves with the practice, discussions became heated, the wireless keyboard got confiscated by the headstrong students and the classroom got engulfed in loud noises. They would even object to other groups' inputs, insisting that they wanted the story to unfold differently, fully aware that they would have to accept the final outcome. All these arguments and disputes between and within the groups might sound counterintuitive to the objective of this group work activity, but the teacher actually welcomed them. She shared that:

"I actually prefer them getting genuinely involved in the lesson rather than having them sulk the entire time."

The students predominantly preferred the AI reading activities over the traditional ones. According to the teacher, students would look forward to that day of the week, and even plan what they would write ahead of time. Not only that, they would ask the assistant teacher to promise that they would be the first to start the activity the following week. Before the intervention began, the teacher predicted that students would like using AI but the level of enthusiasm students displayed was much higher than she had expected. Quoting students, the teacher mentioned that some would go as far as to say:

"I want to have this AI activity instead of regular English lessons for the rest of the semester. – Student 11"

Certain student comments drew comparisons between the AI reading intervention and other activities they did in their regular English lessons, such as puzzles, roleplay or web-based activities (MEB, 2024). Based on the teacher's inquiries, students preferred reading skill instruction with AI activities because it was gamified. This may be one of the reasons that could explain the high level of engagement (Seixas *et al*, 2016). The students were happy to have been given a chance to engage with technology as well as with their classmates, as one of the students' comments indicates.

"It is fun to control a character in a story our group writes even though other groups mess things up. The story gets crazy sometimes and all we can do at that point is laugh." – Student 5

Students labeled as "headstrong" above, were intrinsically motivated well enough that the flow of the discussions was rarely hindered, and each group had at least one such student. Sometimes, some groups would get so passionate about their inputs that discussions would exceed the time limit, resulting in the disruption of the flow and other groups

getting impatient. They would be very eager to engage with the tool to the point that they would not deem it necessary to summarize what happened in the previous week and get straight to writing. Enthusiasm was welcomed but coherence and cohesion suffered occasionally, which disappointed students who pre-tailored their prompts. Occasionally, when their L2 proficiency failed them, they would resort to using L1 during their discussions. This behavior was generally overlooked since this conduct was observed when discussions would get particularly heated. Exceeding the time limit, despite the warnings of the assistant teacher, was unavoidable because the group members could not reach a decision they all agreed upon. It is worth noting that delays in making a decision were not due to insufficient L2 proficiency, rather, group members could not agree on how they should continue the story or what the next image they should have AI create. A student shared the following statement with the teacher:

“We would eagerly wait for our turn to write, but other groups would discuss for what seemed like hours and exceed their time limit. After a while, it would get boring just waiting to do something.”

– Student 17

Having a classroom full of enthusiastic students who eagerly participate in discussions and interact with the AI may not be a realistic prospect. Even if every member of the class was excited to have a narrative experience, as most of the students in this class were, few students were anxious to speak. Both the classroom and the assistant teacher were aware of such students. The teacher felt that these students needed to be eased into joining group discussions and interacting with the AI so that they could produce more L2 input and output. Subsequently, the meaningful communication they establish with both their group and the AI would reflect on them positively (Fryer *et al.*, 2019). The assistant teacher would purposefully roam around the classroom and join in the discussions of the groups waiting for their turn so that he could ask what should be inputted into AI next. He would ask all the students’ opinions in that group, especially those who were either silently listening or not paying attention. Another method the assistant teacher utilized was purposefully selecting passive students as the keyboard keepers. As the one writing the inputs, such students had to listen to the discussions even if they were not actively speaking, so that they would be exposed to their peers’ L2 output. These actions on the part of the assistant teacher aimed to give those students extrinsic motivation in the hopes of getting them engaged in group work and/or overcoming their L2 anxiety (Liu, 2006). The teacher recognized the assistant teacher’s effort to include everybody and commented

that it was remarkable. In addition, she approved the fact that the assistant teacher would not insist on direct participation. It is not always easy for a teacher to be aware of what a student is going through. It is more difficult for a non-permanent teacher, such as the assistant teacher who was a guest in this class, to know each student and what s/he may be going through. The teacher approved the methods employed by the assistant teacher to include all the students in the activities. She shared her observation of previously inactive students participating more in her classes.

According to the headmaster, some students reported that the AI reading activities were entertaining, and they would have liked to have more AI in their other reading activities. One student who was particularly fond of reading mentioned that AI itself was the reason why she had fun during the lessons, not because of the fact that it was a reading activity.

4.2.2 Impressions on AI and AI-based Reading Activity

As information technology became increasingly mainstream and practical, a period that has transformed humanity's way of living, known as the digital age (or information age), has emerged in the 21st century. This period not only shaped the cognition, personality, and cultural values of Generation Y and Z but also exacerbated a paradigm shift from traditional teaching methods to virtual learning environments, redefining the roles of teachers and this new generation of learners in classrooms (Sarkar, 2012). Employing ICT and derivative tools allowed instructors to enhance learning and standards while augmenting students' ability to learn (Sharma *et al.*, 2018). To that end, the school's headmaster and classroom teacher were asked to evaluate the AI-based reading activity that we have designed while considering contemporary education and ICT developments. The following sections will expand upon this topic.

Both the headmaster and the teacher agreed that the students favored AI activity over traditional methods of teaching. The majority of the students in the experimental group expressed their positive views about the intervention to both the headmaster and the classroom teacher. To them, participating in an experiment involving cutting-edge and popular technology, such as artificial intelligence, was much more interesting and enjoyable than regular classes. According to the headmaster, while being a member of a group during an activity is more engaging for students than solo work, regardless of the material, AI itself was also a tool with which none of them had any experience. AI allowed the students to experiment with the limits they can push to create silly and surreal

narratives which made the activity so engaging for them. This particular student who shared these sentiments with the headmaster, regularly read recreational literature that could be considered a level above a 7th grader before the experiment. As such, it could be assumed that students who are predisposed to reading could have potentially benefitted from and enjoyed using AI to create their own narratives, even if said narrative would get regularly disrupted by other groups' inputs, shifting it to a different direction.

The teacher raised the issue of how AI reacts to various syntax and grammar errors in student inputs. Its ability to comprehend structurally flawed sentences is deemed more than adequate by the teacher. Recent LLM modules have advanced enough so that minor spelling or grammar mistakes students made did not disrupt AI's narration. As such, students had the chance to correct their grammar mistakes through implicit instruction as AI wouldn't outright reject the erroneous inputs. Instead of fixing the grammar in the input given, AI would act as if there were no mistakes at all and provide grammatically accurate responses, preserving the flow of the story despite the erroneous prompts given by the students. Although a mechanical flaw in the input would not result in a failed interpretation by the AI, a contextually incompatible input, however, could have a disruptive effect on the narrative. AI would dismiss the previous flow or narration and continue to tailor the story according to the context given by the new input. For example, the plot of the story written by the classroom is unfolding through a character called "Mike". To progress the story, the groups craft their input using a third-person point of view, meaning that the story is told by an omnipresent narrator that no character acknowledges. This means that student groups would use sentences such as: "Mike opens the wooden door and the creak it makes wakes up his brother Archie.". If a student group inputs a sentence like "I talk to my awake brother.", AI would sometimes give responses that shift the point of view to an undefined character's first-person view, disregarding the previous third-person narration. The response given to that sentence could be: "As I talked to my brother Archie, Mike's phone started to ring.". The supposed main character, in this case Mike, is then relegated to a secondary or a tertiary character, and a new, previously not mentioned character, becomes the main character due to a misphrase. The teacher noted that contextual mistakes like these might cause disruptions in the cohesion of the story.

The headmaster's, teacher's, and, according to the teacher, the students' impressions of AI-based reading activities were generally positive. When asked their thoughts on AI, both the headmaster and the teacher said that it is only a matter of time before AI becomes a part of our daily lives, adding that it might already have. Both esteemed educators have seen English instruction evolve through numerous decades, shifting from a focus on mechanics to an all-around integrated approach to language teaching. Change of emphasis in lesson planning paved the way for the introduction of innovative content delivery methods and facilitative tools to language education. They both recounted the gradual inclusion of new tools and methods when they first met with computers:

"I remember how we got used to cassette and CD players when computers got popular. It seemed like such a novelty at first. Now, we see them in every classroom. AI too will become prevalent like that sooner or later." – Classroom Teacher

Considering the experience they have gathered throughout years of teaching, teachers have come to expect and welcome paradigm shifts and social changes transforming education for the better. This is not only a respectable mindset to have but a critical attribute for a teacher who attends to shifting landscapes and social values. However, this belief does not always translate into observable behavior. Although both the headmaster and teacher are aware that constantly evolving technology is a tremendous facilitator in language teaching, they were unwilling to be the pioneers introducing innovative implementations, such as AI, to their classrooms. When asked whether she would use AI in her lessons from now on, the teacher said:

"There is no way I can figure out how it works, let alone use it in a lesson. I can barely deal with projecting the coursebook onto the screen. I do support its development though."

Why would such a well-esteemed teacher, who is experienced in her field so much so that she holds the title of the head teacher of the English teaching committee of the school, think that she would not be able to figure out how an AI application works? The answer may lie within the accumulated experience itself. At a glance, it might look like a benefit to be experienced. However, it could be the reason why teachers lack the enthusiasm to adopt new and unfamiliar implementations. The belief in one's ability to perform the action necessary to achieve an outcome is a stronger driver compared to simply knowing what action leads to which reaction (Bandura, 1997). As teachers get more experienced, they are more unwilling to innovate and more likely to believe that they are externally

limited in bringing changes to their teaching methods and innovations to student learning (Ghaith & Yaghi, 1997).



5 DISCUSSION AND CONCLUSION

This study investigated the effect of AI-based interactive storytelling tools on general reading comprehension. The experimental group received a 10-week intervention in which they used AI to generate a story with the prompts they provided. The students were fully engaged in the AI reading activities, using English with the members of the group to brainstorm and discuss the details of the story they were creating. The comparison between the pre- ($M = 87.72$) and post-test scores ($M = 88.22$) for the experimental group showed that there was no significant difference in the performance of the students, suggesting that AI-based reading activities did not have a positive effect on the students' performance. The control group's performance, however, decreased significantly (from $M = 75.67$ to $M = 70.20$). Although the students in the experimental group performed similarly in both the pre- and the post-test, the significant decline in the control group's performance suggests that the intervention helped students sustain their level of performance and that the AI-based intervention had a positive effect.

Nonetheless, the lack of an increase from pre-test to post-test scores for the experimental group needs to be explored. There may be several reasons. First, the pretest and posttest may not have been the right tools to assess students' reading skills. In this study, the English Assessment tests administered by the school before and after the intervention served as the pre-test and post-test, respectively. These tests consisted of five parts: three with short-answer questions, one with multiple-choice questions, and one writing task. The intervention was not designed to prepare the students for such tests. Rather, the students created their own storyline with events and characters they chose, not necessarily what they would be examined on. Additionally, as they were school-administered tests, the pretest and posttest were not the same tests. Thus, the difficulty or the cognitive load of the tests may not have been perceived to be similar by the students. The fact that the significant decrease in the control group's performance can be considered as evidence.

Second, although the students were immersed in the activities and used L2 throughout most of the activities, it may be the case that the intervention was not appropriate for improving reading skills. Third, the duration of the intervention may not have been sufficient. Students received a total of 20 hours; this may not have been enough to improve their reading skills and they may have needed more activities. Also, they may

have needed more practice with AI-based activities as they were not familiar with such implementations.

Furthermore, the experimental group's lack of difference in reading comprehension skills may be attributed to insufficient coverage of diversely organized input and the level of texts that lack in complexity (Payne *et al.*, 2006; Heyns *et al.*, 2022). Throughout the study, students were responsible for inputting the directive prompts the AI needed to progress the story. Each prompt differed from the other either in terms of accuracy or relevance to the overall story. But they were still accepted by the AI and accommodated to the story. Although AI generated a text that is tailored to the level of students based on the parameters set by the researcher, it has been observed that the students were not careful about the grammatical accuracy of their input and disregarded the implicit feedback given by the AI response. Grammatical accuracy was not the main objective of this specific group practice. Thus, it should be noted that while there are benefits to implementing contemporary technology in language education (Morgan, 2012), the benefits of traditional teaching methods should not be neglected. The contribution of the AI-based reading activities manifested itself in the form of supportive supplementation. It did not necessarily yield a significant increase in reading comprehension. However, it might have helped the experimental group in preventing the potential decline that the control group showed. Nevertheless, factors that might have played a role in the insignificant increase in the experimental group's reading comprehension must be considered in future studies. One factor that needs exploring is students' exposure to AI learning tools. In this study, students had relatively limited exposure to the AI learning tool. This could have posed some limitations and demotivated some of the students due to its dissimilarity to traditional methods (Lin *et al.*, 2017). Similarly, Mageira *et al.* (2022) did not achieve a significant language performance increase, but the findings provided promising results that indicated AI chatbots could be effective in teaching English and its culture. Difficulties experienced in Mageira's study were similar to those encountered in this study, as there would be problems with the flow of the story if students provided contextually incompatible prompts. Sudden changes students made to the narrative perspective could be given as an example of such prompts. Also, the time allocated to such activities may be increased as the period of the intervention might be a reason why it did not translate into improved reading scores in this study.

Even if the AI activity did not reinforce the reading skills of the participants, it might have had a positive effect on other skills, such as creativity, brainstorming, critical thinking, and negotiation with the members of the group and classmates. The design of the activity had students grouped with and, unintentionally, against each other, resulting in constant interaction with each other. Effects of the inclusion of social elements in AI-based activities should be further investigated to explore AI activities' possible effects on not only language skills but also other skills, i.e., communicative and social skills.

Analysis of the interviews regarding AI integration into the lessons shed light on the perceptions of the stakeholders: the headmaster, the classroom teacher, and the students. It was pivotal to examine their impressions on it as studies on cooperative groupwork, reading activity using AI, especially for 7th graders, are limited. There are studies that have demonstrated both positive and negative effects of collaborative and individual-focused implementation of computer-assisted learning and AI in various ways (Chun, 1994; Nurzaman, 2020; Ammanabrolu *et al.*, 2021; Mageira, 2022). Yet, these studies either did not include the contemporary AI technology or they focused on either young learners or high schoolers. Therefore, this study focused on middle school students to discover the potential effects of AI, specifically on reading. Generally, a clear consensus was reached on the engaging and motivating aspect of the activity. Although there was criticism regarding the implementation of the application, the tool itself was mostly praised by all parties, highlighting the development of new methods of interaction with the AI tool (Teo *et al.*, 2016). Putting groups in charge of how the story develops allowed students to determine their own goals and the ways of achieving them, allowing a much more student-oriented instruction and substantiated learning experience (Kolling & Shumway-Pitt, 2019). The inclusion of technology in otherwise regular language instruction was met with an increase in motivation and engagement by the students, confirming the inferences made by Nassim (2018).

The impressions of all the stakeholders were positive. The thematic analysis of the interviews revealed that the study succeeded in increasing student motivation and engagement and was generally favored over traditional instruction methods. Even though the groups were difficult to manage when the discussions heated up, the students said that they enjoyed the experience of creating a story together and competing with the other groups. There were a few students who criticized the tool, stating that they were

disillusioned at first because they expected a more human-like narrator who would appear on the screen and speak. Having students read their inputs and the responses by the AI was an intentional part of the implementation, as this study focused on reading skills, but having a synthetic voice read the AI-generated responses out loud might have been more motivating for the students. Although such technology is available in other AI platforms, AI Dungeon, the application used in this study, lacks voice narration. Rapid progression within the field could yield a platform that could satisfy the need for a narrator voice, preferably one that could be selected from a variety of different voices, depending on student preferences.

Thematic analysis of the interviews also revealed that the stakeholders are open to the inclusion of AI in language classrooms and not as pessimistic as determined by Xue & Wang (2022). Presumably, rapid improvement of AI technology might have played a part in a shift in general opinion of it as it quickly became a part of many teachers' life thanks to its ability to increase efficiency by quickly organizing a wealth of information a teacher might need for material preparation, general literacy, etc. (Chen *et al.*, 2020). 21st-century solutions such as AI are not only tools for teachers to rely on but catalysts that could trigger a paradigm shift from contemporary methods to a whole new understanding of instruction (Sarkar, 2012). AI could, potentially, create a foundation that education will be built upon in the future. Yet, teachers might not be ready for such a change to take place just yet. Even if they agree on the potential benefits of AI, especially on student motivation and engagement, some might be reluctant to adopt it due to factors such as self-efficacy. The class teacher, who had more than 10 years of experience, felt that she would not be able to implement AI effectively in her lesson plans as she feels more comfortable using her well-tested methods over brand-new methods that she had not seen before. Ghaith & Yaghi (1997) explain this phenomenon as experienced teachers inadvertently believing that they are externally limited in bringing drastic changes to their routines. Nevertheless, the positive perceptions of AI implementation were shared by the school's headmaster and the classroom teacher. Both agreed that the intervention was very effective in motivating students to participate in the English lessons. Although they claimed that they lacked the necessary knowledge to implement AI-based tools, they expressed their support for its development and eventual implementation not only in English curricula but also potentially in education in general. But when the teacher is asked if she would consider using it in her lessons, she stated that she would not be able

to use it effectively because of her age. This answer raises the question of how many teachers who support the use of AI as a tool of instruction are reluctant to develop lesson plans using it due to a lack of self-esteem or know-how. Such advances in technology highlight the necessity of in-service teacher training and/or a revision in the material development curriculum for pre-service teachers. No matter how substantial an innovation is, unless educators are incentivized and trained to be aware of and use it, teachers cannot actualize it through educational practice (Lieberman & Pointer Mace, 2008).



REFERENCES

- Ammanabrolu, P., Cheung, W., Broniec, W., & Riedl, M. O. (2021). Automated Storytelling via Causal, Commonsense Plot Ordering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(7), 5859-5867. <https://doi.org/10.1609/aaai.v35i7.16733>
- Apel, K. (2011). What is orthographic knowledge? *Language, Speech, and Hearing Services in Schools*, 42(4), 592–603.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 83, 191- 215.
- Beauducel, A. (2017). Parallel-Forms Reliability. https://doi.org/10.1007/978-3-319-28099-8_1337-1.
- Belda-Medina, J.; Calvo-Ferrer, J.R. (2022). Using Chatbots as AI Conversational Partners in Language Learning. *Applied Sciences*, 12(17), 8427.
- Burnett, R. (1975). Accurate estimation of standard deviations for quantitative methods used in clinical chemistry. *Clinical chemistry*, 21 13, 1935-8. <https://doi.org/10.1093/CLINCHEM/21.13.1935>.
- Burtenshaw, B. (2020). AI Stories: An Interactive Narrative System for Children. *arXiv preprint arXiv:2011.04242*.
- Cervetti, Gina & Hiebert, Elfrieda. (2015). The Sixth Pillar of Reading Instruction. *The Reading Teacher*. 68. 548-551. 10.1002/trtr.1343.
- Chen, X., Zou, D., Xie, H. R., & Su, F. (2021). Twenty-five years of computer-assisted language learning: A Topic modeling analysis. *Language Learning & Technology*, 25(3).
- Cherry-Cruz, T. (2001). Tell Me a Story: Enhancing Literacy Through the Techniques of Storytelling. *The ASHA Leader*, 6(23), 4–18. <https://doi.org/10.1044/leader.ftr1.06232001.4>
- Chun, D. (1994). Using computer networking to facilitate the acquisition of interactive competence. *System*, 22, 17-31. [https://doi.org/10.1016/0346-251X\(94\)90037-X](https://doi.org/10.1016/0346-251X(94)90037-X).
- Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>.
- Council of Europe (2001a) *Common European Framework of Reference for Languages: Learning, teaching, assessment*, Cambridge: Cambridge University Press.
- Cromley J. G., Snyder-Hogan L. E., Luciw-Dubas U. A. (2010). Reading comprehension of scientific text: A domain-specific test of the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, 102, 687-700.
- Friedberg, R. D. (1994). Storytelling and cognitive therapy with children. *Journal of Cognitive Psychotherapy*, 8(3), 209–217.
- Fletcher, J. (2006). Measuring Reading Comprehension. *Scientific Studies of Reading*, 10, 323 - 330. https://doi.org/10.1207/s1532799xssr1003_7.

- Fletcher-Flinn, C. M., & Thompson, G. B. (2004). A mechanism of implicit lexicalized phonological recoding used concurrently with underdeveloped explicit letter-sound skills in both precocious and normal reading development. *Cognition*, 90(3), pp. 303–335.
- Fryer, L.K., Nakao, K. and Thompson, A. (2019), “Chatbot learning partners: connecting learning experiences, interest and competence”, *Computers in Human Behavior*, Vol. 93, pp. 279-289
- Gallagher, K. (2009). *Readicide: How Schools Are Killing Reading and What You Can Do About It* (1st ed.). *Routledge*. <https://doi.org/10.4324/9781032682198>
- Ghaith, G., & Yaghi, H. (1997). Relationships among experience, teacher efficacy, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education*, 13(4), 456.
- Graham, S., & Perin, D. (2007). Writing next: Effective strategies to improve writing of adolescents in middle and high school. *A report to Carnegie Corporation of New York: Alliance for Excellent Education*
- Güneş, F., & Tanju D., (2020). Yetişkin Eğitimi ve Hayat Boyu Öğrenme. *Pegem A. Yayınları*.
- Gunser, V.E., Gottschling, S., Brucker, B., Richter, S., Gerjets, P. (2021). Can Users Distinguish Narrative Texts Written by an Artificial Intelligence Writing Tool from Purely Human Text?. In: Stephanidis, C., Antona, M., Ntoa, S. (eds) *HCI International 2021* (pp. 520–27). Springer, Cham. https://doi.org/10.1007/978-3-030-78635-9_67
- Heyns, N., & Van Zaanen, M. (2022). Detecting Multiple Transitions in Literary Texts., 3375-3381.
- Payne, S., & Reader, W. (2006). Constructing structure maps of multiple on-line texts. *Int. J. Hum. Comput. Stud.*, 64, 461-474. <https://doi.org/10.1016/j.ijhcs.2005.09.003>.
- Kendeou, P., McMaster, K., & Christ, T. (2016). Reading Comprehension. *Policy Insights from the Behavioral and Brain Sciences*, 3, 62 - 69. <https://doi.org/10.1177/2372732215624707>.
- Kolling, D., & Shumway-Pitt, K. (2019). Lesson Planning 101. *Creative Commons Attribution ShareAlike*.
- Kozminsky, E., & Kozminsky, L. (2001). How do general knowledge and reading strategies ability relate to reading comprehension of high school students at different educational levels. *Journal of Research in Reading*, 24, 187-204.
- L. Chen, P. Chen and Z. Lin, (2020). "Artificial Intelligence in Education: A Review," in *IEEE Access*, 8, pp. 75264-75278, 2020, doi: 10.1109/ACCESS.2020.2988510.
- Landay, James & Oliver, Nuria & Song, Junehwa. (2019). Conversational User Interfaces and Interactions. *IEEE Pervasive Computing*. 18. 8-9. [10.1109/MPRV.2019.2921176](https://doi.org/10.1109/MPRV.2019.2921176).
- Lee, Y., & Jonson-Reid, M. (2016). The Role of Self-Efficacy in Reading Achievement of Young Children in Urban Schools. *Child and Adolescent Social Work Journal*, 33, 79-89. <https://doi.org/10.1007/S10560-015-0404-6>.

- Lieberman, A., & Pointer Mace, D. H. (2008). Teacher learning: the key to educational reform. *Journal of Teacher Education*, 59.
- Lin, C. C., Liu, G. Z., & Wang, T. I. (2017). Development and usability test of an e-learning tool for engineering graduates to develop academic writing in English: A Case study. *Educational Technology & Society*, 20(4), 148-161.
- Liu, M. (2006). Anxiety In Chinese EFL Students at Different Proficiency Levels. *System*, 20(3), pp. 301-16. <https://doi.org/10.1016/j.system.2006.04.004>.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413-468). New York: Academic Press.
- Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI Chatbots for Content and Language Integrated Learning. *Applied Sciences*. <https://doi.org/10.3390/app12073239>.
- Mashuri, S., Rasak, M.S., Alhabsyi, F., & Syam, H. (2022). Semi-structured Interview: A Methodological Reflection on the Development of a Qualitative Research Instrument in Educational Studies.
- Morgan, H. (2012). Using Digital Story Projects to Help Students Improve in Reading and Writing. *Reading Improvement*, 51(1).
- Nassim, S. (2018). Digital Storytelling: An Active Learning Tool for Improving Students' Language Skills. *PUPIL: International Journal of Teaching, Education and Learning*.
- Nurzaman, I., Gandana, G., & Wahidah, A. S. (2020). Developing interactive storytelling model to facilitate young learners' speaking skills. In *International Conference on Elementary Education* Vol. 2, No. 1, pp. 64-69.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18, 22-37.
- Piaget, J. (1926). The Language and Thought of the Child. *New York: The Humanities Press INC*.
- Richards, J. C., & Rodgers, T. H. (2014). Approaches and methods in language teaching. *The USA: Cambridge University Press*.
- Royer, J., & Cunningham, D. (1981). On the theory and measurement of reading comprehension. *Contemporary Educational Psychology*, 6, 187-216.
- Santi, E. (2023). Reading and narrative fiction: Understanding ourselves and others. *BERA*. <https://www.bera.ac.uk/blog/reading-and-narrative-fiction-understanding-ourselves-and-others>
- Sarkar, S. (2012). The Role of Information and Communication Technology (ICT) in Higher Education for the 21st Century. *The Science Probe*, 30-41.

- Seixas, L.D., Gomes, A.S., & Filho, I.J. (2016). Effectiveness of gamification in the engagement of students. *Comput. Hum. Behav.*, 58, pp. 48-63.
- Sharma, B., Lauano, F., Narayan, S., Anzeg, A., Kumar, B., & Raj, J. (2018). Science teachers accelerated programme model: A joint partnership in the Pacific region. *Asia-Pacific Journal of Teacher Education*, 46(1), 38–60. doi:10.1080/1359866X.2017.1359820.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1(2), 147-170. <https://doi.org/10.1177/026553228400100203>.
- Speaker, K., Taylor, D., & Kamen, R. (2004). Storytelling: Enhancing Language Acquisition in Young Children. *Education 3-13*, 125, 3.
- Teo, T., Yurdakul, I., & Ursavas, Ö. (2016). Exploring the digital natives among pre-service teachers in Turkey: a cross-cultural validation of the Digital Native Assessment Scale. *Interactive Learning Environments*, 24, 1231 - 1244. <https://doi.org/10.1080/10494820.2014.980275>.
- Tighe, E., & Schatschneider, C. (2016). Examining the Relationships of Component Reading Skills to Reading Comprehension in Struggling Adult Readers. *Journal of Learning Disabilities*, 49, 395 - 409. <https://doi.org/10.1177/0022219414555415>.
- Vaughn, S., Swanson, E. A., Roberts, G., Wanzek, J., Stillman Spisak, S. J., Solis, M., & Simmons, D. (2013). Improving reading comprehension and social studies knowledge in middle school. *Reading Research Quarterly*, 48, pp. 77-93.
- Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. *Harvard University Press*.
- Welch, C., & Piekkari, R. (2006). Crossing language boundaries: Qualitative interviewing in international business. *Industrial Marketing Management*, 35(4), pp. 429-438.
- Xue, Yajing, Wang, Yijun, [Retracted] Artificial Intelligence for Education and Teaching, *Wireless Communications and Mobile Computing*, 2022, 4750018, 10 pages, 2022.
- Yenduri, Gokul & Murugan, Ramalingam & Govardanan, Chemmalar & Supriya, Y & Srivastava, Gautam & Reddy, Praveen & Raj, Deepti & Jhaveri, Rutvij & B, Prabadevi & Wang, Weizheng & Vasilakos, Athanasios & Gadekallu, Thippa. (2024). GPT (Generative Pre-Trained Transformer)— A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. *IEEE Access*.

APPENDIX

A1: Interview Questions and Translations

"Q1: In your opinion, was the AI-based text generation activity more interesting for students than the traditional reading activities found in their books?" (Sizce yapay zekaya dayalı metin üretme etkinliği öğrenciler için kitaplarında bulunan geleneksel okuma etkinliklerinden daha mı ilgi çekiciydi?)

"Q2: How did the activities involving artificial intelligence affect the students' foreign language performance compared to the traditional ones?" (Yapay zeka içeren faaliyetler geleneksel olanlara göre öğrencilerin yabancı dilde performansını nasıl etkiledi?)

"Q3: After this experiment, do you think that students are more inclined to read English texts?" (Bu deneyden sonra öğrencilerin İngilizce metinleri okumaya daha meyilli olduklarını düşünüyor musunuz?)

"Q4: What are your thoughts on the procedural content created by artificial intelligence?" (Yapay zekanın oluşturduğu prosedürel içerik hakkındaki düşünceleriniz nelerdir?)

"Q5: Were the answers provided by artificial intelligence accurate and logical enough to be used regularly in language classes?" (Yapay zekanın verdiği cevaplar dil sınıflarında düzenli olarak kullanılabilecek kadar doğru ve mantıklı mıydı?)

"Q6: What are your thoughts on artificial intelligence?" (Yapay zeka hakkındaki düşünceleriniz nelerdir?)

"Q7: Would you like to use artificial intelligence in your future lessons?" (Yapay zekayı derslerinizde kullanmak ister misiniz?)