T.C.
BURSA ULUDAĞ UNIVERSITY
INSTITUTE OF EDUCATIONAL SCIENCES
DEPARTMENT OF FOREIGN LANGUAGE EDUCATION
ENGLISH LANGUAGE TEACHING PROGRAM

# IMPROVING YOUNG ADULT EFL STUDENTS' SPEAKING SKILLS THROUGH AN ARTIFICIAL INTELLIGENCE CHATBOT: A QUASI-EXPERIMENTAL STUDY

**MASTER'S THESIS**

**Emre Can AYDEMİR**
**ORCID: 0009-0009-8346-3441**

**BURSA - 2025**

T.C.
BURSA ULUDAĞ ÜNİVERSİTESİ
INSTITUTE OF EDUCATIONAL SCIENCES
DEPARTMENT OF FOREIGN LANGUAGE EDUCATION
ENGLISH LANGUAGE TEACHING PROGRAM

# IMPROVING YOUNG ADULT EFL STUDENTS' SPEAKING SKILLS THROUGH AN ARTIFICAL INTELLIGENCE CHATBOT: QUASI-EXPERIMENTAL STUDY

**MASTER'S THESIS**

**Emre Can AYDEMİR**
**ORCID: 0009-0009-8346-3441**

**Danışman**
**Prof. Dr. A. Amanda YEŞİLBURSA**

**BURSA - 2025**

## BİLİMSEL ETİĞE UYGUNLUK

Bu çalışmadaki tüm bilgilerin akademik ve etik kurallara uygun bir şekilde elde edildiğini beyan ederim.

Emre Can
AYDEMİR
16.06.2025

**TEZ YAZIM KILAVUZUNA UYGUNLUK ONAYI**

"Improving Young Adult Efl Students' Speaking Skills Through An Artifical Intelligence Chatbot: Quasi-Experimental Study" adlı yüksek lisans tezi Bursa Uludağ Üniversitesi Eğitim Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırlanmıştır.

| Tezi Hazırlayan | Danışman |
| --- | --- |
| Emre Can AYDEMİR | Prof. Dr. A. Amanda YEŞİLBURSA |

Yabancı Diller Ana Bilim Dalı Başkanı
Prof. Dr. A. Amanda YEŞİLBURSA

## JÜRİ ONAY SAYFASI

Yabancı Diller Eğitimi Ana Bilim Dalı'nda 802293004 numara ile kayıtlı Emre Can Aydemir'in hazırladığı "Improving Young Adult EFL Students' Speaking Skills through an Artifical Intelligence Chatbot: Quasi-Experimental Study" konulu Yüksek Lisans çalışması ile ilgili tez savunma sınavı, 29/08/2025 günü 10:30-12:00 saatleri arasında yapılmış, sorulan sorulara alınan cevaplar sonunda adayın tezinin **başarılı** olduğuna **oy birliği** ile karar verilmiştir.

Sınav Komisyonu Başkanı

Doç. Dr. Kutay UZUN

Trakya Üniversitesi

WoS ResearcherID: P-8767-2017

Scopus Author ID: 57211437792

ORCID: 0000-0002-8434-0832

Üye (Danışman)

Prof. Dr. A. Amanda YEŞİLBURSA

Bursa Uludağ Üniversitesi

WoS ResearcherID: JAC-5894-2023

Scopus Author ID: 48663301600

ORCID: 0000-0002-4884-7388

Üye

Dr. Öğr. Üyesi Umut Muharrem SALİHOĞLU

Bursa Uludağ Üniversitesi

WoS ResearcherID: ABB-9052-2020

Scopus Author ID: 55588018000

ORCID: 0000-0002-5321-6556

**ACKNOWLEDGEMENTS**

First and foremost, I would like to express my deepest gratitude to my esteemed thesis supervisor, Prof. A. Amanda Yeşilbursa, for her invaluable guidance, unwavering support, and insightful feedback throughout every stage of this thesis. Her positive outlook and steadfast encouragement have been instrumental in helping me navigate the challenges of this research, and I am sincerely thankful for the opportunity to work under her mentorship.

I am also profoundly grateful to one of my closest friends, Lecturer Ömer Faruk Kaya of Bursa Technical University, for his continuous encouragement, wisdom, and companionship—not only during the thesis writing process but also throughout my academic journey. His presence has been a source of motivation and strength.

My heartfelt thanks go to my dear friend, Attorney Cüneyt Fidan, for his patience and companionship during the countless long study sessions we endured together. His moral support and understanding made this demanding journey more bearable.

Lastly, I would like to extend my appreciation to all of my friends whose names I cannot list here individually, but who have been there for me with their unwavering support, kindness, and encouragement when I needed it the most. Your presence has meant more to me than words can express.

Emre Can AYDEMİR

# ÖZET


| | |
|---|---|
| Yazar Adı ve Soyadı | Emre Can AYDEMİR |
| Üniversite | Bursa Uludağ Üniversitesi |
| Enstitü | Eğitim Bilimleri Enstitüsü |
| Ana Bilim Dalı | Yabancı Diller Eğitimi Ana Bilim Dalı |
| Bilim Dalı | İngiliz Dili Eğitimi |
| Tezin Niteliği | Yüksek Lisans Tezi |
| Sayfa Sayısı | XII + 147 |
| Mezuniyet Tarihi | 29 / 08 / 2025 |
| Tez Danışman(lar)ı | Prof. Dr. A. Amanda YEŞİLBURSA |


## YAPAY ZEKA SOHBET ROBOTU ARACILIĞIYLA YABANCI DİL OLARAK İNGİLİZCE ÖĞRENEN GENÇ YETİŞKİN ÖĞRENCİLERİN KONUŞMA BECERİLERİNİ GELİŞTİRME: YARI DENEYSEL BİR ARAŞTIRMA


Bu çalışma, yapay zekâ destekli bir sohbet botunun, İngilizceyi yabancı dil olarak öğrenen Türk genç yetişkinlerin konuşma becerilerini nasıl destekleyip geliştirebileceğini incelemiştir. Bu amaç doğrultusunda, açımlayıcı sıralı yarı deneysel desen kullanılarak nicel ve nitel veriler toplanmıştır. Nicel veri toplama süreci, ön test–son test uygulamaları ve kontrol grubunu içeren bir yarı deneysel desen çerçevesinde yürütülmüştür. Nitel veriler ise yansıtıcı günlükler, sohbet botuyla yapılan etkileşim kayıtları ve yarı yapılandırılmış görüşmeler aracılığıyla elde edilmiştir. Katılımcılar, yaşları 15 ile 24 arasında değişen ve dil yeterlik düzeylerine göre eşit şekilde deney ve kontrol gruplarına ayrılan 20 Türk İngilizce öğrenicisinden oluşmuştur. Sekiz hafta boyunca deney grubundaki katılımcılar, ders dışı etkinlik kapsamında haftalık olarak yapay zekâ sohbet botuyla etkileşime girmiştir; buna karşın kontrol grubuna yalnızca geleneksel yöntemlerle eğitim verilmiş ve herhangi bir sohbet botu etkileşimi sunulmamıştır. Nicel bulgular, her iki grup içinde ön test ve son test karşılaştırmalarına göre konuşma becerilerinde anlamlı bir gelişme olduğunu göstermiştir. Ancak deney ve kontrol gruplarının son test puanları arasında istatistiksel olarak anlamlı bir fark bulunmamıştır. Bu durum, çalışma süresi boyunca yapay zekâ destekli uygulamanın geleneksel öğretime kıyasla ölçülebilir bir avantaj sağlamadığını düşündürmektedir. Nitel bulgular ise, öğrencilerin üretken yapay zekâ sohbet botlarıyla ilgili teknik, pedagojik ve sosyal alanlarda çeşitli güçlü yönler, sınırlılıklar ve geliştirme önerileri deneyimlediklerini ortaya koymuştur. Ayrıca, dil yeterliği ve bilgi düzeyinde artışa dair öğrenci ifadelerine de bulgular arasında yer verilmiştir. Bu çalışmanın, elde ettiği sonuçlarla yapay zekâ temelli literatüre katkı sağlaması ve öğretmenler, sohbet botu geliştiricileri ile gelecek araştırmacılar için çeşitli çıkarımlar sunması amaçlanmaktadır.



***Anahtar Sözcükler:*** İngilizce konuşma becerileri, İngilizcenin Yabancı Dil Olarak Öğretimi, Mobil Öğrenme, Yapay Zeka destekli Sohbet Botu.

## ABSTRACT


| | |
|---|---|
| Name and Surname | Emre Can AYDEMİR |
| University | Bursa Uludag University |
| Institution | Institute of Educational Sciences |
| Field | Department of Foreign Language Education |
| Branch | English Language Program |
| Degree Awarded | Master |
| Page Number | XII + 147 |
| Degree Date | 29 / 08 / 2025 |
| Supervisor(s) | Prof. Dr. A. Amanda YEŞİLBURSA |


### IMPROVING YOUNG ADULT EFL STUDENTS' SPEAKING SKILLS THROUGH AN ARTIFICAL INTELLIGENCE CHATBOT: QUASI-EXPERIMENTAL STUDY


This study examined how employing a generative artificial intelligent (GenAI) chatbot could support and improve the speaking skills of young Turkish adult learners of English as a foreign language. Quantitative and qualitative data were collected for this purpose using an explanatory-sequential quasiexperimental design. Quantitative data collection followed a quasiexperimental design involving pre-and post-tests and a control group. To collect qualitative data, the study employed reflective diaries, recordings of interactions with the chatbot, and semi-structured interviews. The participants included 20 Turkish learners of English, aged between 15 and 24, equally divided into experimental and control groups based on their proficiency level. During eight weeks, the experimental group interacted weekly with the GenAI chatbot outside the classroom as an extracurricular activity, whereas the control group received traditional instruction without any chatbot interaction. The quantitative findings revealed significant improvement in speaking skills within each group, according to pre-test and post-test comparisons. However, no statistically significant difference was found between the experimental and control groups in their post-test scores. This suggests that the AI-supported treatment did not lead to a measurable advantage over traditional instruction within the study period. The qualitative findings also indicated that learners experienced a range of technical, pedagogical, and social strengths, constraints, and suggestions for enhancements of GenAI chatbots. Reports of increased language proficiency and knowledge were also present in the findings. The study aims to contribute to the growing body of GenAI literature with its results, and it has implications for teachers, chatbot makers, and future researchers.



***Keywords:*** AI chatbots, English as a foreign language, MALL, Speaking skills

## LIST OF CONTENTS

*Page*

**BİLİMSEL ETİĞE UYGUNLUK** ..... **i**
**TEZ YAZIM KILAVUZUNA UYGUNLUK ONAYI** ..... **ii**
**JÜRİ ONAY SAYFASI** ..... **iii**
**ACKNOWLEDGEMENTS** ..... **iv**
**ÖZET** ..... **v**
**ABSTRACT** ..... **vi**
**LIST OF CONTENTS** ..... **vii**
**LIST OF TABLES** ..... **x**
**LIST OF FIGURES** ..... **xi**
**LIST OF ABBREVIATIONS** ..... **xii**

**CHAPTER 1**
**INTRODUCTION**

1.1. Background to the study ..... 1
1.2. Statement of the problem ..... 3
1.3. Purpose of the Study and Research Questions ..... 4
1.4. Significance of the Study ..... 5
1.6. Assumptions ..... 6
1.7. Limitations of the Study ..... 6

**CHAPTER 2**
**LITERATURE REVIEW**

2.1. Introduction ..... 8
2.2. Theoretical background of GenAI Chatbots in EFL ..... 8
2.2.1. Sociocultural Theory ..... 8
2.2.3. Self-Regulated Learning ..... 10
2.3. Intelligent Computer- and Mobile-Assisted Language Learning ..... 12
2.4. Historical Overview of Chatbots ..... 14
2.4.1. Generative AI Chatbots ..... 15
2.5. Artificial Intelligence in Education ..... 16
2.6. Benefits of Generative AI Chatbots in EFL ..... 18
2.6.1. Generative AI Chatbots for Developing Language Skills in EFL ..... 18
2.7. Challenges on the Use of Generative AI Chatbots in EFL ..... 23

**CHAPTER 3**

**METHODOLOGY**

3.1. Design of the Study ..... 26
3.2. A Quasi-Experimental Pre-test Post-test Group Research Design ..... 26
3.3. Context and Participants ..... 28
3.3.1. Context ..... 28
3.3.2. Participants ..... 28
3.4. Data Collection Tools ..... 29
3.4.1. Speaking Pre-test–Post-test ..... 30
3.4.1.1 Reliability and Validity of the Speaking Test ..... 32
3.4.2. Semi-Structured Interviews ..... 34
3.4.3. Reflective Diaries ..... 35
3.4.4. Credibility and Trustworthiness ..... 35
3.4.5. ChatGPT ..... 37
3.5. Piloting ..... 37
3.6. Data Analysis ..... 38
3.7. Quality of the Research ..... 39
3.8. Ethical Considerations ..... 40
3.9. Limitations of the Methodology ..... 41

**CHAPTER 4**

**RESULTS**

4.1. Results Pertaining to the Comparison of Speaking Skill Development Between the Experimental and Control Groups ..... 43
4.1.1. Descriptive Statistics of the Speaking Pre-tests ..... 43
4.1.2. Descriptive Statistics of the Speaking Post-tests ..... 43
4.1.1. The Normality Tests of the Speaking Scores ..... 44
4.1.3. Comparision of the Speaking Pre-test and Post-Test Scores ..... 46
4.2. Strenghts, Challenges, and Suggestions for Improvement Identified in Learner Diaries ..... 49
4.2.1 Technological Strengths of ChatGPT Identified in Learner Diaries ..... 54
4.2.2 Pedagocial Strenghts of ChatGPT Identified in Learner Diaries ..... 61
4.2.3 Social Strenghts of ChatGPT Identified in Learner Diaries ..... 64
4.2.4 Technological Challenges of ChatGPT Identified in Learner Diaries ..... 67
4.2.5 Pedagogical Challenges of ChatGPT Identified in Learner Diaries ..... 72

4.2.6 Social Challenges of ChatGPT Identified in Learner Diaries ..... 75
4.2.7 Technological Suggestions of ChatGPT Identified in Learner Diaries ..... 77
4.2.8 Pedagogical Suggestions of ChatGPT Identified in Learner Diaries ..... 80
4.2.9 Social Suggestions of ChatGPT Identified in Learner Diaries..... 82
4.3. Findings Regarding Learners' Experiences from the Semi-Structured Interviews ..... 83

**CHAPTER 5**

**DISCUSSION**

5.1. Discussion of Speaking Skill Gains in the Experimental Versus Control Groups ..... 99
5.2. Interpreting the Technological Benefits, Challenges, and Suggestions from Learner Experiences ..... 101
5.3. Interpreting the Pedagogical Benefits, Challenges, and Suggestions from Learner Experiences ..... 104
5.4. Interpreting the Social Benefits, Challenges, and Suggestions from Learner Experiences ..... 109

**CHAPTER 6**

**CONCLUSION**

6.1. Summary of Findings ..... 114
6.2. Implications for English Language Teachers ..... 115
6.3. Implications for Chatbot Developers ..... 116
6.4 Suggestions for Future Research ..... 117
6.5. Concluding Remarks ..... 117

**REFERENCES ..... 119**

**APPENDICES ..... 132**

APPENDIX 1: Institutional Research Approval Document..... 133
APPENDIX 2: Ethics Committee Approval Certificate..... 135
APPENDIX 3: Experimental Group Consent and Demographic Information Form..... 137
APPENDIX 4: Control Group Consent and Demographic Information Form ..... 139
APPENDIX 5: Speaking Rubrics – Oral Presentation and Spoken Interaction..... 141
APPENDIX 6: Student Weekly Reflective Diary Template ..... 143
APPENDIX 7: Semi-Structured Interview Questions ..... 146

**CURRICULUM VITAE ..... 147**

# LIST OF TABLES

*Table* *Page*


1 The Design of the Quasi-experiment ..... 27
2 Gender Distribution in Control and Experimental Groups ..... 29
3 Descriptive statistics of the overall speaking score pre-test ..... 43
4 Descriptive statistics of the overall speaking score post-test ..... 44
5 Skewness and kurtosis of the experimental and control groups in speaking pre-test exam results ..... 44
6 Tests of normality for pre-tests scores ..... 45
7 Skewness and kurtosis of the experimental and control groups in speaking post-test exam results ..... 45
8 Tests of normality for post-test Scores ..... 46
9 Paired Samples t-test findings for the speaking scores of the experimental group ..... 47
10 Paired samples t-test findings for the speaking scores of the control group ..... 48
11 Independent samples t-test results for speaking pre-test scores ..... 48
12 Independent samples t-test results for speaking post-test scores ..... 49
13 Overview of frequencies related to the use of ChatGPT identified in learner diaries ..... 50
14 Frequencies for the technological strengths of ChatGPT identified in the learner diaries ..... 56
15 Frequencies for the pedagogical strengths of ChatGPT identified in the learner diaries ..... 62
16 Frequencies for the social strengths of ChatGPT identified in the learner diaries ..... 64
17 Frequencies for the technological challenges of ChatGPT identified in the learner diaries ..... 68
18 Frequencies for the pedagogical challenges of ChatGPT Identified in Learner Diaries ..... 73
19 Frequencies for the social challenges of ChatGPT Identified in Learner Diaries ..... 75
20 Frequencies for the technological suggestions of ChatGPT Identified in Learner Diaries ..... 78
21 Frequencies for the pedagogical suggestions of ChatGPT Identified in Learner Diaries ..... 80
22 Frequencies for the social suggestions of ChatGPT Identified in Learner Diaries ..... 82

# LIST OF FIGURES

*Figure* *Page*


1 Zone of Proximal Development and Scaffolding Model....9
2 Learning analytics process map of establishing a framework for SRL model for GenAI Chatbot....10
3 The REPOST Model.....22

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AI | : Artificial Intelligence |
| AIEd | : Artificial Intelligence in Education |
| AIML | : Artificial Intelligence Markup Language |
| AR | : Augmented Reality |
| ASR | : Automatic Speech Recognition |
| CALL | : Computer-Assisted Language Learning |
| CEFR | : Common European Framework of Reference for Languages |
| EFL | : English as a Foreign Language |
| ELCC | : English Language Course Center |
| ELT | : English Language Teaching |
| GenAI | : Generative Artificial Intelligence |
| GPT / ChatGPT | : Generative Pre-trained Transformer (OpenAI Chatbot) |
| HCI | : Human-Computer Interaction |
| ICALL | : Intelligent Computer-Assisted Language Learning |
| IMALL | : Intelligent Mobile-Assisted Language Learning |
| IPAs | : Intelligent Personal Assistants |
| L2 | : Second Language |
| LLMs | : Large Language Models |
| MALL | : Mobile-Assisted Language Learning |
| MAXQDA | : Name of qualitative data analysis software |
| MKO | : More Knowledgeable Other |
| NLP | : Natural Language Processing |
| PCs | : Personal Computers |
| REPOST | : Reciprocity, Evaluation, Personalization, Omnipresence, Socio-cultural, Teachableness |
| SDT | : Self-Determination Theory |
| SPSS | : Statistical Package for the Social Sciences |
| SRL | : Self-Regulated Learning |
| UNESCO | : United Nations Educational, Scientific and Cultural Organization |
| WTC | : Willingness to Communicate |
| ZPD | : Zone of Proximal Development |

# CHAPTER 1
# INTRODUCTION

## 1.1. Background to the study

While English language education continues to evolve, the integration of artificial intelligence (AI) technologies, such as AI chatbots, is becoming increasingly popular. The development of communication systems on computers since the 1970s has brought chatbots on display. These communication systems have been called by a range of terms including personal digital assistant, virtual personal assistant, conversational agent, chatbot, and conversational user interface (McTear, 2020). As claimed by McTear, there are varying interpretations to define chatbots, and it creates ambiguity in its definition.

AI chatbots, also known as chatterbots, are computer programs which are designed to participate in humanlike conversations using numerous AI techniques such as natural language processing (NLP), machine learning, (Kohnke et al., 2023; Wu & Yu, 2023). Chatbots, which utilize AI, impersonate a person's speech, so users of AI chatbots are often provided with a natural language interface which is similar to communication with a real person (Kim et al., 2019). Intelligent chatbots manage to do a variety of tasks, including retrieving information and playing music, but their interactions are mostly limited to transactional exchanges (Dizon, 2021). ChatGPT, Google Assistant, Replika, Amazon Lex, and Apple's SIRI are examples of popular AI-enabled chatbots available online.

Intelligent personal assistants (IPAs), also known as “human-computer dialogue systems, machine conversation systems, virtual agents, dialogue systems, or chatterbots” (Jia & Chen, 2009, p. 903), are more commonly known today as digital assistants or chatbots (Lee et al., 2020). Widely used IPAs are Amazon's Alexa, Apple’s Siri, and Google Assistant. These systems are employed to replicate neural networks in the human nervous system, aiming to copy the processes by which humans remember, think, and speak (Koç, 2024).

There are clear distinctions between traditional and AI chatbots. Traditional chatbots, often referred to as rule-based systems, were limited to performing specific tasks and had numerous constraints (Kohnke et al., 2023; Shumanov & Johnson, 2021). Early dialogue systems were often rigid and prone to errors upon receiving unexpected inputs as they were typically designed for narrow purposes (McTear, 2020). Additionally, they focused only on spoken or written modes, neglecting other forms of communication. However, modern AI chatbots have overcome many of these limitations. They are now more accessible and functioning across various devices such as smartphones, PCs, and smart speakers. AI chatbots

can now be integrated into social robots and smart devices, providing a much more natural and dynamic user experience (McTear, 2020).

According to Lee et al. (2020), the classification of chatbots can be based on their purpose: assistant chatbots and conversation chatbots. Assistant chatbots can assist users with everyday activities such as scheduling, web searches, and making phone calls, and are usually embedded in smartphones (e.g., Siri, Google Assistant). These types of chatbots often have more straightforward and rigid responses (McTear, 2020). In contrast, conversation chatbots offer more detailed, dynamic, and natural exchanges resembling human dialogues. Notable examples of conversational chatbots are ChatGPT and Mitsuku (also known as Kuki), which are both designed to engage in human-like and extended interaction.

The need for chatbots in various fields has enabled their further development. The chatbots with the ability to mimick human interaction can understand and react to user inputs through text or voice, which makes them indispensable tools for different applications such as education, customer service, personal assistance, and healthcare (Lee et al., 2021). They are developed for entertainment, mimicking characters or entertainers. They are commonly employed as virtual assistants, whether guiding users on websites, answering frequently asked questions, pretending to be virtual sales or support agents, or administering surveys and quizzes (Kim et al., 2019). As highlighted by Kerly et al. (2006), the advancement of chatbot technology has led to its expanded use across various domains, including entertainment, business, and public services.

In English as a Foreign Language (EFL) classes, AI-powered chatbots can be great tools to offer students more interactive and personalized learning opportunities by promoting learner engagement (Fryer et.al, 2006; Mohamed, 2023). Their relevance extends to both mobile-assisted language learning (MALL) and subsequently intelligent computer-assisted language learning (ICALL), providing flexible, accessible, and on-demand learning sessions (Koç, 2024). While activities are predetermined in traditional computer-assisted language learning (CALL) systems, ICALL makes use of various AI technologies for better comprehension and response to the needs of individual learners in real time (Schmidt & Strasser, 2022). Such systems can perform analysis on language patterns, offer tailored feedback, simulate natural conversations, and adjust learning tasks based on students' learning pace along with their needs.

While AI continues to be integrated into education, understanding the relationship between students' perceptions toward AI and their communication behaviors becomes notably important (Al-Murtadha, 2019; Dewaele & Pavelescu, 2021; Zhi & Wang, 2024). Using AI chatbots in ELT provides students with overall opportunities to foster their motivation, language

skills, and language learning experiences by creating a safe and supportive environment (Zhi & Wang, 2024; Peng, 2020). According to Zhai and Wibowo (2022), the application of specifically trained second language (L2) chatbots may even alleviate the level of language learning anxiety, which is considered one of the most common barriers among many students.

Moreover, AI-assisted language learning, including the use of chatbots, significantly improves writing skills among EFL students, enhancing various aspects of writing such as organization, coherence, grammar, and vocabulary (Song & Song, 2023; Lin & Chang, 2020). Similarly, AI chatbots have been shown to effectively improve speaking competence (Yang et al., 2022; Yin & Satar 2020; Hsu et al., 2023), willingness to communicate (WTC) (Fathi et al., 2024; Tai & Chen, 2020), and listening skills (Dizon, 2020) among EFL learners by providing consistent practice and immediate feedback.

In summary, as these tools along with the ICALL educational technologies continue to evolve, the utility of chatbots in facilitating language education is becoming prominent every day, making them an essential component of novel language education strategies. This study will investigate the effectiveness of incorporating chatbots into English language teaching, with a focus on offering numerous benefits, such as increased learner engagement with a supportive space for speaking skills practice.

### 1.2. Statement of the problem

In the previous chapter, the pivotal role of AI chatbots in improving the language skills of EFL learners was discussed. Many students, particularly in environments where English exposure is minimal, like Türkiye, find it hard to practice speaking due to some societal factors, such as a lack of English language input in their daily lives, homes, or schools (Kara et al., 2017). Therefore, EFL learners have limited opportunities to develop oral language skills outside of the classroom. Giving students access to ICALL technologies that can be used for language learning is one method teachers can use to help language learners (Koç, 2024).

EFL learners frequently experience speaking anxiety, which can be attributed to various factors such as limited opportunities for practice in authentic environments and constrained classroom time (Tanveer, 2007). Additionally, environmental factors, including the fear of negative evaluation by peers and instructors, further worsen speaking anxiety among EFL learners (Öztürk & Gürbüz, 2014). Studies related to AI-mediated interactions have shown that AI can provide learners with a low-anxiety environment, thereby enhancing their speaking proficiency and willingness to communicate (Fathi et al., 2024; Koç, 2024; Tai & Chen, 2020).

Computer-mediated learning has already become an integral part of EFL classrooms because both teachers and students receive a myriad of benefits (Al-kadi, 2018). For instance,

using an online language-learning platform, like Duolingo, to supplement speaking and listening practice is one way to implement mobile-assisted language-learning practices in an EFL classroom. The teacher can give students specific modules to complete at home, emphasizing conversational skills. These types of online platforms offers real-time corrections and immediate feedback on language skills like pronunciation, grammar, and vocabulary, allowing students to practice speaking in low-pressure settings unlike the classrooms (Fathi et al., 2024). This would eventually promote autonomy and enable students to use English in authentic contexts.

With the advent of generative artificial intelligence (GenAI), Platforms with AI have started to produce a considerable volume of high-quality content in various multimedia formats, receiving widespread recognition from the public (Cao et al., 2023; Wang et al., 2024). This innovative paradigm for human-computer interaction offers opportunities to challenge the limits of conventional MALL and CALL practices in classrooms. For instance, one of the GenAI chatbots and operated by OpenAI, ChatGPT has been gaining vast recognition and is considered the beginning of a new AI revolution (Cao et al., 2023; Pack & Maloney, 2023). This recent technological advancement along with the integration of natural language processing, text-to-speech functions, and automatic speech recognition (ASR) has created novel opportunities such as AI-mediated chatbots for students to practice speaking English beyond the traditional classroom, contributing to more dynamic and effective language learning (Fathi et al., 2024; Koç, 2024).

While some research has been done on AI-mediated chatbots in the context of EFL, there is still a significant gap in the literature when it comes to their impacts and applications (Kim et al, 2020; Koç, 2024). As a result, little research has been done on the usefulness of chatbot-mediated instruction in improving the speaking abilities of EFL learners, despite the increased interest in this approach (Fathi et al., 2024). This research aims to fill this gap by investigating the impact of AI chatbots on improving the speaking skills of young adult EFL students with limited access to English input in their daily environments. The study draws upon the work of Koç (2024), who investigated the development of listening and speaking skills through a similar AI-based approach in a quasi-experimental design.

### 1.3. Purpose of the Study and Research Questions

This study seeks to determine if interaction with a GenAI chatbot as an extracurricular activity can facilitate the improvement of speaking skills among young adult EFL learners, particularly in fluency and coherence, grammatical accuracy and range, vocabulary, and pronunciation. Through a quasi-experimental methodology, the study will determine the

effectiveness of this novel technology in improving speaking skills. In addition to quantitative experimental data, qualitative methods will define EFL learners' perspectives on the advantages, challenges, and suggestions for improvement of using Chatbot-mediated interaction. It will set the foundation for future research on the integration of AI technologies in language education.

RQ1: Do significant differences exist in speaking proficiency scores among Turkish young adult EFL learners in the control group compared to those in the experimental group who interacted with a GenAI chatbot?

RQ2: Do significant differences exist in the speaking proficiency test scores between the pre-test and post-test within the control and experimental groups after the treatment period?

RQ3: Regarding the participants' perceptions:

a. What are the perceived positive aspects of employing GenAI chatbots in English language learning?

b. What are the perceived negative aspects of employing GenAI chatbots in English language learning?

c. What are the participants' suggestions for improving the use of GenAI chatbots in English language learning?

**1.4. Significance of the Study**

The education world is beginning to acknowledge the importance of AI-enabled chatbots in developing the speaking abilities of EFL students. Compared to educational chatbots in their early stages, modern chatbots have evolved to establish meaningful conversations, providing space for exchanging ideas and developing conversations to create deep interactions. In their review of articles published between 2015 and 2021, Sharadgah and Sa'di (2022) highlight the promising future of AI in English Language Teaching (ELT), emphasizing that AI has a positive impact on enhancing English skills, translation, and assessment.

In addition, using AI in EFL classrooms offers advantages such as a personalized learning experience, immediate feedback, and a more interactive learning environment (Huang et al., 2022). These digital resources continue to provide fresh approaches to many problems that students encounter when honing their speaking skills. Fathi et al. (2024) reported in their recent study that chatbot-mediated interactive speaking activities considerably improved the EFL learners' speaking skills and willingness to communicate by increasing their intrinsic motivation. Kim et al. (2021) found that AI chatbots helped students improve their English-speaking skills and provide a comfortable learning environment for practice.

Collectively, the findings above highlight the potential of AI for transforming the teaching and learning of English, offering personalized and accessible support to learners. Although there have been such contributions in this area, studies on the use of artificial intelligence chatbots for EFL still remain limited, with only a handful of chatbots developed to aid language users in engaging through voice recognition or multimodal interactions (Koç, 2024). This study, incorporating the latest voice-enabled GenAI chatbot technologies, aims to showcase their effects on the improvement of speaking skills among Turkish young adult EFL learners.

**1.6. Assumptions**

1- It is assumed that students in the experimental group will follow the instructions and regularly interact with ChatGPT during the treatment period outside of class.

2- It is assumed that participants have enough digital skills to use the chatbot on their own without needing extra help.

3- It is also assumed that ChatGPT can hold meaningful and coherent conversations that are suitable for speaking practice.

4- Any major changes in speaking scores are expected to result mainly from using the chatbot—not from outside or uncontrolled factors.

5- It is assumed that students will write honest, thoughtful reflections in their weekly diaries.

6- The speaking test used in this study is assumed to be a valid and reliable way to measure key aspects of speaking, like fluency, coherence, grammar, vocabulary, and pronunciation.

7- Before the treatment starts, both the experimental and control groups are assumed to be at a similar level in terms of speaking ability and English exposure.

8- Lastly, it’s assumed that ChatGPT will stay accessible and work properly throughout the study, without serious technical problems.

**1.7. Limitations of the Study**

Despite careful planning and implementation, this study is subject to several limitations that may affect the generalizability and interpretation of the findings:

1- This study was carried out with a small group of young adult EFL learners from a specific region and classroom setting in Türkiye. Because of that, the results might not apply to all EFL learners, especially those from different places or age groups.

2- The treatment period lasted only eight weeks, which may not have been long enough to fully show the long-term effects of practicing speaking with a chatbot.

3- Since the chatbot practice happened outside regular class time, personal factors like students' motivation, time management, and discipline may have affected how often and how seriously they engaged with the tasks.

4- The study heavily used reflection diaries for qualitative data collection. Students might not always give accurate descriptions of how they feel or exaggerate it.

5- Some students might have faced issues like unstable internet, which could have made it harder for them to use the chatbot consistently. Any technical problems or chatbot downtime may have interrupted the experience.

6- While the study focused mainly on speaking—especially fluency, accuracy, vocabulary, and pronunciation—it did not look closely at how using the chatbot might affect other skills like listening, reading, or writing.

7- Because ChatGPT is a new and interesting tool for most learners, their excitement at the beginning might have temporarily boosted motivation. This initial excitement might not last or show up the same way in future studies.

8- Finally, as helpful as ChatGPT can be, it sometimes gives responses that include errors, off-topic content, or culturally awkward replies. These are out of the researcher’s control and might have shaped the students’ experiences and results.

# CHAPTER 2
# LITERATURE REVIEW

## 2.1. Introduction

This literature review will first examine the study's theoretical framework that showcases the fundamental learning theories, which underpin the application of Generative AI (GenAI) chatbots in English as a Foreign Language (EFL) instruction. Next, to contextualize this study in the larger perspective of AI-assisted language learning, Intelligent Computer- and Mobile-Assisted Language Learning (ICALL and MALL), followed by an examination of the role of Artificial Intelligence in Education (AIEd) will be presented. Furthermore, a historical overview of chatbots will provide insights into their evolution. This will lead to a detailed discussion of GenAI chatbots and their specific contributions to language learning. Finally, the section on GenAI in English Language Teaching will highlight the current applications and pedagogical implications of this technology, with a specific focus on speaking skills. Through this review, the study aims to establish a comprehensive understanding of how GenAI chatbots can enhance EFL learning and contribute to the field of language education.

## 2.2. Theoretical background of GenAI Chatbots in EFL

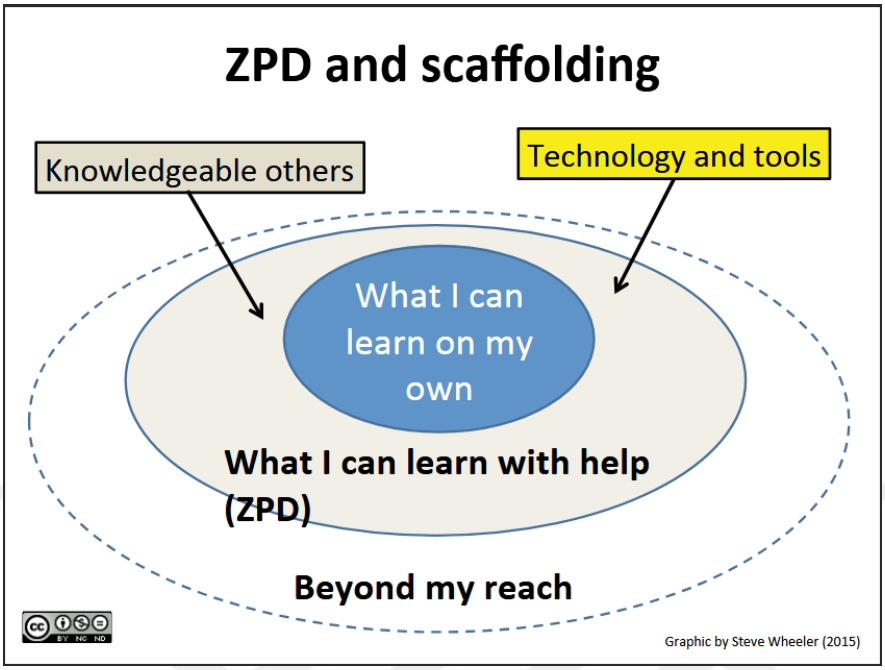
The integration of GenAI technologies, such as Large Language Models (LLMs), including ChatGPT, into educational contexts has opened new avenues for research and practice. These tools align with established learning theories, including the sociocultural theory and self-regulated learning.

### 2.2.1. Sociocultural Theory

Sociocultural Theory posits that cognitive development is deeply rooted in social interactions and mediated by cultural tools, such as language and artifacts (Mahn, 1999). Central to this theory are the concepts of the Zone of Proximal Development (ZPD) and scaffolding (see figure 1). The ZPD refers to the gap between a learner’s independent problem-solving abilities and their potential when guided by a more knowledgeable other (MKO) (Mahn, 1999). Scaffolding, on the other hand, involves the temporary support provided by the MKO to help learners achieve tasks within their ZPD (Shooshtari & Mir, 2014). These ideas emphasize the importance of guided learning and social interaction in fostering higher cognitive development. In this context, GenAI tools such as ChatGPT can serve as mediational means that provide scaffolded support within the learner’s ZPD (Kulaksız, 2024).

**Figure 1**

*Zone of Proximal Development and Scaffolding Model.* Reprinted from Wheeler (2015, p. 68).




Although teachers are usually the ones who provide scaffolding in traditional classrooms, classmates, parents, and even technology systems can also provide scaffolding. GenAI chatbots such as ChatGPT provide learners with real-time feedback, serving as a type of technological scaffolding (Özçelik & Keskin, 2024). In a conversational style with a GenAI chatbot, for instance, students can exchange ideas, clear up misconceptions, and investigate novel ideas. This promotes knowledge internalization through dialogue, resembling classroom interactions. (Liao et al., 2024). This is a clear example of ZPD, in which learners finish activities with assistance.

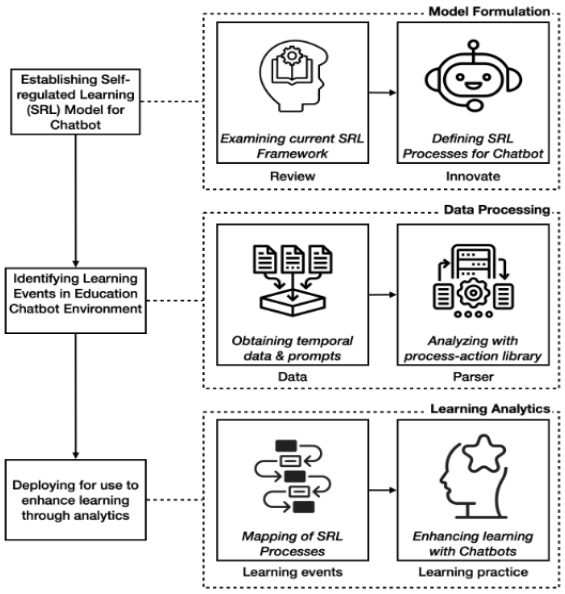
As students engage with these resources, they gradually internalize techniques and become less dependent on outside assistance, ultimately completing assignments on their own (Liao et al., 2024). However, Giannakos et al. (2024) mention that while GenAI tools like ChatGPT offer considerable potential, they cannot fully recreate the emotional and relational components of human care. Effective learning does not only require cognitive assistance. In addition to cognitive help, teachers provide the emotional support, warmth, and trust that are necessary for effective learning (Chew & Cerbin, 2020). As a result, several scholars emphasize that GenAI should be viewed as a supplementary aid rather than a replacement for instructors, who remain essential for providing the emotional and relational support required for effective learning (Chew & Cerbin, 2020; Giannakos et al., 2024).

### 2.2.3. Self-Regulated Learning

Self-regulated learning (SRL) plays a vital role in education, allowing students to take charge of their learning through goal setting, self-monitoring, and self-assessment. Key theoretical perspectives on SRL draw from various frameworks that emphasize cognitive awareness, motivation, and social influences (Lai, 2024). Metacognitive theories highlight the importance of individuals monitoring and controlling their cognitive processes (Flavell, 1979, as cited in Lai, 2024). Zimmerman's cyclical model of self-regulation outlines three interrelated phases: forethought, performance, and self-reflection, each playing a crucial role in guiding learners' self-regulatory behaviors (Zimmerman, 2000). Pintrich's model highlights the role of motivation in SRL, emphasizing how goal orientation shapes learners' ability to regulate their learning (Pintrich, 2000). Bandura's social cognitive theory contributes to this discussion by focusing on self-efficacy, asserting that learners' beliefs in their own abilities influence their motivation and performance (Bandura, 2002). Additionally, Deci and Ryan's self-determination theory highlights the significance of autonomy, competence, and relatedness in fostering intrinsic motivation and effective self-regulation (Deci & Ryan, 2013, as cited in Lai, 2024). These theoretical foundations inform the design of SRL models within digital learning environments. Figure 2 illustrates a learning analytics process map for developing an SRL framework tailored to a GenAI chatbot, highlighting how data, modeling, and learner interaction are integrated into a cohesive support system.

**Figure 2**

*Learning analytics process map of establishing a framework for SRL model for GenAI Chatbot.* Reprinted from Lai (2024).



Advances in generative AI chatbots utilizing large language models have introduced innovative ways to enhance learning by providing individualized support and feedback. AI chatbots are being integrated into the existing SRL frameworks mentioned above, with studies showing that they can analyze student inputs, sort them into learning categories, and generate insights into how students interact with instructional material (Lai, 2024).

Chang et al. (2023) explore how AI chatbots can support SRL by emphasizing three key pedagogical principles: goal setting, feedback, and personalization. Drawing from Zimmerman's SRL framework, the study argues that AI chatbots should not merely provide answers but should be designed to foster self-regulation. The authors propose that effective chatbot integration in education requires teaching students how to prompt strategically, enabling reverse prompting for self-assessment, and utilizing data-driven learning analytics for personalized support. They stress that AI should be seen as a tool to enhance learning rather than a shortcut for academic work, advocating for collaboration among educators, designers, and researchers to refine AI-driven pedagogy (Chang et al., 2023).

Empirical studies have highlighted the impact of generative AI chatbots on SRL. Ng et al. (2024) conducted a pilot study comparing a generative AI-powered chatbot, SRLbot, with a rule-based chatbot, Nemobot, to examine their effects on students' self-regulated learning in science education. The study involved 74 secondary school students in Hong Kong over three weeks. Findings showed that SRLbot significantly boosted students' science knowledge, engagement, and motivation when compared to a rule-based chatbot. Thanks to its personalized suggestions and adaptive responses, SRLbot supported students in building steady study habits, easing their learning anxiety, and achieving better results. Students also shared in their feedback that they appreciated the chatbot's personalized support, and it set it apart from more limited, rule-based systems. Overall, these results show the potential of generative AI chatbots to foster self-regulated learning and provide useful direction for how such tools can be designed and used effectively in the classroom (Ng et al., 2024).

Another empirical study by Chiu (2024) explores the role of GenAI, specifically ChatGPT, in fostering self-regulated learning (SRL) in EFL contexts through self-determination theory (SDT). Using a Delphi design, the research gathered insights from 36 experienced teachers to develop a classification tool for ChatGPT-based learning activities that address SDT needs (autonomy, competence, and relatedness) across SRL phases (forethought, performance, and self-reflection). The findings indicate that ChatGPT effectively supports competence and autonomy needs, particularly in the performance and self-reflection phases, by providing instant feedback, generating practice problems, and facilitating self-assessment. However, it offers

limited support for relatedness. It highlights a gap in fostering student-teacher and peer connections. Overall, the study offers a practical framework for integrating ChatGPT into educational practices and emphasizes the need for future research to explore its role in ICALL learning.

To gain deeper insight into the topic, some AI-assisted chatbots, such as TeacherGAIA, have been designed specifically to support self-directed learning and self-assessment. These chatbots incorporate constructivist learning principles and prompt engineering techniques to encourage independent learning beyond the classroom. Focusing on both cognitive guidance and social-emotional support plays a role in fostering SRL in an AI-assisted manner (Ali et al., 2023).

In light of all this information, the growing presence of AI chatbots in education presents significant opportunities for enhancing SRL. These chatbots are being used to tackle challenges in problem-solving (Cohen & Cohen, 2024). They can be used to assist students in K-12 education in overcoming obstacles related to motivation and cognitive issues through their structured instruction and individualized feedback (Kong & Yang, 2024). This research advances the SRL theory and also provides educators who are looking to improve their students' self-regulated learning abilities with useful applications.

On the whole, by offering personalized support, constructive feedback, and cognitive guidance, these chatbots empower students to take greater control of their learning and improve their academic outcomes. As research in this field continues to evolve, educators and developers will be better equipped to harness the full potential of AI to transform the learning experience. The following chapter will expand on the central theme of this study—the applications of AI chatbots in English Language Teaching (ELT)—by examining their broader integration within larger systems, such as intelligent computer-assisted language learning (ICALL) and mobile-assisted language learning (MALL). This exploration will offer a more comprehensive perspective on the role of AI in language education.

### 2.3. Intelligent Computer- and Mobile-Assisted Language Learning

Intelligent Computer-assisted learning (ICALL) integrates technology with foreign language teaching to develop a highly contextualized and communicative learning environment. Computer-assisted language learning (CALL) and mobile-assisted language learning (MALL) technologies allow students greater exposure to the target language in more culturally relevant ways through easy access to multimedia, online materials, and computer-mediated interaction with native speakers (Amaral & Meurers, 2011). Nevertheless, according to researchers, traditional CALL systems usually cannot go beyond a very simple format, like

multiple-choice or matching exercises, and seriously lack nuanced feedback prepared for each learner. Consequently, there is a call to integrate a linguistic modeling system into CALL platforms to expand their ability to come up with more sophisticated exercises and to offer more comprehensive feedback that considers the specific traits of each user (Amaral & Meurers, 2011). These recent advancements have led to redefining CALL and MALL systems as ICALL and IMALL, reflecting the integration of artificial intelligence (AI), natural language processing (NLP), and other modern technologies (Bibauw, François, Van den Noortgate, & Desmet, 2022, as cited in Koç, 2024).

What ICALL tries to implement in the place of these limitations in use is the integration of NLP tools, such as parsers and semantic analyzers, which could ensure better error detection and feedback tailored to the needs of learners (Amaral & Meurers, 2011). As these technologies advance, ICALL stands as a promising tool in the broader landscape of education, effectively utilizing adaptive feedback to support personalized and life-long, quality learning in foreign language acquisition (Ai, 2017).

One of the important features of ICALL systems is the provision for direct corrective feedback, a sure trait for language learning. Choi (2014) says that even human teachers do not easily manage to provide timely feedback whenever classes have larger numbers, and this is the challenge that ICALL systems are able to handle through their enhanced tutoring capabilities. A system that analyzes learners' uses of language and provides them with individually tailored feedback helps learners recognize their errors and thus correct them to achieve enhancement of their language

Moreover, error analysis plays an essential role in ICALL to provide insight into the learning process. According to Uria et al. (2014) errors are part of learning, any language and such error analysis is also in ICALL systems when creating instructional strategies to offer effective language teaching. This approach is in tune with the wider objectives of ICALL in developing alterable learning environments responsive to individual learners' needs.

Being one of the influential tools in ICALL systems, intelligent conversational agents in the form of chatbots pushed language learning further by allowing learners to practice the language interactively. In fact, Huang et al. (2023) have mentioned that such AI-powered chatbots are able not only to engage students in a dialogic process and analyze their responses but also to provide serious feedback to them, thus allowing for more intensive learning of the foreign language. This interactive attribute of such conversational agents enhances the language skills and even builds up the confidence of learners in using the target language.

Various types of AI chatbots function on computers, smartphones, or tablets, utilizing the embedded mechanisms within these platforms. Thus, classifying these advanced tools solely under ICALL or IMALL would be insufficient (Koç, 2024). This paper aims to explore the broader discussion of ICALL and IMALL research by integrating these concepts in a unified way, particularly in light of recent advancements in intelligent systems embedded within various types of devices.

**2.4. Historical Overview of Chatbots**

Turing's (1950) question, "Can machines think?" is a straightforward and intricate question. According to McCarthy et al. (1955, as cited in Tlili et al., 2023), the term artificial intelligence (AI) was introduced to describe devices and systems that mimic human cognitive functions and decision-making skills. Humans want to communicate with computers in a similar manner to how they do with language. According to Zadrozny et al. (2000), the best way to support Human-Computer Interaction (HCI) is to let users intuitively communicate their goals using speech, typing, or gestures. This idea served as the basis for the creation of chatbots. These advancements led Alan Turing to develop the Turing Test in 1950, formerly known as the Imitation Game, as a novel method to determine whether a machine can display intelligent behavior that is either identical to or indistinguishable from that of a person (Turing, 1950). In this test, a human assessor converses in natural language with a machine and another person, identifying which responses are from the computer and not knowing which is which. As one of the first frameworks for evaluating machine intelligence, the Turing Test is significant because of its fundamental role in the study of AI.

The first chatbot, named ELIZA, was developed in 1960s by Weizenbaum (Adamopoulou & Moussiades, 2020). ELIZA turned the user's remarks into questions, simulating the actions of a psychiatrist. Despite having limited communication capabilities, it was a major source of inspiration for the development of subsequent chatbots. However, it had notable limitations, such as restricted knowledge, which confined its discussions to a specific set of topics. Additionally, ELIZA struggled to maintain extended conversations and lacked the ability to learn or infer context from ongoing interactions (Adamopoulou & Moussiades, 2020).

In 1972, a more sophisticated system was later introduced. PARRY simulated a schizophrenia patient with rudimentary emotional responses. However, it had trouble understanding language and carrying on lengthy conversations (Colby et al., 1971, as cited in Adamopoulou & Moussiades, 2020). Around the mid-1980s, Jabberwacky showed the world the first use of AI in chatbot technology, employing contextual pattern matching (Adamopoulou & Moussiades, 2020), while ALICE (1995) introduced the use of Artificial Intelligence Markup

Language (AIML), significantly expanding its conversational abilities (Marietto et al., 2013). In contrast to ELIZA, which only utilized around 200 keywords and rules, ALICE's knowledge base is said to comprise over 41,000 templates and connected patterns (Heller et al., 2005, as cited in Gumusel, 2024). However, ALICE lacked both emotional expressiveness and intelligent qualities (Adamopoulou & Moussiades, 2020). Along with all these, when SmarterChild was introduced in 2001, it was revolutionary because it included useful features like database retrieval, indicating a move toward useful, task-oriented chatbots (Molnár & Zoltán, 2018).

The 2010s saw the rise of intelligent personal assistants (IPAs) such as Siri, Google Assistant, and Amazon Alexa, which transformed chatbot technology by integrating speech recognition, advanced AI technologies (Smutny & Schreiberova, 2020). These IPAs were able to perform a myriad of tasks, including managing schedules, retrieving information, and home automation. One disadvantage is they sometimes faced challenges with language nuances and comprehension of the context. Chatbots are widely used in marketing, healthcare, education, and entertainment by the middle of the last decade, and their spread was fueled by social media and messaging apps (Dale, 2016). In contrast to their early ancestors like ELIZA, modern chatbots have made tremendous progress in their conversational abilities. They share personal histories and simulate human relationships.

As evidenced by their widespread presence and profound history, many information-driven platforms, such as universities, libraries, and museums, have already integrated chatbots into their online services (Fryer et al., 2020). Fryer et al. also emphasize that they are not a recent innovation but has been used for decades. However, this study specifically focuses on generative AI chatbots, which is a rapidly evolving technology that has recently gained global attention for its transformative impact on education and language learning.

### 2.4.1. Generative AI Chatbots

November 2022 was a milestone for the history of chatbots as ChatGPT (Generative Pre-trained Transformer), a chatbot created to support realistic conversational exchanges, was unveiled by OpenAI, an artificial intelligence research company (Bozkurt, 2023; FathiTlili et al., 2023). Van Dis et al. (2023) describe ChatGPT as a “large language model (LLM), a type of machine learning system that learns on its own from data and is trained on a vast corpus of text to produce sophisticated and presumably intelligent writing.” It uses natural language processing (NLP) to answer follow-up questions, admit mistakes, question erroneous assumptions, and reject inappropriate requests (OpenAI, 2023). The term "generative" in generative AI (GenAI) means that they can generate new data, such as text, images, audio, or

video, instead of only analyzing the existing data (Cao et al., 2023; Wang et al., 2024). In other words, GenAIs can create content by learning different patterns from large datasets, which enables them to produce coherent outputs based on the existing knowledge.

It must also be acknowledged that GenAI is a more comprehensive term than 'ChatGPT.' In addition to ChatGPT, there are other chatbots capable of generating both text and images within the same platform (Cogo et al., 2024). Widely used educational platforms such as Quizlet, Quizizz, Edpuzzle, and Mentimeter are progressively incorporating AI functionalities to assist language teachers and streamline their tasks (Cogo et al., 2024).

Altogether, this signifies that with continuous technological advancements, GenAI features are becoming an integral part of our daily work, making it increasingly difficult to function without them (Cogo et al., 2024). Even while the main objective of GenAI chatbots is to mimic human speech, they are capable of much more than just dialogue. Within their spectrum of capabilities, they can replicate a variety of roles and functions and produce creative output like poetry, stories, and novels (Tlili et al., 2023). For example, with its incredible ability to simulate human thought processes, ChatGPT presents a novel AI technique that may be able to pass the Turing Test (Turing, 1950). In summary, GenAI chatbots are groundbreaking technologies, which causes a profound paradigm change in education as well as in many facets of daily life.

**2.5. Artificial Intelligence in Education**

While ICALL systems stand out as one of the most dynamically developing aspects of language education (Koç, 2024), the evolution of ICALL is closely linked to the growing role of AI in education. Particularly, the emergence of GenAI has further enhanced language learning by enabling more sophisticated, personalized, and autonomous experiences. As conventional machine learning and now deep learning in GenAI raise the number of intelligent services with human-like behavior (Hwang et al., 2020; Huang et al., 2023), artificial intelligence in education (AIEd) has been growing rapidly. Thus, AI has recently been applied in several domains, including decision-making, natural language processing, voice and image recognition, and interlanguage translation (Hwang et al., 2020). Additionally, they are widely used in robotics, computer programs, applications, and equipment control systems (Hwang et al., 2020). Hwang et al. (2020) identified four types of AI-powered education applications. The first type is the intelligent tutor, which can meet the students' needs and provide them with positive learning outcomes. The second type is the intelligent tutee, which provides active learning, encouraging learners to be tutors. The third one introduces intelligent learning tools or partners, gathering and analyzing students' data in order to improve learning. The fourth

category of policy-making advisor applications, in turn, helps administrators understand educational trends and problems and support them in making appropriate decisions effectively.

Building on the conversation about AIEd, one significant advancement in this paradigm is the use of AI chatbots as teaching tools. Unquestionably, the use of AI chatbots is increasing since they can simulate human-like interactions and offer students individualized support and conversational practice (Fathi et al., 2024). AI chatbots enhance the learning process by imitating human conversation patterns, making them valuable for language learning, as the feedback is always contextualized (Zhang & Huang, 2024). Within the framework of Hwang et al.'s (2020) four categories of AI applications in education, chatbots are strongly related to the functions of "intelligent tutor" and "intelligent learning tool or partner." By adapting their responses according to the learner's input and development, chatbots can serve as intelligent tutors and meet individual learning needs, promoting a learner-centered approach (Ifelebuegu et al., 2023). Because chatbots can target particular language difficulties, track progress, and encourage language fluency through repeated practice, this flexibility has important ramifications for personalized learning (Harry, 2023).

AI chatbots also function as "intelligent learning partners," gathering information from every conversation (Hwang et al., 2020). Based on real-time performance data, educators can improve curriculum design and instructional practices by analyzing this data to get insights into common issues faced by students (Hwang et al., 2020). By monitoring trends and spotting learning gaps, the use of chatbots in English language teaching (ELT) not only helps individual students but also advances broader educational ideas (Hwang et al., 2020).

However, the widespread adoption of AIEd comes with certain challenges, particularly regarding the need for robust data privacy and ethical frameworks to protect student information (Ifelebuegu et al., 2023; Harry, 2023). Furthermore, although chatbots show promise, their effectiveness is mostly dependent on the caliber of NLP models and the chatbot's capacity to react appropriately to intricate inputs. Therefore, even if chatbots have many benefits, their use should be carefully considered in light of ethical and technological constraints in order to optimize their potential as useful teaching tools.

Chatbot use in EFL raises the possibility that chatbots driven by AI could be essential supplements to conventional instruction, giving students tailored practice and feedback that was previously challenging to obtain in traditional classrooms (Fathi et al., 2024). Thus, this study will examine students engaging with a GenAI chatbot to determine whether these anticipated benefits hold true in practice.

## 2.6. Benefits of Generative AI Chatbots in EFL

While chatbots have been utilized for many decades now (Adamopoulou & Moussiades, 2020; Huang et al., 2022), GenAI chatbot studies are obviously restricted to the past few years because it is a relatively new research issue (Özçelik & Ekşi, 2024). According to the researchers, the transformative role of ChatGPT in English Language Teaching (ELT) comes with both its potential benefits and inherent challenges (Bozkurt, 2023; Dwivedi et al., 2023; Kostka & Toncelli, 2023; Marzuki et al., 2023). In fact, a special issue in English Language Teaching Journal (Volume 78, Issue 4, 2024) highlights the evolving role of technology, including generative AI, in language education. The articles in the issue provide foundational discussions on the transformative potential of GenAI in ELT, the implications of integrating AI tools into EFL classrooms, and explore how GenAI fosters learner autonomy and inclusivity while addressing both the opportunities and challenges educators face. All this means that GenAI technologies are widely acknowledged within the ELT community.

Moreover, in their study, Kostka and Toncelli (2023) argue that while GenAI tools, such as ChatGPT, encourage reflection on new learning approaches, educators must model and guide responsible relationships with these tools. As García-Peñalvo (2023) notes, banning or denying AI will not halt its influence, making it imperative for educators to explore its applications thoughtfully. Thus, the evolving role of educators presents an opportunity to redefine pedagogical strategies that align with 21st-century learning needs while maintaining the core values of language education (Kostka & Toncelli, 2023).

### 2.6.1. Generative AI Chatbots for Developing Language Skills in EFL

The integration of generative AI technologies, particularly chatbots have introduced new opportunities for various aspects of language learning. Much of the existing literature on generative AI chatbots explores specifically writing (Dwivedi et al., 2023; Fitria, 2023; Lin & Chang, 2020; Liu et al., 2023; Özçelik & Ekşi, 2024; Song & Song, 2023; Su et al., 2023). In contrast, studies focusing on other essential language skills, such as speaking and listening, remain relatively scarce (Koç, 2024).

Marzuki et al. (2023) explored the impact of various AI writing tools on the quality of EFL students' writing. EFL teachers focused on their content and organization, and they shared their perspectives. The study utilized prominent AI tools such as Quillbot, ChatGPT, WordTune, Essay Writer, etc. The findings indicated enhancements in students' academic performance. Furthermore, teachers agreed on the positive influence of these tools in improving the clarity and logical flow of students' writing.

Similarly, According to Dwivedi et al. (2023), ChatGPT can provide basic educational materials and eliminate the need for extensive internet searches. ChatGPT provides personalized feedback on writing tasks by analyzing grammar, organization, and content, which helps students grow academically. It can also handle administrative work like grading or answering common questions, which eventually gives teachers more time to focus on actual teaching. In language learning, ChatGPT supports students with grammar, vocabulary, and translation, offering one-on-one help and adapting its explanations to match each learner's understanding. Beyond that, it also brings a more interactive and engaging feel to online education. It makes digital learning environments more dynamic and student-friendly.

Other researchers even suggested implementing ChatGPT into classes. Dos Santos et al. (2023) propose integrating ChatGPT into writing instruction through the framework of the revised Bloom's Taxonomy (Anderson et al., 2001) as it supports the development of higher-order thinking skills. Teachers can use ChatGPT by designing activities based on learners' needs, goals, and tasks. At the remember and understand levels on Bloom's Taxonomy, it helps with tasks like brainstorming and summarizing. At the apply level, it supports discussions by generating ideas and encouraging reflection. At the analyze level, it aids in breaking down complex topics, reflecting on AI-generated responses, and improving critical thinking. Their study highlights how ChatGPT can support students at different stages of the writing process, and it shows the chatbot's potential to support both foundational and higher-level writing skills in English language teaching.

Similarly, Kohnke et al. (2023) highlighted ChatGPT's role as a vocabulary-learning tool in their study. Students can learn the definitions of unknown words through ChatGPT. It can generate dialogues that are tailored to different proficiency levels. Lastly, they receive explanations in their first language. The study also suggests that the chatbot can provide customized support for language learners, making it a valuable scaffold in L2 development.

In their study, Su et al. (2023) investigate the integration of ChatGPT into argumentative writing classrooms, highlighting its dual role as both a writing evaluator and a virtual peer. The authors demonstrate how the chatbot can support students across various stages of writing—pre-writing, during-writing, and post-writing—by providing feedback on outlines, assisting with content revision, offering proofreading, and facilitating reflection through chat history. While ChatGPT shows promise in enhancing structural, dialogical, and linguistic aspects of argumentative writing, its effectiveness is contingent on the quality of user prompts and evaluation rubrics.

In a similar case study, Özçelik and Ekşi (2024) investigated the impact of ChatGPT on learning register in various writing tasks, involving 11 participants. They were all active users of ChatGPT for purposes such as poetry writing, chatting, problem-solving, and research assistance. The study explored students' perceptions of writing difficulties, and most participants identified writing as a challenging skill due to factors like grammar complexity, vocabulary demands, and variations in register. The research highlighted two key questions: the effectiveness of ChatGPT in enhancing self-editing skills and students' opinions on its use as a learning assistant. Findings, which align with previous research by Su et al. (2023), revealed that ChatGPT positively influenced students' self-editing abilities, particularly in formal writing, by providing grammar, punctuation, and structure corrections, which elevated their writing proficiency.

While much of the existing literature focuses on the role of generative AI chatbots in enhancing writing skills (Liu et al., 2023), language learning extends beyond written mode of communication (Koç, 2024). In addition to supporting writing and vocabulary learning, GenAI chatbots have also been recognized as a prominent trend in language education, with applications in speaking, reading, listening, and writing skills (Al-Obaydi et al., 2023; Fathi et al., 2024; Koç, 2024).

In a study exploring the effects of voice-based AI chatbots on Korean EFL middle school students, researchers found significant improvements in learners' speaking competence, including pronunciation, fluency, language use, and task completion (Han, 2020). The experimental group, which engaged in conversation with AI chatbots, outperformed the control group, suggesting that chatbot-assisted practice can enhance communicative competence. Additionally, the study reported positive shifts in students' affective domains such as interest, belief in language improvement, motivation, and reduced anxiety. These findings highlight the potential of voice-based chatbots to provide a low-stress, engaging, and autonomous learning environment that supports both performance and affective outcomes in language learning.

Similarly, Fathi et al. (2024) investigated the impact of AI-mediated interactive speaking activities on EFL learners' speaking skills and willingness to communicate (WTC), using Engeström's (1987) activity theory as a framework. Their findings showed that AI interactions significantly improved both speaking skills and WTC. These improvements were attributed to learners' enjoyment of the novelty of AI communication and the reduced anxiety from speaking without fear of judgment. The study highlighted that learners, already accustomed to using technology in daily life, were more inclined to engage with AI tools for language learning. The AI Chatbots provided a low-pressure environment, fostering fluency,

confidence, and self-efficacy through immediate feedback and structured prompts. While learners expressed positive attitudes towards the AI-mediated activities, appreciating the personalized feedback and varied speaking tasks, some limitations were identified, such as occasional errors made by the Chatbots and the absence of human-like interaction. The study concluded that while AI-mediated Chatbots effectively enhance speaking skills and WTC, they should complement, rather than replace, human instruction for a more comprehensive language learning experience.

Similarly, Koç (2024), in her doctoral dissertation, conducted a mixed-methods study to evaluate the impact of AI chatbots on Turkish EFL learners' speaking and listening skills. While Fathi et al. (2024) observed significant improvements in speaking skills through AI-mediated interactions, their findings contrast with Koç (2024), who reported that the quantitative results revealed no significant differences between the experimental and control groups' language proficiency. The researcher ascribes this to the short, two-month intervention period or a novelty effect that diminished over time. Nevertheless, the qualitative findings highlighted numerous pedagogical, technological, and social affordances of AI chatbots. Learners reported improvements in vocabulary acquisition, grammar knowledge, and confidence in speaking, alongside increased opportunities to practice interactional strategies such as clarification requests and semantic expansion. The study also emphasized that AI chatbots offer learners a non-judgmental, anxiety-reducing environment, fostering greater willingness to communicate. These findings align with Fathi et al. (2024).

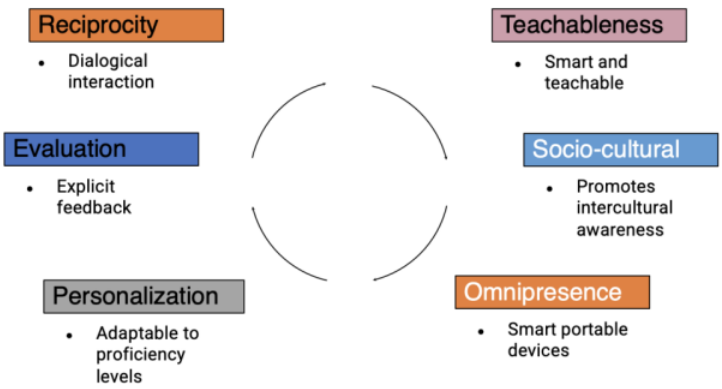
To further investigate the benefits of generative AI chatbots for developing language skills in EFL learners, Du and Daniel (2024) conducted a systematic review of articles published in SCCI journals demonstrates the multifaceted benefits of AI-powered chatbots for improving the speaking abilities of EFL students. Chatbots are incredibly useful tools as they promote fluency, reduce speaking anxiety, and offer real-time feedback on pronunciation. For example, students who engaged with chatbots such as SpokenBot and Google Assistant reported significant gains in speaking confidence and pronunciation correctness (Chen et al., 2020; Wu et al., 2023), which align with previous studies (Fathi et al., 2024; Koç, 2024). Additionally, chatbots create a helpful and immersive language learning environment by increasing student engagement through individualized feedback, interactive discussions, and multimedia materials. These resources have also been shown to be successful in increasing students' willingness to communicate by lowering anxiety and offering a practice environment free from criticism.

In line with the growing interest in exploring the effectiveness of AI chatbots in developing the speaking abilities of EFL learners, several conceptual models have been proposed in recent years to define the pedagogical characteristics of an ideal chatbot. Zhang, Zou, and Cheng (2023), for instance, introduced the RAISE model, which emphasizes Repetitiveness, Authenticity, Interactivity, Student-centeredness, and Enjoyment. According to this framework, AI chatbots can construct a motivating and learner-driven environment that fosters engagement through repetitive and authentic practice. Similarly, Jeon, Lee, and Choe (2023) proposed a model centered on Goal-orientation, Embodiment, and Multimodality, highlighting the importance of intentional design features such as task alignment, avatar interaction, and visual feedback tools. These elements have been shown to enhance learner experience and improve self-monitoring. Additionally, Topsakal and Topsakal (2022) suggested a design framework specifically for young learners, integrating gamification, voice interaction, and augmented reality (AR) within communicative language teaching principles. While these models offer important design insights, many remain theoretical or literature-based.

Building on these conceptual developments, Koç (2024) proposed a model specifically designed for pedagogical AI chatbot integration in language education, titled the REPOST model, which includes six core dimensions: Reciprocity, Evaluation, Personalization, Omnipresence, Socio-cultural awareness, and Teachableness. Unlike earlier models that are often generalized across educational domains or tailored for specific learner types, the REPOST framework was developed from empirical research on AI-mediated language learning and directly reflects learner expectations and challenges encountered during interaction with chatbots (see figure 3).

**Figure 3**

*The REPOST Model.* Reprinted from Koç (2024, p.244).



Thus, it provides a comprehensive blueprint for what an ideal AI chatbot should offer to support language acquisition in an EFL context. The present study, which similarly draws on learner feedback and diary reflections, refers to this model in the Discussion chapter to interpret and contextualize participants' suggestions. The alignment between learner experiences in this study and the dimensions of the REPOST model further strengthens its relevance as a practical framework for future chatbot design and implementation in language classrooms.

In summary, the integration of AI-powered chatbots into EFL education has demonstrated significant benefits, including enhanced speaking fluency, improved pronunciation, increased learner confidence, and reduced speaking anxiety (Chen et al., 2020; Du & Daniel, 2024; Fathi et al., 2024; Koç, 2024; Liu et al., 2023; Wu et al., 2023). Despite these promising outcomes, the adoption of chatbots in language education is not without its complexities. To fully understand their role in EFL instruction, it is essential to examine the challenges that may hinder their effectiveness in the long run. The following section will address these challenges, offering a valuable perspective on the limitations and potential obstacles associated with the use of generative AI chatbots in language learning contexts.

### 2.7. Challenges on the Use of Generative AI Chatbots in EFL

Although there are many benefits to incorporating GenAI into EFL, there are certain drawbacks as well. Concerns have been raised about the practical use of AI-driven chatbots in learning as instructors and students use them more and more (Dwivedi et al., 2023; Kostka & Toncelli, 2023; Marzuki et al., 2023; Özçelik & Ekşi, 2024; Su et al., 2023). Knowing what obstacles might come with GenAI allows educators to thoughtfully integrate it into their lessons, making sure it improves language learning instead of complicating it.

Özçelik and Keskin (2024) highlight that while ChatGPT is valued for its time efficiency and capacity to improve writing professionalism, concerns about its reliability, didactic limitations, and tendency to alter intended meanings were noted. The study also emphasized the important role of prompt engineering, as well-crafted prompts resulted in more precise feedback, which enhanced the learning process in the long run.

Kostka and Toncelli (2023) explore the applications of ChatGPT in English Language Teaching (ELT), highlighting both its potential benefits and challenges such as overreliance on chatbots and academic integrity issues. They argue that by encouraging genuine interaction, improving writing abilities, and offering prompt feedback, ChatGPT can promote language acquisition and learner autonomy. Nonetheless, issues were identified, including an excessive dependence on information produced by artificial intelligence, worries regarding academic integrity, and possible errors in the results. The authors stress the value of teacher direction and

advise a critical approach to incorporating ChatGPT, making sure its use supports meaningful learning and is in line with pedagogical goals. These findings align with Marzuki (2023), who also mentioned concerns about students becoming overly dependent on AI, potentially hampering their critical thinking and problem-solving abilities.

Kostka and Toncelli (2023) also highlight serious ethical implications of ChatGPT, including technological equity and language biases that privilege certain varieties of English. Concerns about access to AI tools and the need for a clear vision of responsible AI use are emphasized (UNESCO, 2021; Eke, 2023). Critical thinking and a focus on the learning process are essential to mitigate these challenges (Dwivedi et al., 2023; Zhai, 2022).

Similiary, Su et al. (2023) identify ethical issues as a primary limitation of using ChatGPT in writing classrooms. Authorship and plagiarism are significant concerns, as AI-generated texts blur traditional definitions of plagiarism. Despite OpenAI's efforts to filter harmful content, biased texts related to violence, sexism, and racism might still be generated (Godwin-Jones, 2021; Hutson, 2021). The study implies the ethical concerns associated with ChatGPT, and suggests that its role in argumentative writing should be carefully managed.

ChatGPT's inability to meet expectations in general speaking is another notable limitation. Zhang and Li (2021) and Thorp (2023) note that ChatGPT tends to guess the meaning of ambiguous expressions rather than seek clarification, revealing a lack of logical reasoning. Similarly, Jamshed et al. (2024) report that learners found the tool unable to grasp cultural nuance and often generating repetitive or shallow responses, which further undermines the quality of spoken interactions and reduces opportunities for meaningful communication. This eventually leads to plausible-sounding but incorrect content. A similar problem was noted by Du and Daniel (2024) as frequent disconnections during conversation were noted as a significant difficulty, especially given the diverse age range of users and their differing levels of attentiveness.

Another difficulty is that the format of the interaction is quite repetitive. According to Hsu et al. (2023), chatbots such as Alexa frequently offer limited English speaking opportunities. This makes it challenging to get students interested in interactive sessions in the long run. Furthermore, high-proficiency learners may benefit more from chatbots than their low-proficiency counterparts, as they may not be able to accommodate all proficiency levels (Chen et al., 2020).

Lastly, Dwivedi et al. (2023) point out several challenges linked to ChatGPT, including biased data, privacy concerns, and a lack of transparency in how its responses are formed. The model often reflects the biases found in its training material, and because its reasoning process

is not fully clear, it becomes harder to trust or trace how it generates certain outputs. This raises important concerns around student data privacy and security. On top of that, cheating or spreading misinformation cases show the need for stricter guidelines and responsible policies.

In summary, while GenAI chatbots like ChatGPT offer promising opportunities for enhancing EFL education, they also present significant challenges related to reliability, pedagogical effectiveness, ethical issues, data quality, and privacy. Addressing these challenges requires a critical, informed approach that emphasizes responsible AI integration and continuous evaluation of its impact on learning outcomes.

# CHAPTER 3
# METHODOLOGY

## 3.1. Design of the Study

This study employed a quasi-experimental pretest-posttest experimental and control group design using mixed methods data collection to investigate the impact of an AI chatbot, ChatGPT on the speaking skills development of Turkish young adult EFL learners. The research design employed in this study is adapted from Koç (2024), who used Replika AI chatbot to facilitate the development of speaking skills in EFL learners. Similar to Koç's approach, participants in this study will interact with the chatbot for an 8-week period, while their speaking proficiency is assessed through achievement tests. Such a design, falling into a category of complex mixed-methods designs (Creswell & Creswell, 2018), consisted of both quantitative and qualitative data so as to enhance the depth of understanding of the intervention's effects. The quantitative experiment formed the primary design, with qualitative data serving as a secondary source to corroborate quantitative findings. Qualitative data gave fuller contextual details of the intervention before, during, and after implementation.

## 3.2. A Quasi-Experimental Pre-test Post-test Group Research Design

The quasi-experimental study had two groups: an experimental group interacting with the AI chatbot and a control group receiving conventional instruction. Both groups were on the same level and received the English courses at the course center. For eight weeks, the experimental group carried out activities in self-directed learning by using ChatGPT outside the classroom. On the other hand, the control group continued classes as usual with no interaction with the AI chatbot.

The study followed a structured process. At the beginning, a pre-test was conducted to gather quantitative data, while qualitative insights were obtained from students' reflective diaries during the study and through interviews after the post-test phase. In a pre-test-post-test design with a non-randomized control group, participants were divided into either an experimental or a control group (Creswell, 2012). The pre-test results helped measure key dependent variables and allowed for statistical comparisons between the two groups before the intervention began.

In quasi-experimental research, a key step is deciding how the treatment conditions will be managed. Unlike fully experimental studies, participants in quasi-experiments are not randomly placed into groups. Instead, the researcher intentionally creates different conditions or versions of the independent variable to see how these changes affect the outcome. The

outcome, known as the dependent variable, is then closely observed to understand the effects of these deliberate adjustments (Creswell, 2012).

Later, the researcher introduced an AI chatbot as an intervention tool for one group while keeping the other group in a traditional learning environment. The goal was to compare the effects of this AI-assisted practice on students' speaking skills. Both groups were similar in terms of language proficiency, age, English learning experience, and academic background. Importantly, the lessons were delivered in an identical manner by the same instructor in both classes. However, students in the experimental group engaged in additional chatbot-based activities outside of class over an eight-week period, whereas those in the control group continued with their standard curriculum taught by the researcher (see Table 1).

**Table 1**

*The Design of the Quasi-experiment*

| Groups | Proficiency Level | Before Treatment | Treatment (8 weeks) | After Treatment |
|---|---|---|---|---|
| Experimental | Pre-Intermediate | Speaking Pre-test | Talking to ChatGPT through mobile devices | Speaking Achievement Post-test |
| Control | Pre-Intermediate | Speaking Pre-test | No additional activity other than the usual methods | Speaking Achievement Post-test |

*Note.* Table concept inspired by the quasi-experimental design outlined in Koç (2024).

Data in this study were collected through speaking pre-tests and post-tests, reflective diaries of language learners, and semi-structured interviews over two months at a state-affiliated language course in Türkiye. Throughout the eight-week treatment phase, the instruction and support for the students were provided by the researcher. The quantitative data, namely pre- and post-speaking tests, were used to assess learning outcomes before and after the eight-week intervention period. The pre-tests measured baseline speaking proficiency and allowed for a comparison of the groups at pre-treatment. The qualitative data, on the other hand, were acquired from the experimental group through reflective diaries during the treatment and semi-

structured interviews after the treatment phase ended. Pre-test and post-test scores were analyzed using statistical tests to compare the groups.

In the experimental group, students were instructed to interact with ChatGPT for approximately 20 minutes per week over an eight-week treatment period, yielding a target of 160 minutes per participant. Usage records indicated, however, that the actual time-on-task varied across individuals. On average, students engaged with the chatbot for 146 minutes (*SD* = 44.23), with totals ranging from 95 to 237 minutes. While this demonstrates general adherence to the planned schedule, the variation also reflects differing levels of engagement among participants.

In addition to these interactions, students submitted weekly voice recordings, on which they received formative feedback from the instructor. These recordings served as supplementary practice and monitoring tools but were not included in the formal data analysis, as the present study focused exclusively on speaking proficiency test scores and qualitative diary entries.

## 3.3. Context and Participants

### 3.3.1. Context

The experiment was conducted at a state university-affiliated English language course center (hereafter, ELCC) in Bursa, Türkiye. The course center applies an English proficiency test when grouping students. The students were assigned to their classes through the test, which had been designed by the state university according to the Common European Framework of Reference for Languages (CEFR) levels (Köse, 2004).

### 3.3.2. Participants

The participants of this study consisted of 20 Turkish young adult learners of English as a Foreign Language (EFL), aged between 15 and 24 years. These English language learners (hereafter, ELLs) were enrolled in a state university-affiliated ELCC in Bursa, Türkiye. They were selected through purposeful qualitative sampling, as they were placed into the classrooms through a CEFR-base proficiency test. A non-probability form of sampling, purposeful qualitative sampling is used when the members of the target population meet predefined criteria relevant to the study's objectives (Creswell, 2012; Given, 2008).

**Table 2**

*Gender Distribution in Control and Experimental Groups*

| | Control Group | Experimental Group | *N* |
|---|---|---|---|
| Women | 4 | 4 | 8 |
| Men | 6 | 6 | 12 |
| Total | 10 | 10 | 20 |

The ELLs were placed into two classes: an experimental group ($n$ = 10) and a control group ($n$ = 10). In terms of gender, both the control group and experimental group consisted of four women and six men. Overall, the combined total for both groups was eight women and twelve men (see Table 2). The experimental group engaged in weekly conversational interactions with ChatGPT over an eight-week period, completing reflective diaries after each interaction. The control group, in contrast, did not interact with ChatGPT and only participated in pre- and post-study speaking proficiency tests.

All ELLs had prior experience learning English but varied in their proficiency levels. Their English proficiency was assessed through the institution's oral placement exam, which was routinely used for grouping students into levels. This ensured a baseline comparison between the two groups. Informed consent and demographic information were secured from the experimental group (see Appendix 3) and control group (see Appendix 4), ensuring ethical standards before the study began. Participants were assured of their anonymity and the confidentiality of their responses throughout the research process.

Moreover, to protect participant confidentiality, all real names were replaced with pseudonyms throughout the thesis. These names were carefully chosen to reflect the participants' gender and cultural background, while ensuring their anonymity. Effort was made to select names that are not similar in sound or spelling to the participants' real names, so that their identities remain fully protected. This procedure was done to maintain the readability of the results chapter of the thesis.

### 3.4. Data Collection Tools

In this quasi-experimental mixed-methods study, speaking proficiency tests, reflective diaries, and semi-structured interviews were employed in accordance with the research objectives. The speaking proficiency tests were adapted from the ELCC's standardized assessment materials, which were developed by experienced instructors to ensure validity and

reliability (Köse, 2004). Prior to the main study, the researcher piloted the tests to confirm their appropriateness for the proficiency levels of the ELLs and the contextual characteristics of the language classrooms. The rationale for using these standardized tests was to provide both the experimental and control group participants with an assessment format they were already familiar with, as they had prior experience taking these tests in their regular coursework. Therefore, rather than designing a new, context-specific test, the researcher opted for an established assessment that aligned with the proficiency levels and prior experiences of the participants.

These tests were given both before and after the ChatGPT interaction activities to see whether there were any statistically significant changes in the speaking proficiency levels of ELLs in the experimental and control groups. To keep the evaluation process consistent and fair, the same proficiency test rubrics used by the language course center were applied during both testing periods.

During the study, learners in the experimental group took part in weekly speaking sessions with ChatGPT and kept reflective diaries to track their experiences. These diaries offered valuable qualitative insights into how they perceived AI-assisted language learning. To explore their thoughts more deeply, semi-structured interviews were carried out in the final stage of the research, giving learners the chance to share their experiences and attitudes toward using an AI chatbot to practice their speaking skills.

### 3.4.1. Speaking Pre-test–Post-test

A pre-test is a fundamental component of experimental research, which assesses participants' initial characteristics before introducing an intervention. Following the intervention, a post-test was conducted to evaluate any changes in the measured factors (Creswell, 2012). In this study, the pre-test was administered to determine the ELLs' initial speaking proficiency levels and to establish whether there were any significant differences between the experimental and control groups before the treatment.

The ELLs were students enrolled in an intensive English language course at a university-affiliated language center. Upon their initial enrollment, they were placed into classrooms based on their performance on the institution's standardized proficiency test (Köse, 2004), which aligned with the Common European Framework of Reference for Languages (CEFR). Each CEFR level (e.g., A1) lasted for approximately three months and was divided into three sub-levels (e.g., A1.1, A1.2, A1.3), with end-of-term exams determining students' progression to the next level. At the beginning of the treatment period, both the experimental and control groups were at the A2.3 level. Over the course of two months, students who

continued their regular coursework advanced to B1.2 while following the standard curriculum using Speak Out coursebooks by Pearson (SpeakOut 2nd Edition, 2015).

The pre-test was administered at the beginning of A2.3 to determine the ELLs' initial speaking proficiency levels and to establish whether there were any significant differences between the experimental and control groups before the treatment. After the eight-week intervention, the post-test was conducted at the beginning of B1.2 to measure students' progress and assess the impact of the AI chatbot on their speaking proficiency.

The speaking exam was conducted in a structured format, with students arriving in pairs to ensure a natural and interactive assessment environment. The exam consisted of two parts. In the first part, the oral presentation section, the researcher posed a series of open-ended questions to each student individually, requiring them to respond in an extended and coherent manner. Their performance in this section was evaluated using the oral presentation rubric (Köse, 2004), which assessed grammar, vocabulary, fluency, coherence, and pronunciation. The second part focused on spoken interaction, where students engaged in role-play activities using structured prompts. These prompts were carefully designed to simulate real-life conversational scenarios and were developed by the university to which the course center belongs. The students' spoken interaction skills, including grammar, vocabulary, fluency, listening skills, communication, and their ability to sustain dialogue were assessed using the spoken interaction rubric (Köse, 2004). The rubrics for assessing oral presentation and spoken interaction are provided in Appendix 5.

The pre-test and post-test rubrics for speaking assessment were developed by the university to which the course center is affiliated (Köse, 2004). The speaking proficiency was assessed on a 100-point scale, with oral presentation constituting 40 points and spoken interaction accounting for 60 points. This structure provided greater granularity, particularly in evaluating spoken interaction, which was central to the study's focus on the impact of the AI chatbot on students' speaking skills.

The oral presentation rubric included five equally weighted components—Grammar, Vocabulary, Fluency, Coherence, and Pronunciation—each allocated 8 points. The spoken interaction rubric consisted of Grammar (8), Vocabulary (12), Fluency (12), Listening Skills (8), Communication (8), and Sustaining the Dialogue (12). This configuration allowed for a comprehensive and balanced evaluation of both individual presentation skills and interactive communication proficiency.

Since both groups were enrolled in the same English course, there was a possibility that ELLs from the control group could interact with those from the experimental group and learn

about the chatbot. If control group learners had informally engaged with the chatbot outside the study, it could have reduced the contrast between the two groups. To mitigate this risk, learners in the experimental group were explicitly instructed not to share their experiences with their peers in the control group.

#### *3.4.1.1 Reliability and Validity of the Speaking Test*

Ensuring the reliability and validity of the speaking tests was a crucial aspect of this study, as these tests served as the primary instruments for measuring students' speaking proficiency before and after the intervention. The assessment process was carefully designed to maintain reliability and to ensure that the tests accurately measured students' speaking skills (Dörnyei, 2007).

Reliability underscores how consistent the test results are when repeated across different times and evaluators (Creswell, 2012). In this study, several steps were taken to improve the reliability of the speaking tests, and this ensured that the outcomes were stable and trustworthy. First, the use of a standardized test structure ensured that all students were assessed under the same conditions. Both the pre-test and post-test followed an identical format, which included an oral presentation section and a spoken interaction section. This consistency helped minimize differences between testing sessions and made the results more reliable.

Another important factor that supported reliability was the use of prompts and questions designed by the institution itself. These familiar and well-established materials contributed to a fairer evaluation process across all ELLs who took part in the study. The open-ended questions for the oral presentation section and the structured prompts for the spoken interaction section were developed by the university to which the ELCC belongs. Since these prompts had been thoughtfully created and used in an academic context, they provided a solid, standardized foundation for evaluating students' speaking performance (Köse, 2004). Their institutional origin also ensured that the content was appropriate and aligned with recognized language assessment standards.

To further support reliability, clearly defined rubrics (see Appendix 5) were used during assessment, which were developed by the university to which the ELCC belongs. These rubrics offered a more structured scoring framework. The oral presentation rubric, worth 40 points, focused on grammar, vocabulary, fluency, coherence, and pronunciation. The spoken interaction rubric, valued at 60 points, assessed grammar, vocabulary, fluency, listening skills, communicative ability, and dialogue maintenance. By breaking down speaking proficiency into specific, measurable components, these rubrics helped reduce subjectivity in scoring, thereby increasing internal consistency.

To assess inter-rater reliability on the speaking test scores, a second rater—an MA student in English Language Teaching who was also a coordinating teacher at the ELCC—independently evaluated half of the participants in both the experimental group ($n$ = 5) and the control group ($n$ = 5). Cohen's Kappa was then calculated using SPSS. The analysis revealed a substantial level of agreement between the two raters, $\kappa = .741$, $p < .001$, indicating strong consistency in scoring. This procedure is in line with recommendations for establishing reliability in qualitative and mixed-methods research (Miles & Huberman, 1994).

Alongside reliability, the study also took careful steps to ensure the validity of the speaking tests—that is, to make sure the assessment genuinely measured students' speaking proficiency. Content validity was achieved by including all key components of spoken language, such as grammar, vocabulary, fluency, coherence, pronunciation, listening skills, communication, and the ability to maintain a dialogue.

The structured format of both the oral presentation and spoken interaction sections allowed students to demonstrate a wide range of speaking abilities. On top of that, because the test prompts and questions were developed by the university, they were already in line with recognized academic standards for EFL assessment. This alignment further supported the validity of the testing process.

Another key aspect was construct validity, which refers to how well the test measured the theoretical construct of speaking proficiency (Dörnyei, 2007). The test format, which combined an oral presentation with a spoken interaction, aligned with internationally recognized frameworks such as the Common European Framework of Reference for Languages (CEFR) (Köse, 2004).

The speaking test also demonstrated face validity, meaning that it appeared to be a relevant and meaningful measure of speaking ability. The inclusion of open-ended questions in the oral presentation section and interactive role-play activities in the spoken interaction section made the test more authentic and realistic (Köse, 2004). Because the test mirrored real-life communication situations, students were likely to perceive it as a fair and appropriate way to assess their speaking skills.

Finally, the test exhibited criterion-related validity, as its scoring system and structure were aligned with established speaking assessment frameworks. Although no direct comparisons were made with standardized tests such as IELTS or Cambridge exams, the rubrics were adapted from the ELCC evaluation system, which shares similarities with internationally recognized speaking assessment criteria (Köse, 2004). This alignment suggests that the test

measured speaking proficiency in a manner consistent with widely accepted evaluation practices.

By integrating these reliability and validity measures, the speaking tests offered a consistent, well-structured, and meaningful way to assess students' oral proficiency. These carefully designed evaluation procedures ensured that the results accurately reflected students' speaking improvements over time, making the study's findings more robust and credible.

### 3.4.2. Semi-Structured Interviews

The semi-structured interviews were conducted with all participants in the experimental group (n = 10), with each interview lasting between 10 and 25 minutes. In total, approximately 148 minutes of audio-recorded data were collected and analyzed for this study.

To ensure the reliability of the information provided, the same question was rephrased and asked at different points during the interview process, allowing for consistency checks (see appendix 7). Throughout the process, the researcher offered guiding questions and prompts while avoiding rigid pre-categorization, ensuring that participants could elaborate on their thoughts freely and in detail. This approach fostered an exploratory discussion, encouraging deeper insights into the research topic (Dörnyei, 2007).

In this study, the interview tool was adapted from Koç (2024) to address the third research question, investigating the perceived benefits, challenges, and suggestions for improvement of using AI chatbots in English language learning. Participants identified the features that they have found useful in an AI-powered language-learning tool. The interviews took place after the experimental phase to provide further explanation and support for the quantitative findings. By examining participants' perspectives on using an AI chatbot, the study aimed to explore how this developing technology might improve students' experiences with speaking English. To ensure that the questions matched the study's goals and fit the research context, they were adapted carefully and reviewed by the thesis advisor before being used.

Even though meticulous measures were taken to ensure the reliability of the speaking tests, it was important to acknowledge that the test structure and rubrics were originally developed for assessing Turkish language learners, not English learners. As detailed in Köse (2004), the materials were designed and piloted with foreign students studying Turkish at ELCC's branches across Türkiye. While the tests align with the Common European Framework of Reference (CEFR), their application to Turkish EFL learners may introduce limitations in construct relevance. Furthermore, since these instruments were created over two decades ago, they may no longer fully align with current pedagogical standards or communicative language

teaching approaches. Therefore, the age and original target population of the assessment materials should be considered limitations in evaluating the study's results.

### 3.4.3. Reflective Diaries

Reflective diaries give participants the chance to take an active role in the research process, almost like becoming co-researchers themselves (Dörnyei, 2007). They serve as personal spaces where students can record their thoughts, opinions, and experiences throughout the study. These entries not only offer insight into the learners' perspectives but also help them make sense of what they're going through.

One key strength of using diaries is that learners write down their reflections while everything is still fresh in their minds, which leads to more detailed and accurate data (Dörnyei, 2007). Diary entries can also help fill in the gaps when research findings don't clearly explain the connections between variables. On top of that, they're a valuable tool for learners to look back and reflect on how far they've come.

In this study, the reflective diary tool was adapted from Koç (2024) to capture participants' experiences using the AI chatbot (see Appendix 6). Students were asked to maintain open-ended diaries, detailing their interactions with the chatbot over eight weeks. The primary goal of incorporating reflective diaries was to document students' ongoing experiences with as much depth as possible. Since interviews were conducted only at the end of the study, diaries provided a means for students to track their progress and challenges over time, ensuring that key observations were not lost or forgotten.

Reflective diaries also played an important role in data triangulation, adding to the reliability of the findings by providing a steady flow of qualitative insights throughout the study. Alongside the diaries, students recorded their voices each week as they interacted with the chatbot. These audio recordings were not analyzed as part of the qualitative data but were instead used as a monitoring tool to make sure participants were consistently engaging with the chatbot during the intervention.

Together, the diaries and weekly voice recordings served a practical purpose as well—they helped track student participation. Without this kind of documentation, it would have been much harder, if not impossible, to confirm that learners were regularly completing the AI-based speaking tasks.

### 3.4.4. Credibility and Trustworthiness

Credibility in qualitative research refers to the degree of alignment between participants' expressions and the researcher's interpretations, ensuring that the findings accurately reflect reality (Given et al., 2008). In this study, credibility was established by

thoroughly analyzing the experiences of Turkish EFL students in engaging with an AI-based chatbot for language learning and by representing their perspectives through multiple qualitative data sources, including weekly reflective diaries and semi-structured interviews (Jensen, 2008).

The interview questions and reflective diary prompts used in this study were directly adapted from Koç (2024) to suit the aims of the current research. While the overall structure and content of the original tools were mostly preserved, certain adjustments were made to better reflect the specific focus on speaking skills. For instance, sections related to listening were removed so that the diaries would concentrate solely on speaking-related experiences.

This adaptation process helped maintain the methodological integrity of the instruments while sharpening their focus, thereby enhancing the construct validity of the data. All necessary changes were reviewed and approved by the thesis supervisor and an additional MA student in English Language Teaching, which added an extra layer of credibility to the data collection tools.

Additionally, data analysis procedures were conducted transparently to maintain consistency. The coding process, along with the formation of themes, was explicitly detailed to illustrate how participants' responses were categorized and interpreted (Jensen, 2008). Another essential aspect of credibility, confirmability, was ensured by minimizing researcher bias in reporting findings. Confirmability in qualitative research refers to ensuring that the findings are not influenced by researcher bias and can be independently verified (Given et al., 2008). To enhance confirmability in this study, qualitative data from the reflective diaries and interviews were systematically analyzed using MAXQDA, a qualitative data analysis software that facilitated transparent coding and categorization of themes. This systematic approach helped to maintain consistency and reduce subjectivity in the interpretation process. Additionally, the coding scheme and analytical decisions were reviewed by the researcher’s supervisor, ensuring that the findings accurately reflected the participants’ perspectives rather than personal biases. This was achieved by presenting a clear coding framework in the findings section, allowing readers to trace how themes emerged from the raw data (Jensen, 2008).

Beyond credibility, this study also considers transferability, which refers to the extent to which findings can be applied to other settings (Given & Saumure, 2008). To facilitate transferability, a thick description of the study’s context, participants, and research design has been provided, allowing future researchers to assess whether similar findings might be observed in comparable conditions.

Lastly, dependability was addressed by ensuring that all methodological procedures were meticulously documented. The methodology section provides a detailed account of data collection instruments and analysis procedures, ensuring that future researchers can replicate the study under similar conditions (Jensen, 2008). By maintaining transparency in the research design and execution, this study upholds the key principles of trustworthiness in qualitative research.

#### 3.4.5. ChatGPT

At the start of the treatment process in December 2024, OpenAI had released an update to GPT-4o, which significantly enhanced its conversational capabilities. Additionally, this update improved the model's ability to process and analyze uploaded content, allowing for more comprehensive interactions (OpenAI, 2024).

GPT-4o was chosen as the primary GenAI chatbot for this study due to its status as the most prominent generative AI chatbot available, offering exceptional voice recognition capabilities. Although there are more personalized and customizable chatbots, such as Character AI or Replika, those platforms were not selected because the participants were high school students, and ensuring an age-appropriate environment was a priority. Additionally, while alternative chatbots designed specifically for EFL learners, such as ELSA Speak and Mondly, exist, they were not viable options due to their high costs and limited usability for research purposes. Thus, ChatGPT emerged as the most practical and effective tool for this study.

OpenAI allows its users to use Gpt-4o version free, although the usage is limited by the number of prompts. Thus, participants used the GPT-4o version, which additionally provided access to an advanced voice mode allowing up to 15 minutes of real-time conversation. This feature was also critical for the study, as it closely simulated real-life spoken interactions and enabled students to practice their fluency, pronunciation, and conversational skills. This mode also contributed to creating a more immersive and interactive learning experience, which aligns well with the study's goal of improving EFL students' speaking skills.

### 3.5. Piloting

A pilot study is defined as “a small-scale implementation of a larger study or part of a larger study” (Schreiber, 2008). It serves as a preliminary test to refine research instruments, identify potential issues in data collection, and assess the feasibility of the main study. Pilot studies help researchers address concerns related to sampling, procedures, and instruments, ensuring that the final study design is well-structured and effective (Schreiber, 2008).

In this study, a pilot phase was carried out to test the speaking assessments and reflective diary tasks before launching the full implementation. Due to limitations in participant access,

the pilot was conducted with three students from a different class rather than a larger sample. Even with this small number, the piloting process offered valuable feedback on how clear and understandable the research tools were.

The speaking tests were given to these students to check whether the test format, instructions, and question prompts made sense and matched the goals of the study. Their experiences during the test, including how well they understood the prompts and followed the structure, were carefully observed. Based on this feedback, small adjustments were made to fine-tune the instruments before the main study began.

Additionally, students were provided with the reflective diaries, which were modified from Koç's (2024) original version. The pilot participants were asked to complete the diaries over two weeks, providing feedback on whether the instructions were clear and whether they could comfortably reflect on their speaking experiences. This feedback helped confirm that the diaries were comprehensible and user-friendly before being distributed to the experimental group in the full study.

**3.6. Data Analysis**

To analyze the quantitative data in this study, IBM SPSS Statistics Version 27 was used. First, descriptive statistics such as means and standard deviations were calculated to get a general picture of the students' speaking scores before and after the treatment in both groups. Then, in order to decide whether parametric tests could be used or not, the normality of the data was checked through skewness, kurtosis, and the Shapiro-Wilk test. These tests helped ensure that the data were suitable for further statistical analysis.

After confirming that the data were normally distributed, as indicated by skewness and kurtosis values falling within the acceptable range of –1 to +1 for the experimental group (skewness = .661, kurtosis = –.779) and within ±2 and ±7 thresholds for the control group (skewness = –1.505, kurtosis = 3.105), as well as non-significant Shapiro–Wilk test results ($p > .05$), paired samples t-tests were run to see whether there was a significant difference between the pre-test and post-test scores within each group. This helped identify whether any progress was made over time. Following that, independent samples t-tests were used to compare the scores between the experimental and control groups. These tests were done both at the beginning (to make sure the groups started at a similar level) and at the end (to see whether the ChatGPT-supported treatment had any effect).

All statistical tests were interpreted using a significance level of $p < .05$. Along with p-values, mean differences and confidence intervals were also considered to better understand the results beyond just statistical significance.

In analyzing the diary data collected from participants, the responses to both questions—(A) general experiences with ChatGPT and (B) turn-taking experiences—were combined under the overarching themes of strengths, weaknesses, and suggestions for improvement. This decision was driven by the significant overlap in participants' responses, as many repeated similar points across both questions. Separating the responses into distinct coding categories would have introduced unnecessary redundancy without yielding deeper insights. Moreover, the combined approach aligns with the study's objective of capturing participants' holistic perceptions of their interactions with ChatGPT. Just as Koç (2024) categorized diary responses into technological and pedagogical affordances, this study adopted a smilar streamlined framework to reflect the natural integration of these aspects in participant feedback. The use of MAXQDA facilitated efficient coding, and merging responses allowed for clearer data synthesis without compromising analytical depth. This methodological choice ensures a more coherent presentation of findings while remaining responsive to the patterns emerging in the data.

The interview recordings were transcribed using Transkriptor, an AI-based tool designed to convert spoken language into written text. This made the transcription process faster and more manageable by giving the researcher more time to concentrate on the next steps of the qualitative analysis. To ensure accuracy, all transcriptions were carefully reviewed and manually corrected before moving on to the coding stage. The researcher coded the interview transcripts using MAXQDA and then compared each participant's interview data with their own diary entries to deepen and contextualize the findings. This comparison aimed to enrich the interpretation of individual learner experiences, even though no new themes emerged exclusively from the interviews. Instead, the interviews served to support, extend, or clarify the patterns already identified in the diaries. For instance, interview excerpts were used alongside diary quotes in the results chapter to highlight consistent learner perceptions over time and across data sources.

During the thematic analysis of the reflective diaries and semi-structured interviews, initial code generation, theme grouping, and refinement were carried out by the researcher. Language support and idea clarification were occasionally supported by AI-based tools (e.g., ChatGPT, Grammarly). However, all analytical decisions, interpretations, and thematic categorizations remained fully researcher-driven.

### 3.7. Quality of the Research

Ensuring the reliability of this research meant careful planning, ethical sensitivity, and close attention to detail at all phases. The study has been purposefully designed to generate

results not only valid and trustworthy, but also relevant to what the learners actually were experiencing when they used an AI chatbot to aid their language learning processes.

Firstly, the use of a mixed-methods, quasi-experimental method facilitated the balancing of various forms of data. By combining quantitative findings from speaking proficiency measures with qualitative information from reflection diaries and interviews, the study sought to provide more rounded insights on the impact of ChatGPT on learners' speaking development. This combination of approaches provided depth and internal validity to the study, rendering the findings more grounded and interpretable from various perspectives.

Each data instrument was selected carefully to match the learners' context and the study's purpose. The speaking test rubrics, for instance, were initially created by the institution but reworked to develop a more detailed and balanced framework for the purposes of the study. The reflection diaries and interview guides were taken from the literature but shaped to cover only speaking events, guided by the researcher's thesis supervisor and another researcher. The collaborative work added robustness to the instruments and made them clearer and applicable.

The research also gave careful attention to the qualitative data's credibility. Credibility was ensured through the use of multiple sources of data, consistent use of the prompts in the diaries, and transparent coding procedures. Confirmability was maintained through recording all coding choices and transparent analysis using MAXQDA. Thick descriptions of the settings, participant backgrounds, and the learning situation were given to ensure transferability—enabling others to see if and how the findings may transfer to similar settings.

### 3.8. Ethical Considerations

This study was conducted in full compliance with institutional ethical standards. Formal approval was obtained from multiple administrative and academic units to ensure the research met ethical, legal, and procedural requirements. First, the Institutional Research Approval was granted by Bursa Uludağ University, dated 15.10.2024 (see Appendix 1).

Subsequently, permission to carry out the study was granted by the ELCC on 07.11.2024 (see Appendix 2). This ensured that the research context and population (young adult learners enrolled in English language instruction) were institutionally approved. In addition, Ethics Committee Approval was issued by Bursa Uludağ University on 03.01.2025, confirming that the research adhered to ethical protocols regarding human participants (see Appendix 2).

Prior to data collection, written informed consent was obtained from all participants. Adult participants signed their own consent forms, while parental consent was obtained for minors. All participants were clearly informed of the study's aims, procedures, data usage, and

their right to withdraw at any stage (see Appendix 3 and 4). To protect participant identity, pseudonyms were used in all reporting. Data were stored securely, accessible only to the researcher. Audio recordings were made with full participant awareness, used solely to monitor engagement levels, and not analyzed for linguistic or content data.

Given the presence of underage participants, particular care was taken to ensure a psychologically safe, inclusive, and developmentally appropriate environment. The selection of ChatGPT over alternative tools was partly motivated by its perceived age-appropriateness and adaptability for educational use. This ethical framework reflects a strong commitment to academic integrity, participant safety, and the responsible implementation of AI technologies in language education research.

### 3.9. Limitations of the Methodology

Even though this research is thoroughly structured to study the ways an AI chatbot can facilitate the speaking ability of EFL learners, it is nevertheless worthwhile to acknowledge several limitations in the conduct of the research. These are not to diminish the significance of the results but to provide a better understanding of the circumstances under which the results were realized.

A key limitation, however, was having only 20 students, 10 per group. Having fewer students allowed the study to become more personalized, but it also means the findings cannot be easily generalized to larger groups. Additionally, the subjects were not randomized to the groups. Given the groups were comprised of existing classroom settings, there is even the possibility that the selection bias had an effect, although the students were meticulously put into classrooms according to their levels and ages.

Another factor to consider is the extent to which students used the chatbot outside of class. Because the ChatGPT tasks were autonomous and completed at their own time and pace, motivation levels would probably have been varied between students. Although weekly diaries and voice recordings were employed to monitor engagement, the research still depended on the honesty and reliability of the students. Some learners would have probably done more work than others, and this may have impacted the results.

The study's short time span of merely eight weeks is also noteworthy. Although it was sufficient to identify initial trends and short-term gains, it may not have been sufficient to take advantage of the maximum potential of employing ChatGPT for long-term speaking skill development (Koç, 2024). A longer study might have revealed deeper changes or helped track whether improvements lasted over time.

An additional limitation was that only the experimental group offered qualitative information via diaries and interviews. Although it made sense with the intervention, obtaining some reflections from the control group would have been beneficial to enable richer comparisons between the student experience and perception. Additionally, even though the reflective diaries were an important source of insight, the extent to which the students expressed themselves clearly or deeply probably varied according to individual writing abilities—and therefore may have had an impact on the caliber of the data.

Last but not least, while ChatGPT had been selected for being accessible, safe, and voice-friendly, having the free version of an external AI system presented challenges of its own. It is still a developing technologies, and modifications to the model GPT-4o would have introduced variations to the way the chatbot answered. As the researcher had limited influence on the technological changes, they represented an unavoidable limitation to the study. Apart from this limitation, the research still yields useful lessons on the potential applications of generative AI to support the learning of languages in authentic settings. Transparency about the limitations of the study actually makes the research stronger and provides future researchers with a solid basis to build further upon. With better study designs, bigger groups of participants, or longer treatment durations, future work can explore the promise of AI-facilitated language education even further and contribute to this rapidly developing field.

# CHAPTER 4
# RESULTS

## 4.1. Results Pertaining to the Comparison of Speaking Skill Development Between the Experimental and Control Groups

This part outlines the results related to the improvement of participants' speaking abilities following the experimental phase. Both descriptive and inferential statistical analyses are thoroughly reported.

### 4.1.1. Descriptive Statistics of the Speaking Pre-tests

Table 3 presents the descriptive statistics for the overall speaking scores in the pre-tests administered to both the experimental and control groups. As stated in Chapter 3, data from one control group participant were excluded from analyses due to significant deviations impacting the normality of the data distribution. Therefore, the control group analyses reported hereafter reflect data from the remaining nine participants. The experimental group ($N$ = 10) had a mean score of 64.40 ($SD$ = 15.85), with scores ranging from 46.00 to 91.00. The control group ($N$ = 9) showed a slightly lower mean of 62.88 (SD = 14.57), with minimum and maximum scores of 30.00 and 81.00, respectively. These results indicate that both groups had relatively similar levels of speaking proficiency prior to the treatment, with comparable means and standard deviations, suggesting baseline equivalency between the groups.

**Table 3**

*Descriptive statistics of the overall speaking score pre-test*

| Group | *N* | *Min.* | *Max.* | *M* | *SD* |
|---|---|---|---|---|---|
| Experimental | 10 | 46.00 | 91.00 | 64.40 | 15.85 |
| Control | 9 | 30.00 | 81.00 | 62.88 | 14.57 |

### 4.1.2. Descriptive Statistics of the Speaking Post-tests

As shown in Table 3, the post-test speaking scores reveal an overall increase in both groups following the eight-week intervention period. The experimental group ($N$ = 10), which engaged in AI-supported speaking practice using ChatGPT, achieved a mean score of 73.10 ($SD$ = 15.68), with scores ranging from 48.00 to 98.00. In contrast, the control group ($N$ = 9), which received no AI intervention, had a mean score of 69.96 ($SD$ = 16.02), with minimum and maximum scores of 34.00 and 89.00, respectively.

**Table 4**

*Descriptive statistics of the overall speaking score post-test*

| Group | *N* | *Min.* | *Max.* | *M* | *SD* |
|---|---|---|---|---|---|
| Experimental | 10 | 48.00 | 98.00 | 73.10 | 15.68 |
| Control | 9 | 34.00 | 89.00 | 69.96 | 16.02 |

### 4.1.1. The Normality Tests of the Speaking Scores

Skewness and kurtosis values provide insight into the distribution characteristics of the data. Skewness measures the asymmetry of the distribution, while kurtosis indicates the "tailedness" or the extent to which data points cluster around the mean (Hair et al., 2022).

Table 5 below presents the skewness and kurtosis values for the speaking pre-test scores. For the experimental group, the skewness was .661 and the kurtosis was –.779, both falling within the optimal range of –1 to +1, which is considered excellent for normality (Hair et al., 2022, p. 66). The control group, however, exhibited a skewness of –1.505 and a kurtosis of 3.105. Although these values slightly exceed the optimal range, they remain within the broader thresholds of ±2 for skewness and ±7 for kurtosis, which are generally accepted as indicators of normality in social sciences (Byrne, 2010; Hair et al., 2010).

**Table 5**

*Skewness and kurtosis of the experimental and control groups in speaking pre-test exam results*

| | Group | *N* | Skewness | Kurtosis |
|---|---|---|---|---|
| Total Score (Speaking) | Experimental | 10 | .661 | -.779 |
| | Control | 9 | -1.505 | 3.105 |

However, given the small sample size (*N*=19), to support these findings, the results of the Shapiro-Wilk and Kolmogorov-Smirnov tests for normality are provided in Table 6.

**Table 6**

*Tests of normality for pre-tests scores*

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| Experimental | .183 | 10 | .200* | .906 | 10 | .256 |
| Control | .218 | 9 | .200* | .862 | 9 | .100 |

The Shapiro-Wilk test, considered more suitable for small sample sizes, showed non-significant results for both the experimental group ($W$ = .906, $p$ = .256) and the control group ($W$ = .862, $p$ = .100), indicating that the data do not significantly deviate from a normal distribution. Taken together, these results—particularly the non-significant Shapiro-Wilk values—indicate that the assumption of normality was sufficiently met, justifying the use of parametric statistical procedures in the subsequent analyses. Although the Kolmogorov-Smirnov test was also conducted, it is generally more appropriate for larger sample sizes and is therefore interpreted with caution in this context.

Table 7 presents the skewness and kurtosis values for the post-test speaking scores of both groups. The experimental group exhibited a skewness of .142 and a kurtosis of .655, clearly falling within the excellent range of –1 to +1, indicating a normal distribution (Hair et al., 2022, p. 66). The control group showed a skewness of –1.447 and a kurtosis of 2.765, which—although slightly outside the optimal range—still remains within the generally acceptable bounds of ±2 for skewness and ±7 for kurtosis, as recommended in applied social sciences research (Byrne, 2010; Hair et al., 2010).

**Table 7**

*Skewness and kurtosis of the experimental and control groups in speaking post-test exam results*

| | Group | N | Skewness | Kurtosis |
|---|---|---|---|---|
| Total Score (Speaking) | Experimental | 10 | .142 | .655 |
| | Control | 10 | -1.447 | 2.765 |

To statistically confirm the assumption of normality for the post tests, the Shapiro-Wilk test was applied due to its appropriateness for small sample sizes ($n < 50$) (Ghasemi & Zahediasl, 2012). As shown in Table 8, the test indicated no significant deviation from normality for either the experimental group ($W = .982$, $p = .976$) or the control group ($W = .888$, $p = .191$). Although the Kolmogorov-Smirnov test was also conducted, it is generally more suitable for larger samples and thus interpreted with caution in this context. Based on these findings, it was concluded that the post-test scores were approximately normally distributed in both groups, supporting the use of parametric tests in further analysis.

**Table 8**

*Tests of normality for post-test Scores*

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | Df | Sig. | Statistic | df | Sig. |
| Experimental | .103 | 10 | .200* | .982 | 10 | .976 |
| Control | .236 | 9 | .157 | .888 | 9 | .191 |

### 4.1.3. Comparision of the Speaking Pre-test and Post-Test Scores

After confirming the assumption of normality across both groups through skewness, kurtosis, and Shapiro-Wilk tests, paired samples t-tests were conducted to examine whether there were statistically significant changes in speaking performance within each group over time. This step allowed for a deeper look into how students' speaking scores evolved from the beginning to the end of the study period. Table 9 below presents the results for the experimental group, which participated in AI-supported speaking practice using ChatGPT.

**Table 9**

*Paired Samples t-test findings for the speaking scores of the experimental group*

| Group | *N* | *M* | *SD* | *t* | *df* | *p* |
|---|---|---|---|---|---|---|
| Experimental pre-test | 10 | 64.40 | 15.85 | -3.252 | 9 | .010 |
| Experimental post-test | 10 | 73.10 | 15.68 | | | |

A paired samples t-test was conducted to examine the effect of ChatGPT-based speaking practice on students' speaking performance. The results indicate a statistically significant difference between the speaking scores before the treatment ($M$ = 64.40, $SD$ = 15.85) and after the treatment ($M$ = 73.10, $SD$ = 15.68); t(9) = –3.252, $p$ = .010. The corresponding effect size was large (Cohen's $d$ = 1.03), suggesting that the magnitude of the improvement was substantial. Based on these results, the null hypothesis that there is no difference between the means was rejected. It can be concluded that the use of ChatGPT had a significant effect on improving the speaking performance of the experimental group.

In addition to the experimental group, the control group also demonstrated a statistically significant improvement in their speaking scores over the study period. The mean score increased from 62.89 ($SD$ = 14.57) in the pre-test to 69.67 ($SD$ = 16.02) in the post-test. The paired samples $t$-test revealed a significant difference, $t(8) = –2.65$, $p$ = .029, with a large effect size (Cohen's $d$ = 0.89). This indicates that some level of speaking development occurred even without AI-supported intervention. Such improvement may be attributed to continued classroom instruction, natural language exposure, or test familiarity. However, the degree of improvement was less pronounced compared to the experimental group, suggesting that additional factors such as targeted AI practice may yield greater gains in speaking performance.

**Table 10**

*Paired samples t-test findings for the speaking scores of the control group*

| Group | *N* | *M* | *SD* | *t* | *df* | *p* |
|---|---|---|---|---|---|---|
| Control pre-test | 9 | 62.89 | 14.57 | -2.654 | 8 | .029 |
| Control post-test | 9 | 69.67 | 16.02 | | | |

An independent-samples *t*-test was conducted to compare the pre-test speaking scores of the experimental and control groups. The control group ($M = 62.89$, $SD = 14.57$) and the experimental group ($M = 64.40$, $SD = 15.85$) had very similar mean scores prior to the intervention. The results confirmed that there was no statistically significant difference between the groups, $t(17) = –0.22$, $p = .832$. The effect size was negligible (Cohen's $d = 0.10$), indicating virtually no baseline difference in speaking proficiency. These results suggest that both groups were comparable at the outset, which strengthens the validity of attributing any subsequent differences in post-test outcomes to the intervention rather than pre-existing disparities.

**Table 11**

*Independent samples t-test results for speaking pre-test scores*

| Group | *N* | *M* | *SD* | *t* | *df* | *p* |
|---|---|---|---|---|---|---|
| Experimental | 10 | 64.40 | 15.85 | -0.215 | 17 | .832 |
| Control | 9 | 62.89 | 14.57 | | | |

To compare the speaking scores of the two groups after the treatment, an independent-samples *t*-test was conducted. The experimental group ($M = 73.10$, $SD = 15.68$) achieved a slightly higher mean score than the control group ($M = 69.67$, $SD = 16.02$) in the post-test. However, the difference was not statistically significant, $t(17) = –0.47$, $p = .643$. The corresponding effect size was small (Cohen's $d = 0.22$), indicating only a modest trend favoring the experimental group. These findings suggest that while the ChatGPT-based practice may have contributed to a slight performance advantage, the effect was not strong enough to reach statistical significance. It is possible that the short treatment duration, small sample size, or external factors limited the observable group-level differences, highlighting the need for further

research with larger cohorts and extended interventions to better determine the potential of AI-supported speaking practice.

**Table 12**

*Independent samples t-test results for speaking post-test scores*

| Group | *N* | *M* | *SD* | *t* | *df* | *p* |
|---|---|---|---|---|---|---|
| Experimental | 10 | 73.10 | 15.68 | -0.472 | 17 | .643 |
| Control | 9 | 69.67 | 16.02 | | | |

### 4.2. Strenghts, Challenges, and Suggestions for Improvement Identified in Learner Diaries

The analysis of learner diaries revealed several strengths, challenges, and suggestions for improvement of ChatGPT that contributed to participants' speaking experiences. These findings derived from recurring themes in the diary entries. Table 13 next page presents an overview of the identified themes along with their frequency of occurrence in the diaries.

**Table 13**

*Overview of frequencies related to the use of ChatGPT identified in learner diaries*

| Category | Sub-Categories | Codes | *f* |
|---|---|---|---|
| Technological Strenghts | User Experience and Interface Functionality | Speech Interaction Features | 4 |
| | | Memory Features | 11 |
| | | Manual Mode | 2 |
| | | Interface Design | 4 |
| | | Enhanced Experience with Advanced Mode | 7 |
| | ChatGPT: Comprehension and Communication Ability | Speed and Efficiency | 28 |
| | | Understanding User Input | 4 |
| | | Handling Complex Input | 3 |
| | | Clear and Relatable Language | 7 |
| | ChatGPT: Knowledge and Depth | Personalized Responses | 8 |
| | | Innovation and Humor | 3 |
| | | Contextual Support and Example Creation | 21 |
| | | Broad Knowlegde Base | 24 |
| | | Generating and Adapting Ideas | 9 |

**Table 14 (Continued)**

*Overview of frequencies related to the use of ChatGPT identified in learner diaries*

| Category | Sub-Categories | Codes | *f* |
|---|---|---|---|
| Pedagogical Strenghts | Language Practice | General Speaking Practice | 13 |
| | | Convenience and Ease of Use | 3 |
| | | Grammar and Vocabulary Development | 8 |
| | | Listening Practice | 3 |
| | | Feedback | 20 |
| Social Strenghts | Fairness and Impartiality | Positive Cultural Comments | 1 |
| | | Real-Life Utility | 2 |
| | | Neutral and Objective Responses | 5 |
| | Emotional Connection | Perceived Warmth and Naturalness | 22 |
| | | Supportive Presence | 29 |
| | | Emotional Support and Motivation | 9 |
| | Total | | 250 |
| Technological Challenges | Communication Challenges with ChatGPT | Topic and Vocabulary Limitations | 6 |
| | | Motivation and Engagement Issues | 2 |

**Table 15 (Continued)**

*Overview of frequencies related to the use of ChatGPT identified in learner diaries*

| Category | Sub-Categories | Codes | *f* |
|---|---|---|---|
| | | Repetition and Irrelevant Responses | 5 |
| | | Speed and Complexity | 14 |
| | | Misinterpretation | 9 |
| | | Comprehension Issues | 27 |
| | | Interruption and Turn-Taking Problems | 36 |
| | Technical Problems | App Behavior and Functional Errors | 11 |
| | | Audio Interface Problems | 3 |
| | | Misbehavior in Speech Recognition | 4 |
| | | Crashes and Freezing | 9 |
| | | Lag and Response Delays | 15 |
| | | Connection Issues | 8 |
| | | Language-Switching Issues | 17 |
| Pedagogical Challenges | Inadequate Feedback | Educational Value | 3 |
| | | Repetition and Simplification | 6 |

**Table 16 (Continued)**

*Overview of frequencies related to the use of ChatGPT identified in learner diaries*

| Category | Sub-Categories | Codes | *f* |
|---|---|---|---|
| | | Error Detection Accuracy | 3 |
| | | Grammar Correction | 8 |
| | | Higher level words | 2 |
| Social Challenges | Unfulfilled Expectations for Interaction | Biased Cultural Comments | 1 |
| | | Lack of human-like emotional and personal responses | 21 |
| | | Overextension of Topics | 4 |
| | Total | | 214 |
| Suggestions for Improvement | Technological Suggestions | Improving Conversational Flow | 3 |
| | | Trusted Source Integration | 2 |
| | | Memory Retention and Continuity | 3 |
| | | Desire for Customization and Interaction Control | 43 |
| | | Enhanced Features Accessibility and Pricing Improvements | 19 |

**Table 17 (Continued)**

*Overview of frequencies related to the use of ChatGPT identified in learner diaries*

| Category | Sub-Categories | Codes | *f* |
|---|---|---|---|
| | Pedagogical Suggestions | Adaptive Language Teaching | 9 |
| | | Desire for Corrective Feedback | 26 |
| | | Topic-Based Conversation Initiation | 4 |
| | Social Suggestions | Desire for Authentic Human-Like Interaction | 42 |
| | Total | | 151 |

The analysis of learner diary data revealed a total of 250 diary entries categorized as strengths, 214 diary entries under challenges, and 151 diary entries classified as suggestions for improvement. The strengths section highlights ChatGPT's technological, pedagogical, and social affordances, with the most frequently mentioned benefits including supportive presence (f=29), speed and efficiency (f=28), perceived warmth and naturalness (f=22), broad knowledge base (f=24), and feedback provision (f=20). Challenges primarily emerged in technological limitations, particularly in turn-taking problems (f=36), comprehension issues (f=27), and response delays (f=15). Additionally, pedagogical and social concerns such as lack of human-like emotional responses (f=21) and inadequate feedback (f=21) were noted. Learners also provided various suggestions for improvement, with the desire for customization and interaction control (f=43) and authentic human-like interaction (f=42) being the most prominent recommendations. The following sections provide a detailed examination of these findings under the respective themes of strengths, challenges, and suggestions for improvement.

### 4.2.1 Technological Strengths of ChatGPT Identified in Learner Diaries

The analysis of reflective diary data revealed that English language learners identified several technological strengths of ChatGPT, emphasizing its effectiveness in facilitating language learning. The most frequently mentioned advantage was ChatGPT's speed and efficiency in responding to user input, which was highlighted as a key factor in maintaining

smooth and engaging conversations. Additionally, participants valued its broad knowledge base, enabling discussions on diverse topics with accurate and informative responses. Another frequently noted strength was ChatGPT's ability to provide contextual support and create relevant examples, aiding learners in understanding language use in different contexts. Learners also acknowledged its user experience and interface functionality, particularly its memory features, speech interaction capabilities, and enhanced experience with advanced mode. Moreover, ChatGPT's comprehension and communication ability was recognized, with participants appreciating its clear and relatable language, personalized responses, and innovation in delivering content with humor. Table 14 next page presents an overview of the technological strengths identified in the learner diaries.

**Table 18**

*Frequencies for the technological strengths of ChatGPT identified in the learner diaries*

| Category | Sub-category | Codes | *f* |
|---|---|---|---|
| Technological Strenghts | User Experience and Interface Functionality | Speech Interaction Features | 4 |
| | | Memory Features | 11 |
| | | Manual Mode | 2 |
| | | Interface Design | 4 |
| | | Enhanced Experience with Advanced Mode | 7 |
| | ChatGPT: Comprehension and Communication Ability | Speed and Efficiency | 28 |
| | | Understanding User Input | 4 |
| | | Handling Complex Input | 3 |
| | | Clear and Relatable Language | 7 |
| | ChatGPT: Knowledge and Depth | Personalized Responses | 8 |
| | | Innovation and Humor | 3 |
| | | Contextual Support and Example Creation | 21 |
| | | Broad Knowlegde Base | 24 |
| | | Generating and Adapting Ideas | 9 |

Hereafer, learner quotes are attributed using pseudonyms (e.g., Lara, Can, Elif) instead of real names to ensure anonymity. The speech interaction features provided learners with positive experiences by enhancing the conversational flow. Students highlighted ChatGPT's ability to wait patiently and manage turn-taking effectively, thus creating smoother interactions. Eren expressed appreciation noting, "when I think what am I gonna say, it waits for me" (Eren, W2). Similarly, Yusuf confirmed this strength by indicating "it usually waits you to stop" (Yusuf, W2). Kerem pointed out the utility of manual mode in mitigating interruptions: "Konuşma kesme durumu oluyor ama manual mod ile önüne geçilebiliyor" [Interruptions happen but can be prevented by manual mode] (Kerem, W4). Overall, these experiences show that ChatGPT significantly supports learners by allowing natural conversational pacing.

Learners consistently recognized the effectiveness of ChatGPT's memory capabilities. They valued its ability to recall past conversations, commands, and mistakes, thereby enhancing continuity in their interactions. Elif noted, "Geçmişteki hatalarımı harfi harfine hatırlıyor" [It remembers my past mistakes exactly] (Elif, W8), underscoring the bot's detailed memory feature. Similarly, Eren found the memory impressive, stating, "The bot remembers what did we talk last week, its memory is huge" (Eren, W2). Emir highlighted memory benefits as facilitating easier communication: "Hafızasında tutması iletişimi başlatıyor" [Keeping things in memory initiates communication] (Emir, W1). Yusuf also pointed out the bot's selective and useful memory storage: "when it thinks it's important, it saves it in its memory" (Yusuf, W1). Such reflections emphasize the crucial role memory features play in enhancing learner engagement and conversational coherence.

Though less frequently noted, manual mode was appreciated for offering users more precise control over interactions. Zeynep succinctly stated the benefit of having manual mode: "have manual mode" (Zeynep, W1), while Yusuf noted that "gpt konuşurken manuel kullanımı seçtiğimizde dinleme kısmına geçiyor" [when GPT is talking and manual mode is selected, it switches to listening] (Yusuf, W4). This suggests that manual mode significantly aids users who prefer a controlled and deliberate interaction process, reducing unexpected conversational interruptions.

Participants highlighted the clear and functional aspects of ChatGPT's interface design. Zeynep found the delineation between text and voice chat effective: "advanced sound modu, text chat, voice chat section" (Zeynep, W1). Eren complimented the quality of the audio component: "Ses yazılımı güzel konuşurken kendi sesini bozmuyor" [The sound software is good, it doesn't distort its own voice while speaking.] (Eren, W1). Emir appreciated the simplicity and efficiency of the interface: "Arayüzü basit" [Interface is simple] and noted

positively on connectivity: "internete bağlanarak kolay bilgi veriyor" [connects to the internet and provides information easily] (Emir, W4). These observations underline the user-friendly, efficient, and accessible interface provided by ChatGPT.

Learners frequently praised the advanced mode for significantly enriching their interaction experience with ChatGPT. Baran emphasized strong positive emotions, stating, "Gelişmiş konuşma çok sarıyor." [Advanced speech is very engaging.] (Baran, W6), and expressed improved feelings during advanced sessions, "Gelişmiş konuşmada kendimi daha iyi hissediyorum." [I feel better in advanced speech] (Baran, W6). Derya appreciated the realism and enjoyment in advanced mode interactions: "15 dk pro modundaki sesi çok gerçekçi" [The voice in the 15-minute pro mode is very realistic] (Derya, W3). Moreover, Derya expressed satisfaction regarding improvements implemented following feedback, highlighting responsiveness: "Daha önceki haftalarda konuşma özelliğinin bilgisayar versiyonuna da getirilmesini önermiştim ve gerçekten geldi" [I suggested earlier weeks for the speech feature to be available in the computer version, and it indeed came] (Derya, W8). Such insights suggest that advanced mode significantly elevates learner motivation, realism, and satisfaction in their interactions with ChatGPT.

Comprehension and communication sub-category encapsulates participants' perceptions of ChatGPT's ability to comprehend user input and respond fluently and appropriately during interactions. The sub-categories reflect how speed, input understanding, complexity management, and clarity of language shaped students’ speaking experiences with the AI chatbot.

Participants consistently appreciated the speed, responsiveness, and efficiency of ChatGPT, emphasizing how quickly it replied to their questions. Lara expressed satisfaction, saying that it is "not interrupting me during conversation" (Lara, W1) and frequently noted that it "responds quickly" (Lara, W1). Elif reinforced this point in multiple reflections, highlighting that "Sorularıma anında cevap veriyor" [it responds instantly to my questions] (Elif, W4) and "Sorularımı hızlı cevap veriyor" [it answers my questions quickly] (Elif, W6). Emir further illustrated efficiency by indicating "İnsanlar gibi düşünmek için beklemiyor seri cevap veriyor" [it doesn't pause to think like people, it responds swiftly] (Emir, W5). Baran emphasized the effectiveness of responses, describing them as "Güzel cevaplar veriyor" [It gives good responses] (Baran, W3) and "Güzel noktalara değiniyo" [It touches on good points] (Baran, W7). Similarly, Yusuf appreciated the thoughtful yet swift responses, noting "it thinks about the best answer and use it" (Yusuf, W1). Derya also found the speed appealing: "It can continue the conversation faster than an average human, so it's fun to have a conversation with it" (Derya,

W2). This overall speed and promptness contributed to increased conversational fluency and user satisfaction.

ChatGPT demonstrated a notable ability to comprehend learners, even when they made language mistakes. Elif affirmed this strength clearly, stating "Cümlelerimi anlıyor" [It understands my sentences] (Elif, W8). Mert also felt understood despite errors: "AI can understand me even if I make mistakes" (Mert, W1). Similarly, Derya highlighted that the chatbot effectively understood even incorrectly structured sentences: "it can understand the sentences that I've made incorrectly" (Derya, W1). These observations reflect ChatGPT's robustness in interpreting learner input accurately, enhancing learners' conversational confidence and comfort.

Learners noted ChatGPT's effectiveness at managing complex and extended input without losing track or clarity. Eren highlighted its capacity, noting, "Uzun cümleleri kolayca anlayabiliyor" [It easily understands long sentences] (Eren, W1). He further detailed ChatGPT's ability to sequentially manage longer interactions, observing, "Bazı anlarda cümlelerimin uzun olmasına rağmen sırayı karıştırmadan sırasıyla ve hızlı cevap verdi" [Although my sentences were long at times, it answered quickly and in sequence without confusion] (Eren, W7). Yusuf similarly observed how it efficiently handled intricate queries, remarking, "karmaşık olayların olduğu bir soru sorduğumda düşünüp 4 dakikalık bir cevap olsa da ayrı ayrı cevap verebiliyor" [when I asked a complicated question, even if the answer was four minutes long, it provided distinct answers separately] (Yusuf, W4). These insights highlight ChatGPT's advanced processing capabilities, crucial in supporting nuanced and sophisticated learner interactions.

Participants found ChatGPT's communication consistently clear, relatable, and easy to comprehend. Lara explicitly mentioned, "Açık ve anlaşılabilir kelimeler, cümleler kullanıyor" [It uses clear and understandable words and sentences] (Lara, W3). Eren confirmed the clarity, stating simply, "Konuşması anlaşılabilir, duyulabilir" [Its speech is understandable, audible] (Eren, W1). Baran emphasized the authenticity and relatability of the responses: "Gerçekçi cevaplar veriyor" [It provides realistic answers] (Baran, W5), and described the responses as "Cevapları Açıklayıcı" [Its answers are explanatory] (Baran, W5). Derya similarly endorsed the clarity of its communication, saying succinctly, "it speaks clear" (Derya, W1). This clarity significantly contributed to learners' ease of interaction, making conversations smooth and more enjoyable.

The broader category knowledge and depth explores the extent to which ChatGPT demonstrates a wide-ranging knowledge base, the ability to provide personalized and relevant

responses, and its adaptability to different conversational contexts. The sub-categories reflect different facets of ChatGPT's knowledge and its relevance to the participants' experiences.

Learners valued ChatGPT's ability to personalize interactions according to their interests and preferences. Lara highlighted how the chatbot engages by asking about personal interests: "ilgili olduğum kısımlar hakkında soru soruyor" [it asks questions about the parts I'm interested in] (Lara, W2) and "öznel olmasını istediğim zamanlarda öyle yanıtlar veriyor" [it gives subjective responses when I want it to] (Lara, W5). Zeynep praised the ability to shape the chatbot's persona, saying, "Kişiliğini istediğiniz gibi şekillendirebilmemiz" [we can shape its personality as we wish] (Zeynep, W5). Similarly, Eren noted how ChatGPT seamlessly adapted to his imaginative scenario: "Kendi arkadaşını yaratmasını istediğim zaman takılmadan ayak uyduruyor" [When I ask it to create its own friend, it adapts without hesitation] (Eren, W3). Derya found personalized interaction beneficial for practical advice, observing, "It can understand my interest quickly and can give advice about it" (Derya, W2). Moreover, Kerem experienced the chatbot acting "Akademik bir koç gibi" [like an academic coach] providing comprehensive support across multiple areas (Kerem, W4). These reflections underscore the value of personalization in making interactions more meaningful and engaging for learners.

Participants also noted ChatGPT's capacity for innovation and humor. Lara appreciated its playful nature through its ability to make jokes (Lara, W1). Zeynep valued its creative potential, noting, "Yeni fikirler üretebilmesi" [its ability to generate new ideas] (Zeynep, W8). Meanwhile, Yusuf acknowledged its technological integration, stating, "Teknolojiye çok bağlı" [It is highly connected to technology] (Yusuf, W6). These aspects contributed positively to a lively and enjoyable conversational experience, enriching the interaction beyond simple information exchange.

Students frequently commended ChatGPT's capability in providing relevant examples, detailed explanations, and contextual support. Lara regularly noted its explanatory clarity: "anlamadığım kısımları açıklayıcı bir şekilde anlatıyor" [it explains clearly the parts I don't understand] (Lara, W6). Zeynep emphasized depth, praising its ability to delve into specifics: "Detaylara inerek konuşması" [It speaks by going into details] (Zeynep, W3). Similarly, Yusuf observed how ChatGPT adopted context-specific roles effectively: "Kendini bir doktor gibi görüp öyle cevap vermesi isteyince bilgilere ekleme yapabiliyor" [When asked to answer as a doctor, it enriches the information accordingly] (Yusuf, W5). Derya praised its practical, scenario-based examples useful in daily interactions, citing it provides "Starbucksta karşılaşılabilecek muhtemel sohbet örnekleri" [probable conversational examples one might encounter at Starbucks] (Derya, W5). Kerem also praised its thoroughness, stating that it offers

"konu hakkında derinlemesine bilgi" [in-depth information on topics] (Kerem, W4). This strong contextual capability enhances learners' understanding, enabling a richer and more practical language-learning experience.

Participants consistently admired ChatGPT's expansive knowledge base and intellectual versatility. Lara described its comprehensive understanding: "konu hakkında derin bilgiler veriyor" [it provides deep information about the topic] (Lara, W6). Elif repeatedly recognized ChatGPT's broad familiarity, stating, "Her konuda bilgi sahibi" [knowledgeable on every subject] (Elif, W3, W4, W5, W8). Eren found ChatGPT impressive in its technological discussions and knowledge capacity, commenting, "Çok güçlü altyapısı var neredeyse her şeye hakim" [It has a very strong infrastructure and is knowledgeable about almost everything] (Eren, W1). Similarly, Yusuf admired its computational ability, noting it can perform advanced tasks such as "matematik işlemlerini hesaplayabiliyor" [can calculate mathematical operations like probability] (Yusuf, W4). Mert and Derya both praised its intellectual breadth, with Derya stating: "It has a powerful vocabulary," which even led to envy due to its extensive lexical resources (Derya, W2). This broad and deep informational background significantly enhances the conversational and educational value of interactions with the chatbot.

Lastly, learners recognized ChatGPT's capability to creatively initiate and steer conversations by generating and adapting ideas spontaneously. Zeynep praised its conversational agility, mentioning "konu açmakta iyi" [good at initiating topics] (Zeynep, W4). Eren observed a strong sense of realism and adaptability, stating it could describe holiday experiences as authentically as a person would: "sanki tatile gitmiş ve deneyimlemiş gibi samimi ve takılmadan anlatabildi" [described experiences authentically as if having been on holiday] (Eren, W4). Baran appreciated its initiative when lacking a topic, noting it "güzel bir konu seçiyo" [chooses a good topic] (Baran, W7). Derya highlighted its visual creativity, reflecting positively on its capability: "I can make pictures with the key words. It's good at it" (Derya, W1). Such dynamic conversational adaptability enhances user engagement by providing fresh, diverse, and relevant content continuously throughout interactions.

### 4.2.2 Pedagocial Strenghts of ChatGPT Identified in Learner Diaries

To explore the educational value of ChatGPT in enhancing EFL students' speaking skills, qualitative data were collected through weekly reflective diaries maintained by the participants in the experimental group. Thematic analysis of the entries revealed several pedagogical strengths of ChatGPT, particularly in relation to language practice and feedback. The emerging categories underscore the multifaceted role ChatGPT played not only in improving linguistic competencies such as grammar, vocabulary, and speaking fluency but also

in fostering a supportive and motivating learning environment. Table 15 shows the pedagogical strengths of ChatGPT identified in the learner diaries.

**Table 19**

*Frequencies for the pedagogical strengths of ChatGPT identified in the learner diaries*

| Category | Sub-Categories | Codes | *f* |
|---|---|---|---|
| Pedagogical Strenghts | Language Practice | General Speaking Practice | 13 |
| | | Convenience and Ease of Use | 3 |
| | | Grammar and Vocabulary Development | 8 |
| | | Listening Practice | 3 |
| | | Feedback | 20 |

General Speaking Practice comprises reflections that highlight ChatGPT's role in facilitating general speaking practice. Learners frequently noted that regular interaction with ChatGPT helped them to improve their fluency, pronunciation, and confidence in English speech.

Lara described a consistent sense of improvement: "İngilizce konuşmayı geliştiriyor" [It improves my English speaking] (Lara, W3), and later added, "konuşmamı ilerlettiğini düşünüyorum" [I think it has improved my speaking] (Lara, W4). She also acknowledged a holistic benefit: "konuşmamı ve dinlememi geliştiriyor" [It improves both my speaking and listening] (Lara, W4).

Elif stated the chatbot helped her "Help learn speak fluently" (Elif, W1), while Eren focused on pronunciation: "konuşma aksanımı geliştirmeye yardım ediyor" [It helps improve

my speaking accent] (Eren, W3). Kerem emphasized the engaging nature of ChatGPT's practice opportunities: "İngilizce konuşma, pratik yapmakta çok keyifli bir araç" [It's a very enjoyable tool for speaking practice] (Kerem, W4). These suggest that ChatGPT, through ongoing dialogue, functioned as a supportive conversational partner, reinforcing learners' confidence and oral fluency over time.

Participants also appreciated ChatGPT's convenience in facilitating English practice without the stress of real-time interaction with humans. Mert highlighted the simplicity of the experience: "I can easily answer" (Mert, W1), and later added, "its easy I had good experience" (Mert, W2). Kerem pointed out its accessibility: "You can practice English easily with this App" (Kerem, W1). This sub-category reflects how ease of use contributed to learner engagement by lowering affective barriers and making English practice more approachable.

A major linguistic benefit reported by students was support for grammar and vocabulary growth. Lara repeatedly emphasized this: "İngilizcemi geliştiriyor" [It improves my English] (Lara, W3, W4), and "ingilizceyi geliştirmeye yararlı" [It is useful for improving English] (Lara, W6). She also noted, "yeni kelime öğrenmeme yardımcı oluyor" [It helps me learn new words] (Lara, W7). Zeynep also expressed the benefit of targeted grammar use: "ingilizcede detaylı gramerleri kullanmasını istediğimde ona göre davranıp konuşmamıza uyarlayabiliyor" [When I ask it to use detailed grammar structures, it adapts to our conversation accordingly] (Zeynep, W7). Participants perceived ChatGPT as not only a conversational tool but also an adaptive tutor that scaffolds their learning of grammatical forms and vocabulary.

Although not the primary focus, some students reported improvements in listening comprehension. Lara shared: "konuşmamı ve dinlememi geliştiriyor" [It improves both my speaking and listening] (Lara, W4). Kerem valued the exposure to auditory input: "Dinleme pratiği sağlaması güzel" [Providing listening practice is nice] (Kerem, W5). While reported less frequently than other areas, these comments show that ChatGPT's voice interaction mode contributed positively to receptive listening skills.

Feedback was one of the most frequently mentioned strengths. Learners appreciated ChatGPT's capacity to provide immediate and clear corrections, suggestions, and clarification during or after speaking. Lara shared several examples: "ingilizcemi düzelt dediğimde düzelterek cümleyi düzelterek söylüyor" [When I ask it to correct my English, it repeats my sentence with corrections] (Lara, W5), and "konuşmalarımı düzeltiyor" [It corrects my speech] (Lara, W8). Zeynep noted its effectiveness in highlighting learning gaps: "Nerede ve hangi bilgimde hata olduğunu söylüyor" [It tells me where and what mistake I made] (Zeynep, W7). Eren described an emotionally supportive correction experience: "feedback vermekle kalmayıp

bütün hikayemi baştan okuyarak düzeltti ve duygusunu belli ederek feedback verdi" [It not only gave feedback but also reread my whole story and corrected it while expressing emotion] (Eren, W7). Derya appreciated that "it explains the words that I can't understand" (Derya, W1), and Kerem emphasized how feedback led to language retention: "Bilmediğin kelimeyi cümle içinde kullanıp öğrenmeni sağlıyor" [It uses unknown words in a sentence to help you learn] (Kerem, W1). This sub-category illustrates ChatGPT's pedagogical potential as an always-available interlocutor that provides personalized, formative feedback in a timely manner.

### 4.2.3 Social Strenghts of ChatGPT Identified in Learner Diaries

In addition to its linguistic affordances, ChatGPT was also perceived to offer significant social strengths throughout the students' reflective accounts. These strengths include the tool's ability to create emotionally engaging interactions, offer support and motivation, and maintain fairness and impartiality in communication. The diaries revealed that learners not only viewed ChatGPT as a language practice tool but also as a socially responsive interlocutor capable of providing emotional warmth and encouragement during conversations. This suggests that the human-like features of ChatGPT contributed positively to learners' affective experiences, fostering a sense of connection, comfort, and confidence. These qualities may enhance learner motivation and reduce the anxiety often associated with speaking practice in a foreign language. The distribution of codes under the main category Social Strengths is presented in Table 16.

**Table 20**

*Frequencies for the social strengths of ChatGPT identified in the learner diaries*

| Category | Sub-Categories | Codes | *f* |
|---|---|---|---|
| Social Strenghts | Fairness and Impartiality | Positive Cultural Comments | 1 |
| | | Real-Life Utility | 2 |
| | | Neutral and Objective Responses | 5 |
| | Emotional Connection | Perceived Warmth and Naturalness | 22 |
| | | Supportive Presence | 29 |
| | | Emotional Support and Motivation | 9 |

The sub-category Fairness and Impartiality captures students' perceptions of ChatGPT as a neutral, culturally aware, and practically useful conversational partner. While this category appeared less frequently than others, the reflections highlighted students' appreciation for the tool's balanced, respectful, and pragmatic responses—key traits for maintaining trust in language interaction and promoting intercultural understanding.

Learners noted ChatGPT's cultural sensitivity and tendency to provide uplifting perspectives even in the face of common stereotypes. Derya pointed out that "Bir ülke hakkında çok fazla kötü yorum olmasına rağmen o övüyor" [Despite the many negative comments about a country, it praised it] (Derya, W4). This suggests that learners recognized and valued ChatGPT's ability to present cultures in a non-judgmental and positive manner, thereby indirectly modeling intercultural respect during conversation. Participants also appreciated ChatGPT's practical application in real-world tasks. Emir commented that it is "plan yapmak için birebir" [perfect for planning] (Emir, W5), while Derya mentioned that "benim için gezi planı oluşturabiliyor" [it can create a travel plan for me] (Derya, W4). These reflections show that learners viewed ChatGPT not just as a conversation tool, but as a functional assistant capable of supporting their daily planning in English, which in turn likely enhanced their motivation to engage with the language in meaningful contexts.

Another sub-category emotional category emerged as many learners expressed that ChatGPT created a social presence that mimicked human interaction in warmth, tone, and emotional sensitivity. Lara stated that it "hızlı cevap veriyor, espri yapıp gülüyor" [responds quickly, makes jokes and laughs], adding, "arkadaşın gibi sohbet ediyor" [chats like a friend] (Lara, W2). She later described it as "insan gibi veya kendi hayat hikayesi varmış gibi konuşuyor" [talking like a human or as if it had its own life story] (Lara, W8). Elif frequently praised the bot's patient attitude, saying "Sorularıma bıkmadan cevap veriyor" [It answers my questions patiently] (Elif, W2, W5) and "Konuşurken daha rahat hissediyorum beni yargılayamaz" [I feel more relaxed while speaking; it can't judge me] (Elif, W5). Eren commented, "The bot can think like it is a human" (Eren, W2), and shared that their conversation about vacations felt lifelike: "sanki tatile gitmiş ve deneyimlemiş gibi" [as if it had actually gone on vacation and experienced it] (Eren, W4). Other students described its speech as "insancıl" [humane] (Eren, W6), "can speak like a human" and "very friendly" (Baran, W1), and even "duygusal olarak gelişmiş" [emotionally developed] (Emir, W8). Yusuf remarked that if asked hypothetical questions about being human, ChatGPT "says humanly things" and responds relevantly (Yusuf, W2), while Mert shared, "It can comprehend my speech" (Mert, W6).

The emotional tone of interaction was further reinforced by Derya: “Plus mode can feel like a real friend,” and “It’s interested about my hobbies” (Derya, W2). She also noted, “Bu haftaki konuşmamda daha insancıl konuşuyordu… modumu arttırdı” [This week it sounded more human; it helped boost my mood] (Derya, W8). Kerem described the experience as “Bir arkadaş gibi konuşabiliyor. Canın sıkılmıyor” [It can talk like a friend. You don’t get bored] (Kerem, W2).

Participants also emphasized the supportive consistency of ChatGPT’s availability and attitude. Elif particularly highlighted this across several weeks, stating “7/24 available,” “it doesn’t break my comfort zone,” and that it is “ideal for students who have social anxiety” (Elif, W1, W2, W3, W4, W5, W6, W7, W8). She appreciated how it “bıkmadan açıklama yapıyor” [explains without getting tired] and “tekrarlamaktan yorulmuyor” [doesn’t get tired of repeating things] (Elif, W3, W4). This emotional dependability provided a strong sense of psychological safety. Elif noted, “Konuşurken sosyal anksiyete sıkıntısı yaşamıyorum” [I don’t experience social anxiety when talking to it] (Elif, W6), and “konuşurken gerilmiyorum” [I don’t get nervous while speaking] (Elif, W8). The repeated emphasis on availability (“7/24 müsait”) across nearly all weeks underlines the importance of constant access for learners.

Lara also appreciated its clarity: “Açık ve anlaşılabilir kelimeler kullanıyor” [It uses clear and understandable words] (Lara, W3), while Emir added that ChatGPT “provides convenience for lonely people or easy access to information” (Emir, W3). Yusuf found value in its calmness: “It can talk so calmly, I think that’s good” (Yusuf, W1), and Derya noted how it agreed with her ideas (Derya, W4), further reinforcing a sense of emotional affirmation. As Kerem echoed: “Bir arkadaş gibi konuşabiliyor. Canın sıkılmıyor” [It can talk like a friend. You don’t get bored] (Kerem, W2).

In addition to emotional warmth and stability, learners described how ChatGPT’s feedback and tone could boost motivation. Baran noted, “ChatGPT bana güzel tavsiyeler verdi” [ChatGPT gave me nice advice] and “konuşmak sıkıcı ama konular bazen güzel olunca sarıyor” [sometimes the topics are boring, but when they’re good, the conversation gets engaging] (Baran, W7, W8). He also appreciated that it “ilk ve son haftamı değerlendirdi” [evaluated my first and last week] (Baran, W8).

Emir highlighted that the bot “yazdıklarıma göre ruh halimi tahmin ediyor ve buna göre tavsiye veriyor” [guesses my mood from what I write and gives advice accordingly] (Emir, W2), while Yusuf mentioned that after saying it was the last week, “bana güzel şeyler söyledi” [it said nice things to me] (Yusuf, W8). Mert simply stated, “It went great this week” (Mert, W4), and Derya appreciated its uplifting tone: “Motive ediyor” [It motivates me] (Derya, W5),

and more powerfully: “feedbackler sayesinde IELTS’ten 7 alabileceğime inanmamı sağladı” [its feedback made me believe I could get a 7 from IELTS] (Derya, W8).

### 4.2.4 Technological Challenges of ChatGPT Identified in Learner Diaries

The analysis of learner diaries revealed that, while ChatGPT was generally perceived as a supportive tool for developing speaking skills, participants frequently encountered a variety of technological challenges during their interactions. These issues, grouped under the main category of Technological Challenges, were classified into two overarching sub-categories: Communication Challenges with ChatGPT and Technical Problems (see Table 17 next page). The sub-category of Communication Challenges with ChatGPT includes issues related to turn-taking disruptions, misunderstanding of input, and ChatGPT’s occasional failure to produce relevant or comprehensible responses. Meanwhile, the Technical Problems sub-category encapsulates the functionality-based limitations of the application, such as crashing, lagging, or failures in speech recognition. The presence of these challenges suggests that despite ChatGPT’s potential to support learners in practicing English speaking, technological shortcomings at both communicative and system levels may hinder learners' overall experience and reduce the quality of interaction. The following paragraphs will provide a detailed examination of these difficulties as reported in the weekly reflections.

**Table 21**

*Frequencies for the technological challenges of ChatGPT identified in the learner diaries*

| Category | Sub-Categories | Codes | *f* |
|---|---|---|---|
| Technological Challenges | Communication Challenges with ChatGPT | Topic Limitations | 6 |
| | | Motivation and Engagement Issues | 2 |
| | | Repetition and Irrelevant Responses | 5 |
| | | Speed and Complexity | 14 |
| | | Misinterpretation | 9 |
| | | Comprehension Issues | 27 |
| | | Interruption and Turn-Taking Problems | 36 |
| | Technical Problems | App Behavior and Functional Errors | 11 |
| | | Audio Interface Problems | 3 |
| | | Misbehavior in Speech Recognition | 4 |
| | | Crashes and Freezing | 9 |
| | | Lag and Response Delays | 15 |
| | | Connection Issues | 8 |
| | | Language-Switching Issues | 17 |

The sub-category communication challenges encompasses a range of interactional problems that learners encountered during their engagement with ChatGPT. These communication-related challenges often disrupted the flow of conversation, reduced learners' engagement, and occasionally caused frustration, thereby hindering the pedagogical value of ChatGPT for speaking practice.

Some students expressed difficulty in finding engaging or suitable topics to sustain a conversation with ChatGPT, especially when personal or culturally specific vocabulary was involved. Mert mentioned, "I can't find a topic to speak" (Mert, W1), indicating an early struggle in initiating dialogue. Kerem frequently emphasized that ChatGPT had difficulty processing proper nouns and Turkish words: "Türkçe özel isimleri anlamadığı zamanlar oluyor" [Sometimes it doesn't understand Turkish proper nouns] (Kerem, W1), "Özel isimleri anlamıyor" [It doesn't understand proper names] (Kerem, W2), and "Oyun isimlerini anlamıyor" [It doesn't understand game names] (Kerem, W2). Furthermore, he added that the repetitive prompts like "How's it going?" based on previous weeks created a feeling of stagnation: "Her hafta önceki hafta konuştuğumuz şeyler hakkında soru soruyor... konuşmak zorlaşıyor" [Every week it asks questions about the last week... which makes speaking harder] (Kerem, W3).

Participants also noted a lack of emotional support or responsiveness from ChatGPT, which affected their motivation. Elif observed, "motive edici tavrı eksik" [It lacks a motivating attitude] (Elif, W8), while Derya found its generic behavior frustrating: "When I force it to hear its hobbies it just says my hobbies, which is annoying" (Derya, W2). These perceptions suggest that emotionally distant or repetitive responses can reduce learner motivation, especially in sustained language interaction.

Several learners reported issues with repetitive or off-topic answers, which disrupted their attempts to maintain coherent dialogue. Elif stated, "Tekrar ediyor cümleleri" [It repeats sentences] (Elif, W2), while Eren pointed out, "geleceğe yönelik sorular sorunca bazen konuyu saptayabiliyor" [It sometimes strays from the topic when asked future-oriented questions] (Eren, W6). Yusuf noted that "Kısa cevap ver diyince vermiyor" [Even when I ask for short replies, it doesn't give them] (Yusuf, W5), and Mert simply said, "it suddenly changes the topic" (Mert, W2). These breakdowns signal a lack of discourse coherence and responsiveness that is crucial for meaningful oral communication.

Speed was a consistent challenge across participants. Elif remarked twice that "Hızlı konuşuyor" [It speaks fast] (Elif, W7 & W8), while Eren added, "Çok akıcı ve hızlı konuştuğu zamanlar oluyor, anlamakta zorlanıyorum" [Sometimes it speaks so fluently and fast that I

struggle to understand] (Eren, W4). Similarly, Baran explained, “Sometimes I can’t understand ChatGPT’s accent” and “because of its speaking speed” (Baran, W1), and even repeated “çok hızlı konuşuyor” [It speaks too fast] in later weeks (Baran, W3). Mert also echoed this issue: “AI don’t let me think” (Mert, W3). Derya added that even the slower-paced speech sometimes felt fast due to the length and complexity of the sentences: “Yavaş konuşması da hızlı gelebiliyor ve cümleleri uzun olduğundan takip etmek zorlaşıyor” [Even its slow speech can feel fast, and it's hard to follow because the sentences are long] (Derya, W6).

Many learners reported that ChatGPT misinterpreted their speech or responded inaccurately. Lara explained, “misunderstood some words while I’m speaking” (Lara, W1), and “konuşmam sırasında anlaşılmayan kelimeleri kendisi yapıyor—çoğunlukla da yanlış yapıyor” [It completes my words when it doesn’t understand—and mostly does it wrong] (Lara, W8). Similarly, Eren noted, “İngilizcedeki konuşma arası kullanılan well, you know gibi şeyleri bazen yanlış anlıyor” [It sometimes misunderstands discourse markers like 'well', 'you know'] (Eren, W1), and Emir stated, “bazı kelimeleri anlamıyor ve yanlış cevaplar veriyor” [It doesn’t understand some words and gives wrong answers] (Emir, W1).

A major concern across diaries was mutual understanding—both learners struggling to understand ChatGPT and vice versa. Lara frequently expressed that it misinterpreted her speech: “konuşmamı anlamıyor kendisi cümlelerimi yuvarlıyor” [It doesn’t understand me and smooths over my sentences] (Lara, W7), and Elif admitted, “I can’t understand what it says” (Elif, W1). Baran noted, “ne dediğimi anlamıyor” [It doesn’t understand what I say] (Baran, W4), and Mert remarked, “it doesn’t understand me sometimes” (Mert, W4). These comprehension issues represent a significant challenge for effective speaking practice, as they disrupt turn-taking, clarity, and confidence.

The most frequently reported communication challenge was ChatGPT's tendency to interrupt or cut off students during speech. Lara reported, “it interrupts my conversation” (Lara, W1), while Zeynep noted, “interrupting your speech when you pause” (Zeynep, W1) and “Konuşma esnasında boşluk olduğu an atlaması” [It jumps in when there’s a pause in conversation] (Zeynep, W2). Elif consistently recorded the issue for multiple weeks: “sözümü kesiyor” [It cuts me off] (Elif, W1–W7). Eren reflected that it “konuşma aralarında düşünmeme zaman tanımıyor” [does not give me time to think during conversation] (Eren, W1), and Mert stated plainly, “AI bot breaks my speech” (Mert, W3). These accounts illustrate that ChatGPT's turn-taking mechanics often disrupt learner speech, reducing opportunities for fluency development and increasing anxiety during production.

The sub-category technical problems includes the functional and system-based issues experienced by learners while using ChatGPT for speaking practice. These issues, ranging from crashes and lagging to speech recognition errors and unexpected language switching, disrupted the continuity and quality of conversational interaction. Despite ChatGPT's overall usefulness, these recurring technical failures impacted learners' engagement and satisfaction negatively.

Several learners experienced unpredictable or dysfunctional behavior from the app interface, which affected their ability to manage conversations. Lara noted, “it listens even if I don’t press the volume button while talking” (Lara, W1), indicating a problem with control over speech activation. Zeynep remarked, “Manueli kullanmama rağmen söze atlıyor” [Even when I use manual mode, it jumps in] (Zeynep, W3), and added that the app was unresponsive at times: “Butona basıldığında hemen aktif olmaması” [It doesn't activate right away when I press the button] (Zeynep, W7). Emir mentioned broader usability issues: “kullanış sorunları çıkabiliyor” [It can have usability problems] (Emir, W4), while Yusuf stated, “If you exit the app and rejoin all of the saves reset” (Yusuf, W1). These malfunctions suggest that inconsistent interface responsiveness and control over interactions can break the conversational rhythm, which is vital for speaking fluency.

Eren observed that ChatGPT’s voice output is sometimes fluctuated: “bot konuşurken sesi kısılıyor bazen” [the bot’s voice lowers while speaking sometimes] (Eren, W6) and “Ses arayüzü çok zayıf” [Its audio interface is very weak] (Eren, W8). These sound-related issues compromised the clarity and audibility of ChatGPT’s responses, which in turn hindered listening comprehension and interactive flow.

Speech recognition errors were another major concern. Lara highlighted that “Konuşurken arkadan ufak bir ses geldiğinde direkt dinlemeye geçiyor” [When a small background sound is heard, it starts listening] (Lara, W4). Eren echoed this, stating “Arka plandaki sesleri duyup ona göre saçmalayabiliyor” [It picks up background noise and responds nonsensically] (Eren, W1). These issues means a sort of sensitivity to environmental noise which leads to misrecognition and failed comprehension during speaking tasks.

Crashes, freezing, and malfunctioning during interaction were commonly reported. Lara described prolonged pauses and system unresponsiveness: “çoğu zaman donuyor” [It freezes most of the time] (Lara, W7). Yusuf noted, “When you ask something fast, it crashes” (Yusuf, W2), and added later, “Uygulamayı açtığımda manuel control etmek istediğimde çöküyor” [When I try to manually control it after opening the app, it crashes] (Yusuf, W4). Emir observed that “standart yazı ve konuşma bitebiliyor” [standard voice or text conversation can suddenly end] (Emir, W3). These crashes interrupted speaking turns, leading to frustration and

diminished motivation for sustained use. Many learners were frustrated by ChatGPT's delayed responses, which disrupted the flow of dialogue. Lara frequently reported long pauses: "çok uzun süre bekliyor" [It waits too long] (Lara, W7), while Baran stated, "cevap vermek için iki saat bekliyor" [It takes forever to answer] (Baran, W4). Yusuf explained, "if you ask him the same question again and again it will start thinking like an hour" (Yusuf, W2), and Mert reported, "It doesn't speak when I ask something" (Mert, W6). These delays discouraged spontaneous speaking and created unnatural conversation dynamics, which are particularly detrimental in oral practice contexts.

Unstable internet connectivity was also a recurring complaint, even when students had working connections. Lara said, "internetim olmasına rağmen hata veriyor" [Even though I have internet, it gives an error] (Lara, W5), and Elif confirmed, "When wifi disconnected, it is disconnected too" (Elif, W1). Derya noted repeated disconnections: "bağlantı ile ilgili sorunlar yaşadım, bu hala çözülmedi" [I had problems with connection, this still isn't fixed] (Derya, W8). Kerem added that switching tabs on the phone or computer caused ChatGPT to mistakenly end the conversation (Kerem, W4). These issues negatively affected continuity, making it difficult for learners to complete tasks or maintain focus.

Unexpected language switching—particularly between English and Turkish—was a widespread issue. Lara reported, "Yanlışlıkla Türkçe kelime kullandığımda kendini direkt Türkçeye çeviriyor" [When I accidentally use a Turkish word, it immediately switches to Turkish] (Lara, W3), and "konuşurken araya farklı dilden ses gelince dilleri karıştırıyor" [It mixes languages when another language is heard in the background] (Lara, W6). Elif observed that ChatGPT started speaking Turkish even without a clear reason: "Türkçe konuşmaya başlıyor" [It starts speaking Turkish] (Elif, W5). Emir noted: "Sadece bir Türkçe kelime bile duysa anında Türkçeye dönüyor" [Even one Turkish word makes it switch to Turkish instantly] (Emir, W1). These automatic switches confused learners, interrupted the English-only practice, and reduced the effectiveness of target-language immersion.

### 4.2.5 Pedagogical Challenges of ChatGPT Identified in Learner Diaries

In addition to technological concerns, participants also identified pedagogical limitations in their weekly reflections regarding their interactions with ChatGPT. These challenges, grouped under the category of Pedagogical Challenges, primarily relate to the perceived inadequacy of feedback provided by the AI tool. Although ChatGPT was frequently used as a means to practice speaking and receive immediate responses, learners noted several deficiencies in its instructional value. The sub-category Inadequate Feedback encompasses

issues such as superficial or repetitive correction, lack of detailed explanations, inaccuracies in error detection, and limited support for higher-level language development (See Table 18). These findings underscore the necessity for pedagogically-informed integration of AI tools in language learning environments, particularly when learners rely on such tools for autonomous skill development. The following section presents a detailed analysis of these concerns based on student diary data.

**Table 22**

*Frequencies for the pedagogical challenges of ChatGPT Identified in Learner Diaries*

| Category | Sub-Categories | Codes | *f* |
|---|---|---|---|
| Pedagogical Challenges | Inadequate Feedback | Educational Value | 3 |
| | | Repetition and Simplification | 6 |
| | | Error Detection Accuracy | 3 |
| | | Grammar Correction | 8 |
| | | Higher level words | 2 |

Elif emphasized several times that ChatGPT lacked an instructional role in their speaking practice: “öğretici bir görev üstlenmiyor” [It doesn’t take on a teaching role] (Elif, W3). Even when ChatGPT attempted to correct her, she found the explanations insufficient: “hatalarımı düzeltirken açıklaması tatmin edici değil” [Its explanation while correcting my mistakes is not satisfactory] (Elif, W8). Furthermore, she noted that when she didn’t understand a point, ChatGPT didn’t pause to clarify: “dediklerini anlamadığımda durup detaylı açıklama yapmıyor” [When I don’t understand, it doesn’t stop and explain in detail] (Elif, W8). These insights reveal a need for more scaffolded, intentional teaching behavior, especially when learners are seeking guided support.

Participants reported that ChatGPT tended to repeat content or speak more slowly when asked to simplify, instead of adapting the language to their level. Lara noted, “When I asked her to explain it in a simpler way, it repeated the sentence” (Lara, W1), and in the following week, she wrote, “daha basit söylemesini istediğim yerlerde cümlesini tekrar ediyor ya da daha yavaş söylüyor” [When I ask it to say something more simply, it either repeats the sentence or

just says it slower] (Lara, W2). Kerem also experienced this issue: “Cümleleri basitçe tekrar etmesini isteyince aynısını yavaş sesle söylüyor” [When I ask it to simplify, it just says the same thing slower] (Kerem, W3). Zeynep mentioned a similar experience where the bot kept repeating both correct and incorrect forms instead of providing clarification: “Yanlışlarım varsa düzelt dememe rağmen doğruları da tekrar söylemesi” [Even though I asked it to correct my mistakes, it repeats both correct and incorrect phrases] (Zeynep, W6). These patterns reflect a superficial approach to simplification, which can hinder learners’ understanding rather than enhance it.

Some participants expressed concern that ChatGPT failed to identify their language errors accurately. Elif wrote, “Hatalarımı düzeltmiyor” [It doesn’t correct my mistakes] (Elif, W2) and later added, “hatalarımı tespiti eksik” [Its error detection is inadequate] (Elif, W8). Kerem also pointed out the risk of fossilization, explaining, “Hatalarımı düzeltmediği zamanlar olduğunu fark ettim... yanlış öğrendiğim bir şeyi doğru kabul ederek devam edebilirim” [I realized there are times it doesn’t correct my mistakes… I might keep using something wrong thinking it’s correct] (Kerem, W5). These observations reveal the importance of accurate and timely feedback in preventing persistent errors, which ChatGPT sometimes failed to provide.

The most frequently mentioned pedagogical challenge was the lack of grammar correction. Elif repeatedly stated across several weeks that “Gramer hatamı düzeltmiyor” [It doesn’t correct my grammar mistakes] (Elif, W4–W8), showing that this issue persisted over time. Eren also mentioned, “Bazen kötü gramer kullandığım zaman beni düzeltmeyi unutuyor” [Sometimes when I use bad grammar, it forgets to correct me] (Eren, W3), and Kerem observed that the chatbot prioritizes understanding over correction: “çoğu zaman hatalı cümle kurduğunda düzeltmeye değil anlamaya çalışıyor” [Most of the time, it tries to understand instead of correcting my faulty sentence] (Kerem, W3). These insights suggest that while ChatGPT fosters communication, it may not promote grammatical accuracy unless explicitly prompted.

Lara pointed out that ChatGPT sometimes responded with vocabulary beyond her proficiency level: “It used C2 words even though it knew I was at B1 level” (Lara, W1), and later added, “Seviyem düşük dememe rağmen zor kavramlar kullanıyor” [Even though I told it my level was low, it still uses difficult concepts] (Lara, W3). This inconsistency in level-appropriate language output may discourage learners or lead to misunderstandings, particularly when scaffolded input is essential for learning.

### 4.2.6 Social Challenges of ChatGPT Identified in Learner Diaries

In addition to technological and pedagogical concerns, students also highlighted a set of social challenges in their weekly reflections, particularly regarding the interactional limitations of ChatGPT. These challenges were grouped under the main category Social Challenges, with a focus on the sub-category Lack of Natural Interaction. Many participants reported that ChatGPT often failed to meet their expectations for human-like communication. Complaints included emotionally flat or overly objective responses, repetition or overextension of certain topics, and limited ability to role-play or connect personally. Some students also pointed out inconsistencies in how ChatGPT handled tone and personality, while one participant mentioned cultural bias. These reflections suggest that ChatGPT, while linguistically competent, lacks the social responsiveness and contextual sensitivity needed to fully satisfy learners seeking authentic conversational experiences. Table 19 shows the codes and frequencies for the social challenges of ChatGPT as identified in the learner diaries.

**Table 23**

*Frequencies for the social challenges of ChatGPT Identified in Learner Diaries*

| Category | Sub-Categories | Codes | *f* |
|---|---|---|---|
| Social Challenges | Lack of Natural Interaction | Unfulfilled Expectations for Interaction | 39 |
| | | Biased Cultural Comments | 1 |
| | | Lack of human-like emotional and personal responses | 21 |
| | | Overextension of Topics | 4 |

This table encapsulates learners' social dissatisfaction with ChatGPT's interactional performance, especially its inability to simulate human-like behavior, personalize conversations, or manage topics in a dynamic and engaging way. Students frequently reported that while ChatGPT could respond fluently, its robotic tone, lack of emotions, surface-level responses, and repetitive conversation patterns failed to meet their expectations for authentic

dialogue. This disconnect led to boredom, frustration, and a diminished sense of engagement over time.

A prominent issue voiced by learners was the emotional and social detachment they felt when interacting with ChatGPT. Lara noted, “İnsan olmadığını söylüyor tüm cevapları böyle veriyor” [It keeps saying it’s not human in every response] (Lara, W3), and expressed frustration with its refusal to adopt a human role even during roleplay: “okula gidiyormuş gibi davran dediğimde bile ‘eğer gitseydim’ diye cümleye başlıyor” [Even when I asked it to pretend going to school, it starts with ‘if I went’] (Lara, W3). Similarly, Zeynep stated, “Duygu yok ve insan olmadığı belli oluyor” [It has no feelings and it’s clear that it’s not human] (Zeynep, W3). This sentiment was echoed by Elif, who admitted, “I’m bored to chat with it” (Elif, W1) and added that its replies were “yüzeysel” [superficial] (Elif, W3).

Learners often found ChatGPT repetitive or inflexible. Eren shared, “Sometimes it talks about the same things” (Eren, W2), while Baran remarked, “Standart konuşma sıkıcı oluyo” [Standard conversation gets boring] (Baran, W6). Emir noticed inconsistent behavior in the bot’s identity: “kimi zaman ben bir robotum diyor, kimi zaman düşünce belirtiyor” [Sometimes it says it's a robot, other times it shares opinions] (Emir, W2). These inconsistencies made it difficult for learners to build trust or maintain interest in the conversations.

Additionally, several students, including Derya, were frustrated by ChatGPT’s limited creative input or over-objectivity: “Her konu hakkında bir fikir sorduğumda çok objektif bir cevap yazıyor” [When I ask for opinions, it gives overly objective answers] (Derya, W8), and “Its voice is monotone so it can be boring sometimes” (Derya, W1). Kerem added that ChatGPT failed to engage with his curiosity: “Konuyu değiştirmesini istediğimde bana soruyor” [When I ask it to change the topic, it asks me instead] (Kerem, W3). These reflections underscore the participants’ unmet need for variety, emotional nuance, and personalized conversation.

Only one student explicitly raised a concern about cultural bias. Eren commented, “kültürler hakkında yorumlarını taraflı yapıyor” [It makes biased comments about cultures] (Eren, W4). Although rare, this point signals the importance of ensuring cultural neutrality and appropriateness in AI interactions, especially in diverse language learning environments.

Several students were discouraged by ChatGPT’s inability to replicate natural human empathy, tone, and creativity. Zeynep noted, “bir insan gibi fikirlere ve duygulara sahip değil” [It doesn't have thoughts and emotions like a person] (Zeynep, W4), and later said, “Konuşma tekdüze olduğundan sıkıcılaşması” [The monotony of conversation makes it boring] (Zeynep, W6). Elif also felt disconnected: “Bir robotla konuşuyor hissi veriyor, yorum yapmıyor” [It feels like talking to a robot—it doesn’t comment] (Elif, W6), and said, “Cevapları çok sıradan”

[Its answers are too ordinary] (Elif, W7). Eren shared a failed attempt at personalization: “İbrahim isimli bir arkadaşım ol dedim olmadı” [I asked it to be my friend named İbrahim, but it didn’t work] (Eren, W7). Emir experienced similar frustration: “istediğim kişiliğe bürünmüyor” [It doesn’t take on the personality I want] (Emir, W4), and “Şaka anlamıyor” [It doesn’t understand jokes] (Emir, W6). These examples reveal that learners not only desired a learning assistant, but also expected a conversational partner who could embody personality and emotional nuance—expectations that were not met.

Lastly, some learners grew frustrated with ChatGPT’s insistence on staying with a single topic too long or over-explaining. Derya shared, “It wants to continue to speak about the topic for a long time and I just get bored” (Derya, W2), and “Konuyla alakalı çok fazla soru soruyor” [It asks too many questions about one topic] (Derya, W3). Kerem expressed similar fatigue: “Belirli bir konuda çok detaya iniyor” [It goes into too much detail about a specific topic] (Kerem, W2). These interactions highlight the need for better pacing and topic management in AI-assisted speaking practice.

### 4.2.7 Technological Suggestions of ChatGPT Identified in Learner Diaries

While learners identified various challenges in their interactions with ChatGPT, they also offered numerous constructive suggestions for enhancing the tool’s technological capabilities. These suggestions, grouped under the category Suggestions for Improvement, focus on increasing the platform’s usability, responsiveness, and personalization. Students emphasized a strong desire for greater control over the interaction process, improved memory retention across sessions, better conversational flow, and access to more advanced features. A few participants also mentioned the need for the integration of reliable sources to improve content quality. These insights reflect learners’ evolving expectations of AI as an educational partner, not just a language tool. Table 20 next page shows the sub-categories and frequency of technological suggestions identified in learner diaries.

**Table 24**

*Frequencies for the technological suggestions of ChatGPT Identified in Learner Diaries*

| Category | Sub-Categories | Codes | *f* |
|---|---|---|---|
| Suggestions for Improvement | Technological Suggestions | Improving Conversational Flow | 3 |
| | | Trusted Source Integration | 2 |
| | | Memory Retention and Continuity | 3 |
| | | Desire for Customization and Interaction Control | 43 |
| | | Enhanced Features Accessibility and Pricing Improvements | 19 |

Several participants highlighted the need for ChatGPT to manage conversations more naturally. Lara suggested that instead of misinterpreting unfamiliar words, the bot could "anlamadığı kelimeleri farklı yazmak yerine sorabilir" [ask for clarification instead of replacing misunderstood words] (Lara, W8). Eren proposed that ChatGPT should express emotions without sudden tonal shifts, noting, "Duygularını belirtirken bir anda bağırmamalı" [It shouldn't suddenly yell when showing emotion] (Eren, W7). Mert, who experienced several interruptions and misunderstandings, simply suggested: "Fix the challenges" (Mert, W5), summarizing a desire for a more coherent, responsive system. Moreover, A few learners voiced concerns over the credibility and neutrality of ChatGPT's responses. Eren emphasized, "Objektif düşünmesi ve yorumlarını objektif yapması önemli" [It's important that it thinks and responds objectively] (Eren, W4), while Emir pointed out the limited range of source material: "sadece belli başlı sitelerden alınmış bilgiler" [It only draws from a handful of sites] (Emir, W2). These comments suggest learners expect ChatGPT to not only be technically accurate but also balanced in its outputs.

Students showed a strong interest in continuity across sessions. Eren stated, “Ona öğrettiğimiz şeyleri uygulamayı aç kapa yaptığımızda unutmamalı” [It shouldn’t forget what we’ve taught it after restarting the app] (Eren, W1). Similarly, Yusuf recommended, “There should be a save button to save the conversation” (Yusuf, W1), while Derya added that ChatGPT should “Kaydettiği belleği kullanıp soru oluşturabilmeli” [use saved memory to generate questions] (Derya, W3). These suggestions point to a desire for longitudinal learning and a personalized experience that evolves over time.

Desire for customization and interaction control was by far the most frequently mentioned category, reflecting a strong learner demand for personalized, adjustable, and user-controlled interaction. Lara repeatedly suggested options for correcting misunderstandings manually or having the bot confirm unclear parts (Lara, W5–W7). Zeynep proposed features like a speech rate setting and a “signal that the conversation is over” (Zeynep, W1–W2). Elif expressed enthusiasm for subtitle integration—stating “altyazı ile konuşabilir” [it can speak with subtitles] across multiple weeks (Elif, W2–W8)—and repeatedly asked for simpler sentences and slower speech.

Eren recommended improving the mobile interface, saying, “Uygulamanın mobil arayüzü genişletilebilir” [The mobile interface could be expanded] (Eren, W5), and “arayüz mutlaka geliştirilmeli” [The interface must definitely be improved] (Eren, W8). Derya wanted more fun and expressive interaction, suggesting that ChatGPT “should be able to send gifs and make a fun atmosphere” (Derya, W2) and allow personality customization (W4). Kerem proposed several interaction enhancements, including a confirmation button when speech is misunderstood and a button to cut off the AI’s reply (Kerem, W2). These comments consistently reflect learners’ need for more agency and flexibility during interaction.

Lastly, many participants suggested improving access to advanced features, particularly for free users. Lara and Eren both advocated for extending the usage time of premium voice features (Lara, W1–W2; Eren, W1–W2), with Eren stating, “Gelişmiş motor için 15 dakika kısa bir süre daha uzun olmalı” [15 minutes for the advanced engine is too short] (Eren, W1). Baran repeated this sentiment multiple times across weeks, stressing: “Gelişmiş konuşma süresi artırılmalı” [The duration of advanced speech should be increased] (Baran, W6–W7). These suggestions reflect not only practical needs but also a sense of fairness and accessibility in using AI tools for learning.

### 4.2.8 Pedagogical Suggestions of ChatGPT Identified in Learner Diaries

In addition to technological improvements, learners also expressed pedagogical expectations for how ChatGPT could function as a more effective language learning partner. These reflections, categorized under Pedagogical Suggestions, reveal a desire for more intentional and structured educational support from the tool. Participants emphasized the importance of receiving timely and accurate corrective feedback, the need for language input tailored to their proficiency level, and the value of topic-driven interactions that align with their personal interests and classroom learning objectives. These pedagogical insights suggest that learners do not simply view ChatGPT as a tool for practice, but as a potential instructor-like presence that should facilitate more meaningful language development. Table 21 presents the codes and frequencies for the pedagogical suggestions recorded in the learner diaries.

**Table 25**

*Frequencies for the pedagogical suggestions of ChatGPT Identified in Learner Diaries*

| Category | Sub-Categories | Codes | *f* |
|---|---|---|---|
| | Pedagogical Suggestions | Adaptive Language Teaching | 9 |
| | | Desire for Corrective Feedback | 26 |
| | | Topic-Based Conversation Initiation | 4 |

Learners emphasized the need for ChatGPT to offer tailored language support that responds to their level and learning goals. Elif, in particular, proposed several adjustments to enhance pedagogical relevance, such as "Cümlenin yerine alternative örnekler vermeli" [It should offer alternative sentence examples] (Elif, W4) and “Alternatif kelimeleri de söyleyebilmeli” [It should suggest alternative words] (Elif, W8). She also pointed out that the AI could help her finish sentences and explain grammar in more detail (Elif, W8).

Eren offered a similar view, stating, “Konuşma arasında İngilizcede seviyeme göre biraz yukarı kelimeler kullanıp öğretebilir” [It could use slightly more advanced words than my level to help me learn] (Eren, W1). Meanwhile, Kerem suggested designing scenarios for practicing

specific grammar structures: “Belirli roller olabilir... örneğin belirli bir tense ile ilgili” [There could be specific roles to practice, for example, related to a certain tense] (Kerem, W6). These comments collectively reflect learners’ desire for adaptive scaffolding and incremental language input during AI interaction.

One of the most recurring themes in the diaries was the learners’ strong demand for corrective feedback. Lara proposed the idea of an “English development department” within the app and added, “can try to correct the words I use incorrectly” (Lara, W1). She also wished for oral correction capabilities, such as “konuşma sonunda akıcılık, kelime dağarcığı fln analizi de olabilir” [there could be an analysis of fluency and vocabulary at the end of the conversation] (Lara, W8).

Elif’s diary was especially rich in feedback-related requests. She frequently asked ChatGPT to correct grammar errors (e.g., “Gramer hatamı düzeltebilir” – Elif, W4) and help complete her sentences (e.g., “Cümlelerimi tamamlamama yardım etmeli” – Elif, W8). Her entries across weeks also reflected a growing awareness of how ChatGPT could function like a teacher, such as: “cümlelerde anlamadığım yerleri öğretmen gibi açıklayabilir” [It can explain the parts I don’t understand like a teacher] (Elif, W3) and “Yanlışımı anında tespit edebilir” [It can detect my mistake instantly] (Elif, W7).

Kerem suggested a written feedback option where incorrect utterances are underlined and corrected visually: “Konuştuğun yazılı olarak ve hatalı yerler çizilip altına doğrusu yazılsa iyi olur” [If spoken input could be displayed in writing with errors marked and corrections shown, it would be better] (Kerem, W5). Overall, the participants not only desired correction but wanted it to be explicit, timely, and pedagogically meaningful.

Several students expressed a wish for more dynamic and interest-driven conversation topics initiated by ChatGPT. Lara proposed that when a learner says “bu konudan konuşalım” [let’s talk about this topic], ChatGPT should be able to open a related conversation thread automatically (Lara, W2). Eren emphasized the importance of flexibility: “Konuyu değiştirmeyi öğrenmesi lazım” [It needs to learn how to switch topics] (Eren, W3).

Derya highlighted the potential for engaging conversations if the AI could “konuyla bağlantılı farklı konulara geçiş yapabilir” [switch to other related topics] (Derya, W3), while Kerem recommended that ChatGPT “Farklı konularda veya ilgi alanlarında sohbet açmaya çalışabilir” [try to initiate conversations in different topics or areas of interest] (Kerem, W3). These suggestions show a need for intelligent, learner-driven topic variation to maintain motivation and relevance.

### 4.2.9 Social Suggestions of ChatGPT Identified in Learner Diaries

As learners engaged in weekly interactions with ChatGPT, many began to express their aspirations for a more emotionally resonant and natural dialogue experience. These reflections went beyond functionality or accuracy, revealing a deeper need for human-like connection and interaction quality. Students frequently highlighted how the AI's lack of emotional tone, spontaneity, or personalized response patterns affected the authenticity of their conversational experience. This indicates that while learners valued ChatGPT's role in language learning, they also sought a sense of companionship and responsiveness that mirrored real-life human interaction. These insights were synthesized under a single, high-frequency code: Desire for Authentic Human-Like Interaction. Table 22 presents the frequency of this theme as identified in the learner diaries.

**Table 26**

*Frequencies for the social suggestions of ChatGPT Identified in Learner Diaries*

| Category | Sub-Categories | Codes | *f* |
|---|---|---|---|
| | Social Suggestions | Desire for Authentic Human-Like Interaction | 42 |

Lara highlighted her wish for more natural and immersive interaction: “İnsan gibi sohbet kısımları olabilir” [It could include conversation parts like a human] (Lara, W5). She also suggested enhancing engagement through visual or animated elements: “İnsan animasyonlarıyla daha zevkli hale getirebilirler” [They could make it more enjoyable with human animations] (Lara, W6). Similarly, Zeynep proposed, “Duygu ağırlıklı karakterler yapılabilir” [Emotionally expressive characters could be developed] (Zeynep, W3), emphasizing a need for affective presence during the conversation.

Another recurrent suggestion was enabling the AI to adopt and maintain a personality. Elif explained, “Kişisel fikri olmalı” [It should have a personal opinion] (Elif, W5) and “Cevaplarında subjektif olabilir” [It can be subjective in its answers] (Elif, W6), showing a preference for more expressive and individualized dialogue. This theme also appears in Eren’s entries, where he noted, “Kendini bir insan gibi tanıyabilir, istediğimiz kişiye dönüşebilir” [It can introduce itself like a human and turn into the person we want] (Eren, W1), envisioning a

dynamic, role-playing capability. His wish to simulate friendship is further reflected in: “Maybe when we start to talk, it can talk to me like a friend” (Eren, W2).

Beyond emotional expression, students also recommended improvements to auditory and visual realism. Baran stated, “Anadili gibi konuşanlar gibi konuşmalı” [It should speak like a native speaker] (Baran, W8), and Derya suggested, “It should have a more human-like voice and the accent we want” (Derya, W1). These insights underline a strong demand for both vocal authenticity and the ability to fine-tune the AI’s conversational traits.

Moreover, the desire for spontaneity and responsiveness was central to many entries. Students found the interactions dull or robotic when the AI lacked personal continuity. “Belleğindeki hobilerim ile alakalı bilgilerle sohbetlerimizde yeni konular açabilir” [It can open new topics based on the hobbies stored in its memory] (Derya, W8), emphasized the expectation for a memory-based, evolving dialogue similar to human interactions.

In sum, learners clearly favored a shift from mechanistic responses toward emotionally intelligent, personality-driven exchanges. Their suggestions demonstrate that while accuracy and fluency are important, authenticity, individuality, and emotional depth are equally essential for sustaining engagement in AI-mediated language learning. This widespread theme was evident in 42 codes, making it a dominant social expectation across participants.

### 4.3. Findings Regarding Learners’ Experiences from the Semi-Structured Interviews

To complement the reflective diary analysis and gain a deeper understanding of students’ experiences with ChatGPT, semi-structured interviews were conducted with selected participants from the experimental group at the end of the 8-week treatment period. These interviews aimed to further explore key themes that had emerged from the diaries, such as improvements in fluency, confidence, interactional quality, and learner autonomy. The interview format allowed participants to elaborate on their weekly reflections, clarify previous entries, and provide more nuanced insights into their perceptions and development over time. The data presented in this section are organized using a participant-based approach, highlighting individual voices while also demonstrating how interview findings reinforced or expanded upon the diary themes.

Lara’s interview reflections provide deeper insight into the progression she experienced while using ChatGPT, especially in terms of fluency, confidence, and interaction quality. Throughout her diaries, she frequently mentioned that the chatbot helped improve her speaking: “İngilizce konuşmayı geliştiriyor” [It improves my English speaking] (Lara, W3), and later added that “konuşmamı ve dinlememi geliştiriyor” [I think it has improved my speaking and listening] (Lara, W4). These early reflections were elaborated further in her interview, where

she stated: “Özellikle ilk konuşmama bakılınca çok fazla akıcı konuşmuşum... şu anda daha akıcı ve hızlı konuşuyorum” [Compared to my first recording, now I speak more fluently and faster] (Lara, Interview). This change points to a growing sense of fluency and automaticity, which she attributed to repeated exposure and practice: “o cümleyle ilgili konuşmayı çoktan 5-10 kere konuşmuşum” [I've already talked about that sentence 5-10 times].

Moreover, the chatbot contributed to her self-confidence. In her diary, she had emphasized the emotional comfort of talking to an AI: “arkadaşın gibi sohbet ediyor” [It talks like a friend] (Lara, W2). The interview confirms this social-emotional impact: “Özgüvenimi geliştiriyor... çünkü ben buna başlamadan önce bir yabancı gördüğümde onunla konuşmaya çekiniyordum. Fakat şu an konuşuyorum” [It builds my confidence… before, I was hesitant to talk to foreigners, but now I do] (Lara, Interview). These feelings of comfort and reduced judgment allowed her to take more linguistic risks, contributing to both affective and linguistic development.

Nevertheless, Lara also noted challenges. She repeatedly expressed difficulty when the chatbot used complex or abstract vocabulary: “Seviyem düşük dememe rağmen zor kavramlar kullanıyor” [Even though I said my level was low, it used difficult concepts] (Lara, W3). This was echoed in the interview, where she said: “Ben B1 seviyesindeyim, ona göre konuş dememe rağmen… kelimenin tanımını yaptı sadece” [I told it I was B1 level, but instead of simplifying, it just gave definitions] (Lara, Interview). This reveals a gap in ChatGPT’s ability to adapt input to the user’s proficiency level—a limitation that affects the perceived usefulness of the tool in lower proficiency contexts.

Lara’s suggestions across weeks were consistent and thoughtful. She desired more interactive feedback features: “konuşma sonunda akıcılık, kelime dağarcığı falan analizi de olabilir” [There could be an analysis at the end of the conversation on fluency and vocabulary] (Lara, W8), and “anlamadığı kelimeleri farklı yazmak yerine sorabilir” [Instead of changing misunderstood words, it could ask the user] (Lara, W8). These suggestions reflect a need for more transparent and responsive interactions, where the user has agency in the repair process. In her interview, she confirmed this: “cümleyi kendi cümleye göre kendisi düzeltiyor ve çoğunlukla da yanlış düzeltiyor” [It auto-corrects my sentence based on its guess, but often incorrectly] (Lara, Interview). This shows that while the AI provides timely responses, there is still a need for more context-sensitive repair and clarification strategies.

Overall, Lara’s data illustrates both the empowering and limiting aspects of AI-mediated speaking practice. On one hand, it enhanced her fluency, speaking confidence, and access to conversational opportunities. On the other hand, challenges related to simplification, feedback

accuracy, and user control highlighted the need for further pedagogical design in AI tools aimed at language learners. Her reflections offer a comprehensive view of how learners at the B1 level can benefit from AI-driven speaking support—as long as it remains responsive to their level, encourages confidence, and allows more agency in shaping the conversation.

Zeynep's reflections throughout her diaries and interview reveal an evolving awareness of both the strengths and shortcomings of using ChatGPT as a speaking partner. From the outset, she acknowledged the chatbot's well-designed interface and diverse response options: "advanced sound modu" and "text chat, voice chat section" (Zeynep, W1), and appreciated how the bot could handle diverse topics and generate detailed responses: "bir konuda çok yönlü konuşabilmesi" [its ability to talk in depth on a topic] (Zeynep, W2). In the interview, she emphasized this advantage again: "Yeni fikirler ve yollar üretebiliyor... bana tavsiyeler veriyor ve nasıl daha eğlenceli yapabileceğimi gösteriyor" [It generates new ideas and shows me how to do things in a more fun way] (Zeynep, Interview). She saw the bot not only as a speaking tool, but also as a guide for language learning strategies and self-directed learning.

Zeynep especially valued the bot's role in building speaking fluency and creating a safe, low-pressure practice space. In her interview, she shared: "Sanırım buna daha hızlı konuşmam diyeceğim… dediğim gibi yine pratik açısından katkı sağladı" [I'd say it helped me speak more quickly… it really helped with practice] (Zeynep, Interview). This is supported by her diary entry: "İngilizce pratiği açısından faydalı oluyor" [It is useful for English speaking practice] (Zeynep, W7). Additionally, the opportunity to engage in socially grounded dialogue—even about personal interests—made the experience more relatable: "arkadaş gibi konuşabiliyorum" [I can talk like it's a friend] (Zeynep, Interview). The ability to shape the direction of conversation was noted across weeks: "konuyu açabiliyor" [It can open up topics] (Zeynep, W6), and "kullanıcılara tavsiyede bulunması" [It gives personalized suggestions] (Zeynep, W8), highlighting the bot's flexibility and adaptive potential.

Despite these affordances, Zeynep repeatedly pointed out communication flow issues. Throughout her diary, one recurring problem was interruption during speech: "konuşma bitmeden söze atlaması" [It interrupts before I finish speaking] (Zeynep, W3; W4), and "söz bitmeden konuşmanın karşı tarafa geçmesi" (Zeynep, W2). In the interview, she expanded on this with: "konuşabileceğim kişi robot olduğunu fark ediyorsun ve bu biraz işi sıkıcılaştırıyor" [You realize the person you're talking to is a robot, and that makes it boring] (Zeynep, Interview). The lack of emotional depth also surfaced multiple times as a challenge: "bir insan gibi fikirlere ve duygulara sahip değil" [It doesn't have real thoughts or emotions like a person] (Zeynep, W4), further reflected in her comment: "buna ruh katılabilir... ses tonunun olması

gerektiğini düşünüyorum" [It could be given some soul… I think it should have tone while speaking] (Zeynep, Interview).

Regarding feedback, Zeynep expressed a complex relationship. On one hand, she noted how the bot "nerede ve hangi bilgimde hata olduğunu söylüyor" [tells me where I made mistakes] (Zeynep, W7), and confirmed in the interview: "hatalarımı düzelt gibi bir komut verdiğimde gerçekten düzeltiyor" [when I give a command to correct my mistakes, it really does] (Zeynep, Interview). However, she also encountered inconsistencies: "yanlışlarım varsa düzelt dememe rağmen doğruları da tekrar söylemesi" [even when I ask it to correct mistakes, it repeats correct sentences too] (Zeynep, W6), revealing that while feedback was available, it lacked precision and personalization at times.

Zeynep's suggestions largely centered around improving human-likeness and contextual awareness. In her diary she proposed: "duygu ağırlıklı karakterler yapılabilir" [emotion-oriented characters could be created] (Zeynep, W3), and reiterated in the interview: "karşımdakinin robot olduğunu anlıyorum… bu olmaması için ruh katılabilir" [I realize it's a robot… to fix that, it should be given a spirit] (Zeynep, Interview). She also suggested improving memory and continuity: "önceki cümlelerime yanıt vermiyor" [it doesn't reply to earlier parts of the conversation] (Zeynep, Interview), which aligns with her earlier complaint about repetition and disjointed interaction: "geçmişte sorduğu sorunun benzerini sorması" (Zeynep, W2).

Elif's reflections throughout the study show how ChatGPT offered her both affective support and practical gains in speaking fluency, while also exposing persistent limitations in feedback accuracy, personalization, and natural interaction. In her diaries, Elif repeatedly described the tool as emotionally supportive and accessible, writing: "It's available to speaking 7/24… ideal for students who have social anxiety" (Elif, W1), and "konuşurken sosyal anksiyete sıkıntısı yaşamıyorum" [I don't experience social anxiety while speaking] (Elif, W6). In her interview, she reaffirmed this: "Gerilmiyorsun, daha sakinsin, daha rahat oluyorsun" [You don't feel nervous; you're calmer and more relaxed] (Elif, Interview), emphasizing how the chatbot's nonjudgmental nature reduced pressure during practice.

Elif clearly recognized the platform's ability to foster fluency through repetition and availability. She highlighted that speaking became easier with repeated use: "ChatGPT helps learn to speak fluently" (Elif, W1) and "Pratik çok iyi işe yarıyor… konuştukça daha çok hatırlamaya başladım" [Practice works really well… the more I spoke, the more I started remembering] (Elif, Interview). She appreciated that the chatbot was always available to repeat phrases and answer without frustration: "Bıkmadan cevaplıyor, bıkmadan tekrarlıyor" [It

answers and repeats without getting tired] (Elif, Interview), echoing her earlier diary remarks such as “Yorulmuyor tekrar tekrar dinliyor” (Elif, W3).

However, her expectations for more human-like, emotionally intelligent interaction remained high throughout. From Week 1 onward, she suggested improvements like “must speak like a human”, “Gpt must be like a human”, and “kişisel fikri olmalı” [It should have personal opinions] (Elif, W1; W5). In the interview, this sentiment persisted: “Robotla konuşmadığımı hissetmek istiyorum” [I want to feel like I’m not talking to a robot] (Elif, Interview). She even shared how giving it a name (e.g., Elif) and prompting it to behave like a Turkish girl helped personalize the experience. Nevertheless, Elif showed growing critical awareness of these expectations. By the end of the interview, she questioned whether AI should fully simulate humans, noting: “Çok insan gibi olması da biraz sıkıntı yaratabilir… toplumun ahlaksal normlarını da etkiler” [Being too human-like could be problematic; it might affect societal norms] (Elif, Interview). This shift reveals a nuanced understanding of AI’s role—both as a tool and as a boundary of human-machine interaction.

In terms of pedagogical feedback, Elif consistently called for better grammar correction and explanations. Her diaries show multiple entries like “Gramerimi düzeltmeli” [It should correct my grammar] (Elif, W5), “Cümlelerimi tamamlamama yardım etmeli” [It should help me complete my sentences] (Elif, W4), and “Yanlışlarımı düzeltirken açıklaması tatmin edici değil” [Its explanations when correcting are not satisfying] (Elif, W8). The interview further reinforced this need: “Gramer hatasını düzeltmek… olmazsa olmaz bu” [Correcting grammar is essential] (Elif, Interview). She also desired multimodal features like subtitles and visual cues to support comprehension: “Altyazıyla konuşması çok kritik” [Speaking with subtitles is very important] (Elif, Interview), showing a preference for multisensory learning environments that support weaker listening skills.

Yet, Elif also noted technical and interactional limitations. She was often frustrated by comprehension breakdowns, fast speech, and the lack of follow-up clarification: “Çok hızlı konuşuyor… dur ne dedin diye sorsam yapamıyor” [It speaks too fast… when I ask it to repeat or clarify, it can’t] (Elif, Interview). This was consistent with her diary notes about fast speed and misunderstanding: “Bazen dediklerimi anlamıyor” (Elif, W5), and “Gramer hatamı düzeltmiyor” (Elif, W4). She also pointed out a lack of creativity and personalization in the responses: “Konuştuğum konularda yaratıcı soru sormuyor” [It doesn’t ask creative questions about what I’m talking about] (Elif, W6), and “Cevapları çok sıradan” [Its responses are too ordinary] (Elif, W7).

Ultimately, Elif valued ChatGPT as a tool that offered safe, repeated, and accessible speaking practice, especially for recalling forgotten vocabulary and improving fluency over time: “Strong'u unuttuğumu hatırlıyorum… unuttuğum kelimeleri hatırlamamı sağlıyor” [It helped me remember words I had forgotten, like “strong”] (Elif, Interview). At the same time, she remained critically aware of its limitations—especially in delivering meaningful feedback and simulating authentic, adaptive human-like conversation. Her experience shows the value of combining AI with visual and corrective scaffolds for learners who benefit from clear structure, repetition, and emotional safety in language learning.

Eren’s reflections offer a dynamic account of how ChatGPT supported his English speaking practice, both linguistically and metacognitively. In his diaries, he initially expressed surprise at the chatbot’s range and depth: “Çok güçlü altyapısı var, neredeyse her şeye hakim” [It has a very strong infrastructure; it's knowledgeable about almost everything] (Eren, W1). This impression deepened during the interview, where he stated: “Botun bu kadar geniş kapsamlı konuşacağını hiç düşünmüyordum… konu üretebildiği için keyifli, arkadaş canlısı bir varlık gibi oldu” [I didn’t expect the bot to be this capable… it became an enjoyable, friendly presence] (Eren, Interview). These remarks suggest that Eren developed an emotional bond with the chatbot, seeing it not just as a tool but as a communicative partner.

In terms of language development, Eren demonstrated clear awareness of the bot’s pedagogical impact, particularly on his fluency and discourse management. One critical gain was learning to replace Turkish hesitation markers with English discourse fillers: “İngilizce konuşurken takıldığımda ‘ı, e’ yerine ‘well, you know’ gibi ifadeler kullanmaya başladım” [I started using English fillers instead of Turkish ones when I hesitate while speaking] (Eren, Interview). This observation links to his earlier feedback: “When I think what am I gonna say, it waits for me” (Eren, W2), reflecting how real-time interaction with ChatGPT trained him in rhythm and turn-taking—core elements of spoken fluency.

Eren also made frequent use of explicit feedback features, requesting grammar corrections from the bot: “Bazen ben ondan konuşma içinde feedback istiyordum. Gramerimi düzeltiyordu bazı konularda” [Sometimes I would ask for feedback mid-conversation, and it corrected my grammar] (Eren, Interview). These reflections align with diary entries such as “Feedback vermesini istediğimde... bütün hikayemi baştan okuyarak düzeltti ve duygusunu belli ederek feedback verdi” [When I asked for feedback, it reviewed and corrected my full story with expressive tone] (Eren, W7), showing that he valued both corrective input and the chatbot’s expressive delivery.

While Eren expressed overall satisfaction, he also pinpointed several challenges, especially related to consistency and contextual comprehension. He noted: “Arkadaş gibi konuş dediğimde ikinci üçüncü cümleden sonra yine bot gibi devam ediyor” [Even when I said “talk to me like a friend,” it quickly returned to bot-like speech] (Eren, Interview), which resonates with his diary: “When I told speak with me as a friend… it can't do that” (Eren, W2). The lack of persistent personalization and the need to repeat commands were recurring frustrations. Similarly, technical issues like speech delays, freezing, and incorrect command interpretation also emerged: “Gpt sesi çok fazla dondu” [Its voice froze often] (Eren, W5), and “Verdiğim iki komutu algılamadı” [It didn’t understand two of my commands] (Eren, W7).

Eren consistently advocated for a more human-like and context-aware chatbot, repeatedly suggesting that ChatGPT should adjust its tone, memory, and personality according to user intent. His proposals included subtitled speech, emotional sensitivity, and scenario-based adaptation: “Bir insanla İngilizce konuşuyormuşum gibi hissiyat yaratmalı” [It should feel like talking to a real person in English] (Eren, Interview). These preferences also appeared across his diary suggestions, such as: “Konuşma arasında seviyeme göre yukarı kelimeler kullanıp öğretebilir” [It should use slightly higher-level vocabulary in speech to teach me] (Eren, W1), and “Kültürler hakkında konuşurken oraya ait şeylerden bahsedebilir” [It could bring up culturally relevant things when talking about that place] (Eren, Interview).

Finally, Eren described a clear sense of achievement. He recalled moments of sustained interaction with satisfaction: “20 dakikalık kesintisiz akıcı bir şekilde sohbet ettiğimizde başardım hissi oluyor” [When we spoke for 20 minutes fluently, I felt a sense of accomplishment] (Eren, Interview). His future goals include using English with foreign customers at work and leveraging real-world communication opportunities to solidify language knowledge. Although he does not plan a major next step immediately, he continues to view ChatGPT as a valuable support tool in his language journey: “Eksikleri olsa da faydası var yine de totalde” [Even with its flaws, it’s still beneficial overall] (Eren, Interview).

Baran’s experience with ChatGPT evolved notably over the eight weeks, revealing both technical adaptation and growing linguistic awareness. In his diary entries, he emphasized ChatGPT’s emotional resonance early on: “ChatGPT can speak like a human… GPT can speak very friendly” (Baran, W1). This emotional comfort was revisited in the interview, where he described how the bot’s friendliness and flexibility increased his willingness to speak: “Konuşmalarımın arasında biraz beklediğimde o hemen fikrini söylemeye başlıyordu… ama manuel modla çözdüm” [When I paused between words, it interrupted me—but I solved it with manual mode] (Baran, Interview). This shift reflects his adaptation not only to the tool but also

to strategic interaction management—learning how to set up the bot for more fluid communication.

Baran particularly benefited from corrective feedback delivered in context. In one session, he recalled asking ChatGPT to fix his grammar and pronunciation errors mid-conversation: “Bir anda konu alakasız başlıyor… ama sonra söylediğimi daha doğru bir şekilde söylüyor… ve böyle sürpriz bir şekilde düzeltmesi aklımda kalıyor” [It restated my sentence correctly without announcing the correction, and that made it stick more effectively] (Baran, Interview). This aligns with his diary where he appreciated “açıklayıcı ve gerçekçi cevaplar” [clear and realistic responses] (Baran, W5), and with his broader belief that implicit correction can be pedagogically powerful when it’s naturally embedded.

However, Baran also faced comprehension challenges with fast-paced speech and platform functionality. His diary consistently noted: “Sometimes I can’t understand ChatGPT’s accent” (Baran, W1), and “bazı kelimelerimi umursamıyor” [It ignores some of my words] (Baran, W3). These issues were reiterated in the interview: “Bazen cümleyi kursak bile anlamıyorum, farklı şeyler söylüyor” [Sometimes I form a sentence but don’t understand it, or it responds unpredictably] (Baran, Interview). He also expressed frustration with app instability, such as the screen shifting or conversations switching to text unexpectedly: “Telefonumu yan çevirince konuşma kapanıyor, yazı moduna dönüyor” [When I tilt my phone, it ends the voice mode and switches to typing] (Baran, Interview).

A central theme in Baran’s reflections is his desire for more realistic, emotionally nuanced interaction. Across his diaries, he repeatedly highlighted how “Gelişmiş konuşma daha sarıyor… standart konuşma sıkıcı oluyor” [Advanced mode is more engaging; standard mode is boring] (Baran, W6). He explained that tone and emotional variation shape his experience deeply: “Gülerek arada espri yapması bile konuşmanın seyrini değiştiriyor” [Even just joking mid-conversation changes the flow entirely] (Baran, Interview). His suggestion to let the chatbot remember user history and initiate based on personal context further reflects his desire for human-like relational continuity, e.g., “Daha önceden bisikletle ilgili konuşmuştuk… tekrar bana bisikletle ilgili bir şey sorsa çok gerçekçi olur” [It would be realistic if it remembered we talked about biking and asked again later] (Baran, Interview).

Baran also offered detailed reflections on technical and pedagogical design improvements. He advocated for longer access to advanced mode across weeks (Baran, W1; W6), and requested “highlighted grammar feedback and subtitled speech with color codes” during real-time conversation (Baran, Interview). Additionally, he showed interest in memory features and dynamic feedback: “Konunun akışını iyi takip ediyor… belleğe kaydediyor ve

diyalog listesinden geri dönebiliyorum" [It tracks the flow of conversation well, saves it in memory, and lets me go back later] (Baran, Interview). These preferences show how learners like Baran value scaffolded reflection, error awareness, and visual memory support during interaction.

In terms of personal growth, Baran reported a clear sense of progress by the end of the program. In his Week 8 diary, he noted: "ChatGPT bana ilk ve son haftamı değerlendirdi" [ChatGPT evaluated my first and last weeks] (Baran, W8). He elaborated in the interview: "İlk haftama 4-5 puan, son haftama 7-8 puan vermişti… beni biraz övmesi iyi hissettirdi" [It rated my growth and praised me—that felt great] (Baran, Interview). These moments of positive reinforcement supported his motivation and helped him visualize his improvement trajectory.

Emir's reflections throughout the study demonstrate an evolving sense of connection with ChatGPT and its potential to support language learning. From the outset, he appreciated its responsive nature: "Genel olarak güzeldi. Bana birçok şey kattığını söyleyebilirim. Hem İngilizcemi geliştirme açısından, Speaking ya da listening becerilerimi geliştirme açısından bana birçok şey kattı" (Emir, Interview). This aligns with his earlier diary entry praising its memory and speed: "Hafızasında tutması iletişimi başlatıyor" and "verdiği hızlı tavsiyeler iyi" (Emir, W1).

One of the most striking points in the interview was Emir's perception of emotional engagement and simulated friendship: "ChatGPT insan gibi... hatta ilerleyen zamanlarda... daha samimi olduğumuzda... duygu falan da katarak konuşmaya başlamıştı" (Emir, Interview). His sense of fluency also improved over time: "Eskiden İngilizce konuşurken çok duraksıyordum, ee falan diyordum... ama artık daha akıcı konuşmaya başladığımı düşünüyorum" (Emir, Interview). This development parallels his diary entries in later weeks where he noted, "Konuşmalarda insana daha çok benziyor" (Emir, W8).

Despite these strengths, Emir also encountered frustrations. A recurring complaint was the system's automatic language-switching: "Bir tane Türkçe kelime duyduğunda... hemen sohbeti Türkçe'ye çeviriyor. Bu en büyük sorun olabilir" (Emir, Interview), which he had also identified in week one: "Sadece bir Türkçe kelime bile duysa anında Türkçeye dönüyor" (Emir, W1). Interruptions were another source of tension: "Konuşurken sözümü kesiyor" (Emir, Interview); "Sesimi bölmemeli (kesmemeli)" (Emir, W1).

Emir expressed a clear wish for more robust interactional capabilities in AI systems, particularly for error handling and natural conversation flow. He emphasized the importance of understanding incomplete or incorrect utterances: "Bence burada en önemli olay tam oluşmamış soruları anlayabilme" (Emir, Interview). He added that while some of these existed

in ChatGPT, “Konuşma esnasında... burayı yanlış söyledin gibi şeyler dese... daha hoş olabilirdi” (Emir, Interview).

Looking ahead, Emir appeared cautiously appreciative. While he continues to attend an English course, he was unsure if he would rely on ChatGPT as a primary speaking tool: “Bunu bir İngilizce geliştirmeye devam etmek olarak kullanır mıyım emin değilim” (Emir, Interview). Nonetheless, he acknowledged its advantage in providing better responses in English than in Turkish: “ChatGPT ile İngilizce konuşmak daha etkili cevaplar veriyor Türkçeye sanırım” (Emir, Interview).

Yusuf’s reflections throughout the 8-week diary period and the follow-up interview provide a nuanced view of ChatGPT’s technological and pedagogical value, particularly for learners with intermediate-level speaking and comprehension skills. From the outset, Yusuf highlighted the system’s broad knowledge base and impressive memory functions: “It almost knows everything” and “when it thinks it’s important, it saves it in its memory” (Yusuf, W1). However, these strengths were not without caveats, as he also reported that “if you exit the app and rejoin all of the saves reset” and that ChatGPT sometimes “talks as if it knows the topic” even when it doesn’t (Yusuf, W1).

During the interview, Yusuf expanded on these early concerns by noting that while the experience was generally positive, there were moments of frustration: “Oldukça iyiydi ama bazı yerlerde sıkıntı oldu… bazı bilgileri bilmiyormuş ama biliyormuş gibi konuşuyordu… sunucular çöküyordu” (Yusuf, Interview). His frustration with connection instability and system crashes was also documented in several diary entries, where he stated, for example, “When you ask something fast, it crushes” (Yusuf, W2) and “internet iyiyken bile çekmiyor” (Yusuf, W6).

Despite these technical issues, Yusuf described several pedagogical benefits, particularly in vocabulary acquisition and comprehension support. In his own words: “Mesela bir konu hakkında internet hakkında konuşuyorum. Bana internetin bilmediğim bir özelliğini söylüyor… Açıklar mısın diyorum… bazı kelimelerle açıklıyor” (Yusuf, Interview). He valued ChatGPT’s ability to break down unfamiliar terms and explain them conversationally—although he was aware that “bilgi verdi mi konuşurken aynı zamanda çete basamıyorsunuz” due to app design limitations (Yusuf, Interview).

Another frequent complaint Yusuf shared was the system’s struggle with interruption and turn-taking: “It doesn’t wait for your dialog ends” (Yusuf, W1); “It almost always cuts me while I’m talking” (Yusuf, W2); and “Konuşmamı bölüyor” (Yusuf, W5). These patterns were consistent across multiple weeks and were elaborated on in his suggestions: “It should wait around 7 seconds to start speaking without interrupting you” (Yusuf, W1).

Interestingly, Yusuf viewed ChatGPT as particularly useful for A1–B1 level learners, explaining, “A1 mantıklı geliyor… b1 b2 civarında daha fazla grammar’a yönelik olduğunu düşünüyorum” (Yusuf, Interview). He noted that it could explain grammar rules and support productive skills like writing and speaking, although it sometimes gave overly complex or abstract answers. Yusuf emphasized that for more advanced learners (C1/C2), real-time conversation with humans was more beneficial.

Regarding ideal design features, he was clear about the importance of integrating grammar explanations and maintaining topic relevance. “Geniş bir grammar olmalı… konuşma ve yazmadaki akıcılık olması gerek… sonra yenilikçi özellik olmalı” (Yusuf, Interview). He was also one of the few students to raise the importance of innovation and personalization, expressing a desire for chatbots to evolve competitively with other platforms like Character.AI.

He described a specific moment of pride related to his feedback being seemingly considered by the system: “Ben de bir öneride bulunmuştum… sonra böyle evet güzel bir özellik olur… gurur duymuştum böyle olduktan sonra” (Yusuf, Interview). Even if the system had not directly acknowledged his input, this instance was significant to him emotionally and motivationally.

By the final week, Yusuf reported a more emotionally responsive interaction with ChatGPT: “Bu çalışmanın son haftası olduğunu söylediğimde bana güzel şeyler söyledi” (Yusuf, W8). His perception of ChatGPT as more human-like and affective grew, as also noted when he wrote “it can talk so calmly, I think that’s good” (Yusuf, W1) and later described its tone as supportive (W8).

Looking forward, Yusuf mentioned that although ChatGPT had its place, he didn’t plan to rely on it heavily unless its voice interface improved significantly: “Şu anda zor gibi… ama ileride yenilik getirirlerse… belki yardım alabileceğim bir asistan gibi düşünebilirim” (Yusuf, Interview). Instead, his main focus was on participating in Erasmus+ to practice English in an immersive environment.

Mert's experience with ChatGPT as an English speaking partner was shaped by both enthusiasm and a growing awareness of its limitations. From the beginning, he described the tool as “easy to use” and emphasized how he could say what he wanted and be understood, even noting that it could "understand me when I make mistakes" (Mert, W1). This ability seemed to contribute to his sense of comfort and flow in conversations, particularly in spontaneous exchanges: “Kullandığım kolay, hem de istediğimi söylüyorum, anlıyor” [It was easy to use, and it understood what I said] (Mert, Interview).

A notable advantage for him was ChatGPT's topical flexibility. He appreciated that he could initiate conversations on varied subjects and "farklı konularda farklı cümleler kurmam" [form different sentences in different topics], which he viewed as the primary source of benefit for his speaking practice (Mert, Interview). This is echoed in his diaries as well, where he praised how the AI "can create so many topics" (Mert, W4) and "can come up with many questions when I talk about something" (Mert, W5). He rated his experience as "easy" and "a good experience" (Mert, W2), suggesting a degree of fluency development in unstructured conversation.

However, Mert's reflections also illustrate several recurring challenges that affected his interactional flow and comprehension. Chief among them was ChatGPT's tendency to interrupt or "break" his speech. He expressed strong frustration with this: "Konuşurken bir anda sözümü kesip Türkçe'ye çeviriyor. O çok sıkıntı oluyor. Çok sinirleniyorum" [While speaking it suddenly cuts me off and switches to Turkish. It's really annoying. I get very frustrated] (Mert, Interview). This complaint is consistent across diary weeks: "AI breaking my speech while I'm speaking" (Mert, W1; W3; W4; W5). He also struggled with the AI's listening comprehension, frequently stating, "I can't understand everything" (Mert, W1) and "it doesn't speak when I ask something" (Mert, W5; W6), pointing to technical breakdowns and turn-taking issues.

Another critical challenge was ChatGPT's occasional "topic jumping," which Mert mentioned in both interview and diaries. For instance, "It suddenly changed the topic" (Mert, W2), and during the interview: "Devlet hakkında konuşuyoruz. Milletvekili diyorum milletvekiliye geçiyor. Sonra milletvekilinden genel bakana falan geçiyor. Çok hızlı değişiyor" [We're talking about government, I say parliament member, it switches to that, and then quickly to minister—it shifts too fast] (Mert, Interview). This erratic flow interfered with sustained topic development, undermining deeper language practice.

Mert also had strong views on cultural adaptability and emotional resonance. While he didn't demand full human-like emulation, he expressed a desire for the bot to better understand Turkish users' social logic: "Türklerin ne düşündüğünü anlasa güzel olur" [It would be nice if it understood what Turks think] (Mert, Interview). He also emphasized the value of prosody: "Ses tonlama göre ne söylediğimi ne kastettiğimi anlayabilse güzel olur" [It would be nice if it could understand my intent based on tone of voice] (Mert, Interview).

In terms of ideal chatbot design, Mert prioritized three features: "geniş bir kelime dağarcığı," "geniş bir konu hazinesi," ve "dediklerimizi anlayabilme" [broad vocabulary, broad topic capacity, and understanding what we say] (Mert, Interview). However, he was also clear that he does not use ChatGPT for error correction or grammar feedback: "Normal bir insanla

konuşuyormuş gibi konuşuyorum… hatalarına bakmak, geri bildirim vermesi vs. bu tarz şeyleri kullanmayı düşünmüyorum" [I just talk like I would with a normal person… I don't care about feedback or grammar corrections] (Mert, Interview).

In the final part of the interview, Mert reflected that he didn't really experience a "sense of achievement" because the experience felt more like everyday conversation: "Başarı hissi duyduğum pek bir durum yok çünkü zaten normal bir insanla konuşuyormuş gibi konuşuyorum" (Mert, Interview). Nevertheless, he did express interest in improving his English by "konuşarak," especially through real-life interactions with foreigners at university (Mert, Interview).

Derya's overall experience with the AI chatbot was shaped by both her academic and conversational use of the tool. She explained that the chatbot was helpful for academic tasks such as "matematik sorularını, fizik sorularını" [math and physics questions] and that this usefulness extended into English learning. Specifically, when practicing around casual topics like hobbies, she said, "insan gibi cevapları oldu" [it gave human-like responses] (Derya, Interview). In her diaries, she highlighted the chatbot's ability to generate and adapt information, saying, "It can generate a lot of questions" and "I can make a deep information research" (Derya, W1).

One of the most beneficial features for Derya was the chatbot's instant corrective feedback: "anında görünüş verebilmesi oldu... grammar yapısını geliştirmeme yardımcı oldu" [it was able to provide immediate feedback... helped improve my grammar structure] (Derya, Interview). She also appreciated that "it explains the words that I can't understand" and "it can understand the sentences that I've made incorrectly" (Derya, W1), reinforcing how this contributed to her speaking development. According to her, these corrections became even more useful when speaking about "hobiler hakkında biraz günlük… farklı yabancılarla sohbet ederken" [hobbies and casual daily topics, especially when chatting with foreigners] (Derya, Interview).

Nevertheless, she encountered a few persistent challenges. For example, she was frustrated that the bot "konuşmaları bazen çok uzatabiliyor… aynı soruyu birden fazla sormaya çalışıyor" [sometimes prolongs conversations too much… tries to ask the same question more than once] (Derya, Interview). In line with this, her diary noted: "It wants to continue to speak about the topic for a long time and I just get bored" (Derya, W2). Another common issue was interruption, as she wrote: "It doesn't wait me to continue when I think of the rest of the sentence" (Derya, W1), and found the voice "monotone" and emotionally flat, which made the experience less engaging.

From a pedagogical perspective, she valued how the bot handled incomplete input and ambiguous questions. She said: “Bir cümleyi tam bağlamını vermeden anlayabilmesi... olumlu bir şey olduğunu düşünüyorum” [its ability to understand a sentence even when I don’t provide full context... I see that as a good thing] (Derya, Interview). She also mentioned that “yazım hatalarını yönetebilme... göze aşinalık oluşturuyor” [being able to manage spelling mistakes helps build visual familiarity] (Derya, Interview), particularly when it corrected small details like capitalized months or complex spelling.

In terms of design suggestions, Derya emphasized the need for personalization. She believed a more effective AI speaking assistant should “senin olduğun ve ulaşmak istediğin seviyeyi iyice anlayabilmesi lazım” [clearly understand your current level and your target level] and “ilgi alanlarımızı belliğinde daha net bir şekilde kesinleştirmeli” [better recognize and remember our interests in its memory] (Derya, Interview; Derya, W3). She also expressed a desire for a more dynamic voice experience, stating: “eğer insan sesi baz alınsaydı... bu daha iyi bir geri dönüşü olabilirdi” [if it were based on a real human voice, the feedback experience would be better] (Derya, Interview).

Derya’s motivation noticeably increased toward the end of the program. She recalled: “IELTS sınavında sence kaç alabilirim diye sordum... bana yüksek bir puan vermesi beni motive etti” [I asked what IELTS score I might get... and getting a high estimate motivated me] (Derya, Interview). In Week 8, she wrote that the AI “benim hakkımda yaptığı feedbackler gerçekten IELTS’ten 7 alabileceğime inanmamı sağladı” [the feedback it gave me made me believe I could get a 7 in IELTS] (Derya, W8).

Looking ahead, Derya plans to continue using the tool: “Büyük ihtimalle onu bir çeviri olarak kullanmaya devam ederim... çoğu çeviri programlarından daha doğru sonuç verebiliyor” [I’ll likely keep using it for translation... it gives more accurate results than most translation apps] (Derya, Interview). Rather than focusing on using it strictly for speaking, she seems to see its value more broadly as an assistive language tool in everyday learning.

Kerem initially approached the AI-assisted speaking experience with curiosity and openness. In the beginning, he found the concept of interacting with an AI quite novel: "Hocam ilk başlarda böyle enteresan bir deneyim olduğu için akla farklı sorular geliyordu" [At first, it was quite an interesting experience, so many different questions would come to mind] (Kerem, Interview). However, over time, he became less engaged, noting that the responses became repetitive and the system was not specifically tailored for language learners: "Sonradan cevaplar böyle birbirine benzemeye başladıkça biraz açıkça sıkılmaya başladım" [Later on, as the

answers started to resemble each other, I honestly got a bit bored] (Kerem, Interview). This sentiment of decreased motivation was echoed across several diary entries as well.

In terms of language practice, Kerem saw some potential for vocabulary enrichment and exposure to new structures, especially in daily conversation. He noted that the bot helped him consider alternative expressions after sessions, even though the benefit was delayed: "Kendimce bir araştırma yaparken, düşünürken faydalı olur" [It becomes useful when I reflect and research on my own after the conversation] (Kerem, Interview). This reflective strategy was present in his diaries too, such as when he mentioned: “Bilmediğin kelimeyi cümle içinde kullanıp öğrenmeni sağlıyor” [It helps you learn unknown words by using them in sentences] (Kerem, W1).

Regarding feedback and grammar, Kerem consistently expressed dissatisfaction. Despite explicitly requesting corrections, he found the bot reluctant to provide them: "Ben düzelttiğime rağmen böyle birkaç kere komut girmeme rağmen... düzeltmediğini görmüştüm" [Even though I asked it to correct me several times, it didn’t] (Kerem, Interview). This aligns with his diary observations: “Hatalarımı düzelt komutu girdiğim zaman hatalarımı düzeltmediği zamanlar olduğunu fark ettim” [I noticed that sometimes it didn’t correct my mistakes even when I asked it to] (Kerem, W5). He feared this might lead to the reinforcement of incorrect structures.

From a technical and usability standpoint, Kerem faced persistent performance issues—particularly related to audio input and lag. He reported: “Konuşmamı bitirdiğimde hala ‘dinleniyor’ ifadesi geliyor ve GPT cevap hazırlamıyor” [After I finish speaking, it still says 'listening' and doesn’t prepare an answer] (Kerem, W8). These issues were exacerbated when other applications like screen recording were running. In his interview, he attributed many problems to device-specific limitations: “Ses kaydıyla ilgili... video kaydı almadan da deneyeceğim” [I’ll try without recording the screen next time] (Kerem, Interview).

Kerem also criticized the lack of adaptive difficulty in conversation. He emphasized the importance of the system tailoring its output to the learner’s level: “Eğer A1 seviyesindeki... birine çok üst kelimelerle cevap verirse... zorlanır” [If it responds with too high-level vocabulary to an A1 learner, they will struggle] (Kerem, Interview). He reiterated this need in his diaries by suggesting “belirli roller olabilir konuşma yapmak için” [there could be specific roles or scenarios for conversation] (Kerem, W6).

In terms of suggestions for improvement, Kerem offered multiple pedagogical and technical ideas, such as the inclusion of “a button to pause the AI's voice” or a feature that allows the learner to confirm unclear words (Kerem, W2). He also proposed adding “character

types" to simulate conversations with different personalities (Interview), echoing his desire for more human-like, dynamic interaction.

Despite the challenges, Kerem experienced moments of progress and confidence, especially when he noticed spontaneous speech: "Aklından bir cümleyi çevirmediğin zaman direkt İngilizce konuşmaya başlıyorsun" [When you don't translate a sentence in your head but start speaking directly in English] (Kerem, Interview). Although no strong moment of success stood out, he acknowledged a gradual sense of improvement.

Looking forward, Kerem expressed an intention to continue writing in English to support his development, often jotting ideas in a notebook: "Böyle karalama defterimin bir köşesine yazıyorum" [I write in a corner of my draft notebook] (Kerem, Interview). Combined with ongoing coursework, this indicates a self-directed approach to learning, with AI serving as a supplementary tool rather than a central one.

# CHAPTER 5
# DISCUSSION

This chapter discusses the research questions in relation to the study's findings and relevant literature. It begins with quantitative results on speaking performance, followed by key themes from the qualitative data. These include insights into learner interaction and the technological, pedagogical, and social benefits of using AI chatbots.

## 5.1. Discussion of Speaking Skill Gains in the Experimental Versus Control Groups

The findings of this study provide insight into both research question 1 and research question 2. Regarding research question 1, which explored whether there were significant differences in speaking proficiency between the experimental and control groups following the treatment period, the results indicated no statistically significant difference between the two groups at the post-test stage. Despite this, it is worth noting that the experimental group did show a slightly higher average increase in speaking proficiency. This hints at a potential positive trend for those who practiced speaking with the help of ChatGPT.

Concerning research question 2, which focused on within-group differences before and after the intervention, both the experimental and the control groups demonstrated significant improvements in their speaking scores over time. This finding highlights that students in both groups benefited from the learning process, albeit through different modes of engagement. Importantly, the experimental group's interaction with ChatGPT appears to have offered an additional layer of speaking practice, potentially accounting for their relatively greater gain.

These findings mirror those reported by Koç (2024), where no significant difference emerged between the experimental and control groups in post-intervention comparisons, despite higher gains in the experimental group. Such results suggest that even if AI-based tools do not always produce statistically significant outcomes in short-term studies, they can still support meaningful progress in speaking performance. The consistent within-group improvements in the experimental group across multiple studies (Koç, 2024; Fathi et al., 2024; Yin et al., 2020) reinforce the pedagogical potential of AI chatbots in oral skill development.

On the other hand, Fathi et al. (2024) reported statistically significant differences between the experimental and control groups, particularly in lexical diversity and grammatical accuracy, favoring the group that engaged with AI chatbots. Their study, which employed a longer treatment period and more structured task-based interactions, highlights the importance of sustained and pedagogically guided exposure to AI tools.

However, the lack of statistically significant differences between the groups may not necessarily be due to a lack of structure, since the experimental group in this study participated

in weekly guided conversations based on specific topics and prompts carefully designed to match their language level. This structured approach aimed to scaffold learners' interactions and support their engagement with ChatGPT in a focused and manageable way. As observed by Lin and Mubarok (2021), structured pedagogical frameworks often yield more pronounced learning outcomes than unguided activities, which supports the rationale behind the design of the current study. Nevertheless, the speaking activities were conducted outside regular classroom hours as an extracurricular component, which may have affected the intensity and consistency of learners' participation. Even though students recorded their interactions with the bot, without the immediate presence of teacher monitoring or feedback, some students may have interacted with the chatbot in a more mechanical or surface-level manner, potentially limiting deeper linguistic processing or output modification.

Technical factors may also have contributed to the outcome. As reported in both the present study and Koç (2024), issues such as recognition errors, feedback limitations, and occasional communication breakdowns with AI chatbots can hinder interaction quality. Such limitations reduce opportunities for receiving corrective input, a key element in language development through modified output (Swain, 2005, as cited in Koç, 2024). Similarly, studies such as those by Dizon (2020) and Hsu et al. (2023) have pointed to the inconsistent accuracy of speech recognition engines, which can discourage learners or limit their ability to benefit from natural feedback cycles. These technical and communicative challenges were also frequently reflected in student diaries and will be more thoroughly discussed in the subsequent sections of this chapter, particularly in relation to learners' perceptions of affordances and constraints.

Nevertheless, the greater gains observed in the experimental group suggest that certain affordances of AI chatbots—such as reduced anxiety and more frequent speaking opportunities—may have supported students' speaking development over time. Learners were able to engage in conversation outside the classroom setting, providing additional exposure to language in a low-stakes, low-anxiety environment, a benefit frequently emphasized in studies of intelligent personal assistants (Chen, Yang, & Lai, 2020; Hapsari & Wu, 2022).

However, as Fathi et al. (2024) argue that while AI-mediated platforms foster low-anxiety environments and encourage willingness to communicate, they cannot replace the nuanced feedback human instructors provide. Learners in their study also mentioned that they missed the depth of human interaction, especially when the chatbot's replies felt off-topic or couldn't keep the conversation going in a meaningful way. These insights highlight how

important it is to combine AI tools with supportive human guidance in order to get the best results in speaking development.

In line with the findings of Fathi et al. (2024), learners in this study also encountered limitations in the chatbot's feedback, often describing it as inconsistent, generic, or lacking relevance to their input. Although chatbots are quick to respond and easy to access, they often fall short when it comes to offering the kind of detailed feedback that human teachers are able to give. This shortcoming can reduce learners' opportunities to notice and revise their errors effectively—particularly during more complex or extended speaking interactions where subtle linguistic scaffolding is needed. In the present study, while the experimental group engaged in weekly guided interactions with ChatGPT, the absence of individualized corrective feedback may have weakened the potential benefits of those sessions. Without sufficient feedback loops or clarification requests from the chatbot, learners may have continued producing inaccurate forms without realizing it. This lack of dynamic feedback could be one reason why the experimental group did not demonstrate a statistically significant advantage over the control group, even though they engaged in structured speaking practice and had more exposure to the target language.

A notable limitation of the present study lies in the small sample size, with only 19 EFL learners participating in the quasi-experimental design. This restricts the generalizability of the findings and limits the extent to which conclusions can be extended to wider learner populations. While the current study provides valuable exploratory insights, the reduced number of participants inherently limits the statistical power of the analyses. Future studies should consider incorporating larger, more representative samples to enhance the external validity of the findings and to explore how generative AI tools function across diverse EFL learning contexts.

Overall, even though the difference between the two groups was not statistically significant, the experimental group still showed more noticeable progress. This improvement, in conjunction with findings from earlier studies, suggests that continuous, learner-driven AI-supported speaking practice can positively contribute to students' speaking development. Future studies might benefit from combining autonomous AI interaction with structured instructional support to better harness the benefits of generative AI in language learning contexts.

### 5.2. Interpreting the Technological Benefits, Challenges, and Suggestions from Learner Experiences

In order to address research question 3, which explored learners' perceptions of using a GenAI chatbot for English speaking practice—including their views on its positive aspects,

drawbacks, and potential improvements—the present section interprets the findings through a technological lens. To answer research question 3a (Positive Aspects), participants in the their weekly diaries reported a wide range of affordances. These findings are discussed here in light of current literature, with a particular emphasis on EFL contexts.

One of the most frequently reported advantages was the speed and efficiency of ChatGPT in providing responses. Learners appreciated its ability to comprehend and reply to prompts quickly, which enabled more continuous interaction compared to traditional classroom settings. This aligns with Du and Daniel's (2024) systematic review, which emphasized the role of AI chatbots in enhancing speaking fluency, pronunciation accuracy, and user confidence—particularly when used in informal or self-directed environments. In the present study, learners also highlighted the chatbot's broad knowledge base, its use of relatable examples, and the ability to generate and adapt ideas, which supported their speaking engagement across different weekly themes. These features correspond with prior findings by Koç (2024), who observed that learners viewed the chatbot as a flexible and knowledgeable partner capable of stimulating linguistic output.

The personalization of responses and the chatbot's use of natural, moderately complex language were also noted as strengths. These qualities appear to promote learner autonomy and reduce communicative pressure—benefits also emphasized by Moussalli and Cardoso (2016) and Hapsari and Wu (2022), who noted that AI tools foster low-anxiety environments conducive to oral production. Moreover, the availability and accessibility of the chatbot—a strength commonly cited in learner diaries—reflects what Du and Daniel (2024) identified as "EO6: easy access and convenience," a core technological affordance of chatbot-mediated learning. This finding resonates with more previous studies in the literature, where ease of use emerged as a key strength of chatbot-based learning environments. For instance, students in earlier research were able to access chatbots seamlessly through their computers, tablets, or mobile phones, encountering no significant technological barriers (Chen et al., 2020; Lin & Chang, 2020). This suggests that the simplicity and accessibility of these tools will continue to play a crucial role in promoting student engagement and comfort during language practice.

Despite these benefits, to answer the research question 3b (negative aspects), several technical and communicative challenges emerged that may have hindered the full effectiveness of AI-supported practice. A major issue was the chatbot's inability to manage turn-taking and maintain dialogue continuity, especially when learners paused or gave extended responses. These problems mirror the "interruption and turn-taking" difficulties recorded in the learner diaries and also match observations by Koç (2024), who noted frequent communication

breakdowns caused by the chatbot's misinterpretation of pauses or speech hesitations. Similarly, Du and Daniel (2024) identified the "limitations of speech recognition" (EC2) as a core barrier, especially when chatbots struggle to process non-native accents or longer utterances.

Learners also shared that they sometimes faced problems with the chatbot giving off-topic or repetitive answers. There were even moments when it misunderstood what they meant—especially during culturally sensitive or open-ended conversations. These kinds of breakdowns made the conversation feel less natural and, for some students, even caused them to lose interest. This aligns with findings from Du and Daniel's (2024) review, which highlighted the frequent breakdowns in communication due to users' inability to interpret the chatbot's implicit cues or recover from sudden topic shifts (Tai, 2020). Furthermore, the limitations of speech recognition systems, especially when processing longer utterances or accented speech, were widely acknowledged as a core barrier to sustained interaction (Chen et al., 2020; Dizon, 2017; Moussalli & Cardoso, 2016). When the chatbot failed to transcribe accurately or understand what was being said, it did not just interrupt the flow of the conversation. It also stopped learners from getting the timely feedback they needed. As noted by Hsu et al. (2023) and Kim, et al. (2021), the chatbots' repetitive patterns and inflexible content design further diminished their pedagogical responsiveness, particularly for learners at different proficiency levels. These technological shortcomings collectively restricted the chatbot's potential to provide personalized, adaptive support—a feature frequently emphasized as critical in successful language learning environments.As also noted in both Koç (2024) and Fathi et al. (2024), learners often feel that while AI chatbots provide fluent responses, they fail to deliver the kind of nuanced, corrective input necessary for deeper language development. These technological issues may also explain, at least in part, why the experimental group did not significantly outperform the control group in the post-test scores, despite engaging in structured chatbot-mediated practice.

Lastly, to shed light on the research question 3c (Suggestions), students in the present study proposed a range of technology-focused suggestions to improve their experience. The most common suggestion from learners was to add customization options which are things like adjusting the pace of the conversation or choosing their own topics. Many also wished the chatbot could remember their previous chats, so the dialogue could feel more connected over time. These findings are echoed in Du and Daniel's (2024) review, where learners advocated for chatbots that could remember prior exchanges and deliver more contextually aware interaction.

In summary, while the participants in this study acknowledged several technological benefits associated with GenAI chatbot use—including speed, availability, and knowledge richness—they also encountered notable limitations in terms of speech recognition, feedback quality, and interaction breakdowns. Their suggestions emphasize the need for more responsive and socially intelligent chatbot designs, particularly in speaking-oriented EFL contexts. These reflections provide valuable implications for future instructional integration and design of AI tools tailored to language learners' communicative needs.

### 5.3. Interpreting the Pedagogical Benefits, Challenges, and Suggestions from Learner Experiences

In order to address research question 3, which examined participants' perceptions of GenAI chatbot use for English speaking practice—including its perceived benefits, limitations, and areas for improvement—this section focuses specifically on the pedagogical dimension of learner experiences. Drawing on patterns identified in the weekly diary data, the analysis explores how students perceived the chatbot as a tool for language development, what instructional challenges they encountered, and what suggestions they proposed for more effective integration. These findings are interpreted in connection with relevant studies in the literature, with a particular emphasis on pedagogical implications for EFL contexts.

Learners often mentioned that chatting with the bot gave them steady chances to speak English, which they found useful for building fluency, picking up new vocabulary, and reinforcing grammar. The weekly prompts were especially helpful. They provided structure to the conversations and helped students stay on track with meaningful practice. Many learners pointed out that the chatbot made it easier to come up with ideas, stay focused on what they were talking about, and use the language they had recently learned in more connected responses.

These learner reflections resonate with findings in the broader literature on AI-supported speaking instruction. As highlighted by Du and Daniel (2024), chatbots have been shown to enhance EFL learners' oral proficiency by simulating authentic dialogues, offering vocabulary support, and helping learners structure their speech in extended output. The chatbot's role as both conversational partner and resource supplier (Shawar & Atwell, 2007, as cited in Du & Daniel, 2024) allows learners to rehearse language in a way that reinforces both accuracy and fluency. In the present study, learners' ability to generate ideas and maintain topic focus appears to be supported by the system's adaptability and responsiveness—features that are particularly important for sustained oral production.

Additionally, the integration of structured prompts likely contributed to learners' sense of pedagogical coherence, making the AI interaction not only engaging but also goal-oriented.

This aligns with Du and Daniel's (2024) observation that task-based chatbot applications—such as Smart UEnglish (Hwang et al., 2022)—can help students practice English in authentic and contextually meaningful ways, thereby strengthening speaking competence. Moreover, learners in the current study often described the chatbot as supportive, accessible, and responsive—qualities echoed in prior studies emphasizing the motivational impact and psychological safety that AI-mediated tools can provide (Azizimajd, 2023; Tai & Chen, 2023). These attributes help create a practice environment where students can experiment with language and engage in meaningful oral tasks without fear of judgment.

Some learners also expressed that ChatGPT served as a kind of personal tutor—readily available for extra practice outside class hours—especially valuable for learners who felt they had limited speaking time during formal lessons. The chatbot provided a real-time interaction with students without time and space limitations (Wang et al., 2017; Xu, 2020).

This perception of the chatbot as a readily available personal tutor reflects broader trends observed in the literature on AI-mediated language instruction. As Du and Daniel (2024) note, chatbots function not only as conversational partners but also as resource suppliers and adaptive instructional facilitators, providing learners with immediate access to language input and practice beyond formal instructional hours. This affordance is particularly beneficial for learners who experience limited classroom time. This study's statement of the problem was grounded in the recognition that many students, particularly in environments where English exposure is minimal—such as Türkiye—struggle to practice speaking due to societal factors like the absence of English input in their daily lives, homes, or schools (Kara et al., 2017). In this context, ChatGPT offers an avenue for ongoing language engagement and supports learning continuity.

Learners' appreciation for the chatbot's consistent availability parallels Fryer and Carpenter's (2006) argument that chatbots can overcome traditional classroom constraints—such as restricted speaking time and social anxiety—by creating accessible, autonomous, and non-judgmental practice spaces. These out-of-class interactions, when supported by structured prompts, appear to scaffold language use in meaningful ways, reinforcing learner autonomy while maintaining a pedagogical focus.

Moreover, the idea of chatbots acting as digital tutors aligns with findings that highlight their capacity to simulate human-like interaction through real-time dialogue, repetition, and task adaptation (Huang et al., 2022; Qasem et al., 2023). According to Du and Daniel (2024), these tools are increasingly recognized for their role in enhancing learner engagement, motivation, and speaking fluency, particularly when integrated with intentional design. In this regard, the

chatbot's role in this study extended beyond mere conversational support; it offered learners a complementary instructional space, responsive to their individual pace and accessible at any time. This aligns with the broader pedagogical shift toward learner-centered, AI-supported environments where students can regulate their language exposure and build communicative confidence at their own pace.

These pedagogical affordances can be further illuminated through Vygotsky's sociocultural theory, particularly as applied to language development by Shabani (2016). According to this framework, meaningful learning occurs when learners are guided within their Zone of Proximal Development (ZPD) through scaffolded interaction with a more knowledgeable other. In this study, ChatGPT appears to function as such a mediational tool—offering structured prompts, vocabulary support, and feedback that gently extend learners' current abilities. The weekly tasks served not only as linguistic stimuli but also as scaffolds, helping learners build towards independent communicative competence. As Shabani emphasizes, effective mediation fosters self-regulation and autonomy, two qualities that were strongly reflected in learner reflections on ChatGPT's accessibility and responsiveness. The chatbot's role in this context mirrors that of a digital tutor operating within the learner's ZPD: it responds dynamically to learner output, supports error correction through interaction, and allows repeated, self-paced engagement with the target language. Thus, the pedagogical strengths observed in this study align with sociocultural views on language acquisition and further validate the potential of AI-mediated dialogue to scaffold speaking development in EFL contexts.

In addition to the reported pedagogical strengths of ChatGPT in providing speaking opportunities and exposure to target language input, participants in this study also expressed several concerns regarding areas where its pedagogical affordances fell short. These challenges revolved around the lack of accurate feedback, repetitive language, and insufficient lexical or grammatical depth, as identified in the weekly learner diaries and follow-up interviews.

One frequently reported issue was inadequate feedback, particularly regarding grammar and pronunciation. While participants appreciated the accessibility and responsiveness of ChatGPT, they often found the tool lacking in delivering the kind of corrective input necessary for supporting linguistic accuracy. Some learners noted that errors in their speech were ignored, leaving them unsure about the accuracy of their output. Interestingly, this perception was echoed in interviews, where several learners expressed appreciation for instant responses, but still preferred human instructors for more detailed, context-aware feedback. At the same time, learners also demonstrated that when given the right prompts, ChatGPT could serve practical

functions beyond language feedback—for instance, helping them generate travel plans or organize tasks. This utility seemed to foster authentic language use in everyday contexts, even if not directly instructional. However, it is important to recognize that this limitation stems from ChatGPT's general-purpose design; it is built as a conversational AI, not as a dedicated pedagogical agent. As such, placing the full responsibility for feedback limitations on the tool itself would overlook its intended function. Nevertheless, this finding aligns with prior observations that chatbots often fail to provide meaningful, timely, or explicit corrective feedback due to their design as communicative companions rather than instructional systems (Koç, 2024, p. 208).

Another challenge frequently voiced by the learners was the repetition and simplification of responses. This diminished the educational depth of interactions. Learners reported encountering redundant expressions and generic prompts, especially when discussing familiar topics. Similarly, Jamshed et al. (2024) observed that Saudi EFL learners expressed concerns about ChatGPT producing inaccurate or contextually shallow answers. These limitations not only affected the richness of conversation but also raised doubts about the tool's reliability and critical thinking value in more nuanced language learning tasks. These pedagogical concerns are supported by Du and Daniel's theme EC4, "the relatively monotonous nature of course content and format," which was identified in several empirical studies (Hsu et al., 2023; Kim et al., 2021). In Koç's study as well, participants remarked that the chatbot Replika responses sometimes lacked variety and complexity, which made sustained engagement difficult and hindered deeper linguistic processing (Koç, 2024, p. 207).

One common challenge that came up in this study was the chatbot's limited ability to adjust its language level. Learners with lower proficiency often found ChatGPT's replies too difficult to follow, which made it harder for them to understand and disrupted the flow of the conversation. Some participants mentioned feeling overwhelmed by complex vocabulary or long sentence structures, which hindered their ability to respond confidently. This concern resonates with Fryer and Carpenter's (2006) claim that chatbots tend to be more suitable for advanced learners than for beginners. They suggest that lower-level learners may produce inaccurate or incomplete input, which chatbots struggle to interpret effectively—resulting in confusion and unhelpful responses.

Interestingly, this finding is quite the opposite of what Koç (2024) reported. In their study, university-level English language teaching students criticized Replika for having a limited vocabulary, making grammar mistakes, and sounding unnatural—even though it could handle somewhat complex sentences (p. 207). The contrast between the two studies likely

comes from the differences between the learners. Koç's participants were more advanced and needed richer, more refined language input. Meanwhile, the learners in the current study were younger and at a pre-intermediate level, so they needed simpler, clearer responses. Still, both cases point to the same core issue: the chatbot struggled to adapt to the user's language level—just in opposite ways. This also matches what Du and Daniel (2024) noted under challenge EC3, "Cannot fit all levels," which points out that chatbots often fall short when trying to support both beginner and advanced learners effectively (p. 9).

ChatGPT, as a more advanced large language model, produced more sophisticated and structurally rigorous responses. However, as evidenced by the findings of this study, it still fell short in adapting its language output to match the proficiency level of lower-level learners. This mismatch often resulted in comprehension difficulties and disrupted interactional flow. Naturally, this limitation stems from the fact that ChatGPT is not a pedagogically designed tool but a general-purpose conversational AI. Nonetheless, its use in this study offers valuable insights into the potential of AI chatbots in language education. Future designs of pedagogical chatbots should consider integrating adaptive language features to better support learners at varying proficiency levels.

Participants in this study also made it clear that they needed more structured, pedagogically guided tasks—such as grammar-focused activities or organized speaking prompts to support their practice. Although ChatGPT did reply to their messages, many students felt unsure about how to begin or carry on with a practice session that would actually help them improve in targeted language areas. Simply put, they wanted the chatbot to be more than just a conversation partner—they hoped it could guide them, too.

This finding echoes what Koç (2024) discovered as well. In that study, ELLs also suggested adding grammar-focused sections to the chatbot experience. Both studies point to the same issue: when there is no structured support, especially in fully learner-driven interactions, the learning process risks being incomplete (p. 220). Similarly, Du and Daniel (2024) emphasized that AI chatbots, though widespread in language education, are often underutilized when not supported by intentional instructional design, limiting their impact on targeted skill development.

Even though weekly speaking prompts were provided, participants in this study still expressed a clear preference for more explicit pedagogical support from ChatGPT. More specifically, students said they wanted clearer grammar corrections and built-in subtitle prompts to support their language growth. While ChatGPT is mainly designed for general conversation and not tailored for educational purposes, the feedback from learners points to important areas

that need improvement. Since most current chatbot technologies are aimed at a broad audience rather than classroom settings, what these students had to say holds substantial value for shaping future designs.

These pedagogical and technological weaknesses, identified throughout this research, underscore a critical gap between the theoretical promises of Intelligent Computer-Assisted Language Learning (ICALL) and its practical implementation in EFL contexts. ICALL systems, as outlined by Amaral and Meurers (2011), are designed to overcome the limitations of traditional CALL through the integration of NLP tools that support dynamic feedback, error diagnosis, and individualized scaffolding. However, in the present study, the AI chatbot did not always meet these expectations. Although ChatGPT embodies key features of ICALL—such as real-time interaction and personalized language generation—learner experiences revealed inconsistencies in its capacity to deliver accurate and context-sensitive feedback, particularly in response to mispronunciations or syntactic errors. These observed limitations are consistent with prior criticisms indicating that existing ICALL systems may fall short in replicating the nuanced corrective feedback typically provided in human interaction (Katsarou et al., 2023; Koç, 2024).

Overall, these findings suggest that while ChatGPT offers flexibility and conversational practice, its pedagogical utility remains constrained by the absence of systematic feedback, adaptive scaffolding, and advanced input features. Future chatbot designs should consider integrating features such as customizable feedback modes, topic-based grammar tasks, and proficiency-level calibration to meet diverse learner needs and promote deeper engagement with the language (Du & Daniel, 2024; Koç, 2024).

### 5.4. Interpreting the Social Benefits, Challenges, and Suggestions from Learner Experiences

In addressing Research Question 3, this section solely focuses on the social dimension of learner experiences with GenAI chatbot use for speaking practice. In particular, it explores how participants perceived the interpersonal, emotional, and affective aspects of interacting with the chatbot—such as motivation, speaking confidence, reduced anxiety, and the sense of social presence. Drawing on thematic codes from student diaries, this section also examines the limitations of AI-mediated interaction from a social standpoint, including issues related to conversational unnaturalness, lack of emotional nuance, and the absence of human connection. Finally, learners' socially motivated suggestions for improving the interaction experience are discussed in light of recent literature, especially studies that explore how chatbots can replicate or support socially meaningful learning environments.

The reflections documented in their weekly diaries pointed to both positive and negative social experiences, ranging from feelings of companionship and motivation to frustration due to the perceived lack of human-like interaction. These findings correspond to similar themes in previous literature, which has examined the role of chatbots in fostering social comfort, emotional bonding, and anthropomorphism—the attribution of human characteristics, emotions, or intentions to non-human entities—in language learning contexts (Du & Daniel, 2024; Koç, 2024).

One prominent benefit emphasized by learners was the supportive presence and emotional warmth they experienced when interacting with ChatGPT. Participants described the chatbot as non-judgmental, encouraging, and understanding—qualities that contributed to a stress-free speaking environment. For instance, learners noted that they did not feel anxious about making mistakes or being negatively evaluated, which in turn enhanced their self-confidence and willingness to speak. This emotional safety is particularly significant in EFL contexts, where classroom speaking anxiety is often a persistent challenge (Hapsari & Wu, 2022; Koç, 2024). These findings are in line with previous research highlighting the role of AI chatbots in creating psychologically safe environments for language learners (Chen et al., 2020; Du & Daniel, 2024; Fathi et al., 2024; Koç, 2024; Liu et al., 2023; Wu et al., 2023). In these studies, learners frequently emphasized that talking to the chatbot helped lower their anxiety. Since the conversations felt free of judgment, they felt more comfortable speaking and weren't as afraid of making mistakes.

Learners also shared that, over the course of their weekly interactions with ChatGPT, they began to form an emotional connection with it. In their diaries, they often used expressions like "supportive presence," "comforting tone," and "feeling understood," pointing to a strong emotional bond. This affective side of the experience closely mirrors what Koç (2024) observed with Replika. In their study, participants described the chatbot as friendly, patient, and even comforting. It was further noted that learners felt more at ease and less anxious while speaking with the AI because it responded in a respectful and non-judgmental way, without cutting them off or making them feel self-conscious. This kind of stress-free atmosphere helped students open up more, speak with greater confidence, and become more willing to engage in language practice (pp. 223–226). Du and Daniel (2024) similarly observed that learners often start to view chatbots as emotionally intelligent companions—almost like human partners—capable of easing the sense of isolation that sometimes comes with learning a language on one's own (p. 9). This tendency, which they categorize as EO4 ("fostering a tendency to anthropomorphize"), was a common theme across the studies they reviewed.

These findings are also consistent with a broader body of research emphasizing the role of AI in enhancing learner engagement through perceived psychological support and companionship (Dwivedi et al., 2023; Wu et al., 2023; Liu et al., 2023; Chen et al., 2020; Fathi et al., 2024). In particular, Dwivedi et al. (2023) describe how students have begun to regard AI chatbots as collaborative partners or team members rather than mere tools, with some users even reporting feelings of relief and reassurance when engaging with such systems. This emotional scaffolding is especially significant in EFL contexts where anxiety and fear of negative evaluation can hinder spoken output.

Despite the emotional support many learners experienced, several participants also voiced disappointment regarding the chatbot's limited social intelligence. The code "lack of human-like emotional and personal responses" appeared prominently among social challenges in learner diaries. Learners expressed that while ChatGPT was often polite and friendly, it lacked genuine empathy, emotional nuance, or the ability to remember past conversations. These deficiencies weakened the perceived authenticity of interaction and led some learners to feel detached over time.

This echoes the concerns reported in Koç's study, where participants described the AI chatbot as "robotic," "depersonalized," or "unable to comprehend abstract concepts" (Koç, 2024, pp. 225–227). Similarly, Du and Daniel (2024) drew attention to a significant drawback, asserting that speech recognition systems used in chatbots are only capable of processing sentences up to a certain length with accuracy. Once the sentence length exceeds this threshold, the system's recognition capability starts to fall apart, posing a considerable challenge for sustaining effective communication (p. 9).

In parallel, Dwivedi et al. (2023) caution that although some users may develop emotional bonds with AI chatbots, the absence of genuine empathy and contextual memory often results in only a superficial sense of companionship. The chatbot's inability to show real human emotion or remember past conversations sometimes left learners feeling disconnected. This lack of continuity weakened the sense of a real social exchange, and this highlights that current limits of AI when it comes to social intelligence are still present. During the eight-week process, students initially expressed amazement at ChatGPT's responsiveness and conversational capabilities. However, as the intervention progressed, they increasingly voiced dissatisfaction with the chatbot's lack of authentic human interaction, highlighting its "not humanized" nature. Interestingly, conversations with the participants showed a gradual change in how they viewed their interactions with AI. Over time, students began to adjust their expectations, seeing ChatGPT more as a helpful tool or supportive assistant rather than

expecting it to fully act like a human. They started to value it for what it could offer in reaching their language learning goals, instead of treating it as a direct replacement for real-life conversation partners.

As a result, their initial complaints about the chatbot not feeling human enough began to fade. Still, many learners kept suggesting ways to make the interaction feel more natural and emotionally rich in their diaries, showing that the desire for a more human-like experience remained strong.

Lastly, given the mixture of positive and negative experiences, learners proposed several suggestions to improve the social realism and responsiveness of the chatbot. The most frequent suggestion in this category was the desire for authentic human-like interaction as expressed above. Students proposed enhancements such as more natural voice modulation, facial expressions, and continuity in conversations. These preferences are echoed in Koç's findings, where learners suggested the inclusion of more realistic avatars, gesture simulation, and memory retention features to foster stronger social bonds with the chatbot (Koç, pp. 227–228). Similarly, Du and Daniel (2024) suggested that future chatbot design should aim for increased personalization and contextual memory in order to enhance social presence and interaction continuity (p. 10).

Interestingly, learners in this study were also given the opportunity to interact with an advanced-voiced version of ChatGPT for 15 minutes per month. Several students were genuinely surprised by how closely the enhanced voice resembled that of a real person. This, of course, noticeably boosted their interest and motivation during interactions. They clearly stated a strong preference for having more frequent access to this professional-level voice feature, pointing out how its realistic and nuanced quality made a big difference. This suggests that future chatbot developers should seriously consider incorporating more advanced voice technologies into their systems, as a more lifelike voice can play an important key role in keeping learners engaged and making the experience feel more authentic.

Furthermore, learners also pointed out the need for more balanced conversations, saying that the interaction often felt one-sided because they had to do most of the work from starting dialogue to maintaining the dialogue on their own. This sense of imbalance led to frustration over time. A similar issue was reported by Koç (2024), where students experienced conversational fatigue due to the chatbot's lack of follow-up questions or its inability to guide the flow naturally (p. 212). In both studies, participants expressed a clear wish for more goal-driven, dynamic interactions that better reflect how real conversations unfold.

Generally speaking, the suggestions expressed by learners in this study, including the need for clearer feedback, culturally appropriate responses, language level adjustments, and better turn-taking, are in line with the REPOST model proposed by Koç (2024). As already outlined in the literature review, and now reinforced by learner feedback, this conceptual framework outlines six key characteristics an ideal pedagogical chatbot should possess: Reciprocity, Evaluation, Personalization, Omnipresence, Socio-cultural awareness, and Teachableness. These elements collectively emphasize the need for chatbots that not only sustain dialogic interaction and provide corrective feedback but also adapt to learners' proficiency levels, offer culturally relevant content, remain accessible across platforms, and are continuously trainable through user input and multimodal data. In this light, the REPOST model offers a practical and research-informed roadmap for future chatbot development in EFL contexts. The convergence between this study's learner-driven suggestions and the REPOST components highlights the growing consensus on the essential features needed to make AI chatbots genuinely pedagogically effective and socially responsive tools in language learning.

Overall, the findings in this section underscore the need for chatbot designs that can simulate more human-like interaction patterns, offer continuity in conversation, and respect learners' emotional boundaries. Future improvements in these dimensions could significantly enhance learner satisfaction and motivation in technology-assisted language learning environments.

# CHAPTER 6
# CONCLUSION

This study investigated the effects of GenAI chatbot use on Turkish young adult EFL learners' speaking proficiency and their perceptions regarding the pedagogical, technological, and social dimensions of AI-mediated speaking practice. Drawing on a quasi-experimental research design supported by qualitative analysis of student diaries and interviews, the study provided a multifaceted view of how generative AI tools such as ChatGPT can contribute to EFL speaking development. This chapter summarizes the findings in relation to the research questions and offers final reflections on the overall significance of the study.

## 6.1. Summary of Findings

The first research question looked into whether the experimental and control groups showed notable differences in their speaking proficiency scores after the intervention. Although both groups showed improvement over time, the difference between them was not statistically significant. However, the experimental group, which engaged in structured weekly interactions with ChatGPT, demonstrated a greater increase in speaking proficiency scores compared to the control group. These findings imply that even though speaking practice supported by GenAI may not always result in substantial improvements over a brief period, it can still serve as a valuable supplementary tool for enhancing speaking skills when properly structured and supported by learner participation. Still, it is essential to highlight that the relatively small sample size in this research makes it hard to generalize these results broadly. As such, the observed trends must be interpreted with caution and seen as exploratory rather than conclusive.

The second research question addressed within-group differences in speaking performance before and after the intervention. In this respect, both groups made statistically significant gains from pre-test to post-test. This means that speaking proficiency can or may improve over time regardless of the mode of practice. Notably, the improvement observed in the experimental group was more substantial, and this result supports the pedagogical value of meaningful interaction opportunities provided by the chatbot. These findings back up the idea that digital tools, if integrated thoughtfully into teaching practices, can boost learners' language output and help them gradually develop essential speaking skills.

The third research question examined learner perceptions, with sub-questions focusing on the perceived benefits, challenges, and suggestions related to AI chatbot use for speaking development. Thematic analysis of student diaries revealed a wide range of affordances. On the pedagogical side, students perceived the chatbot as a helpful tool for practicing spoken English, improving vocabulary, and receiving instant feedback. Many participants considered the

chatbot as a personal tutor—accessible beyond classroom hours—offering repetitive and safe spaces for language practice. On the technological front, students admired the chatbot mainly for being fast, easily accessible, and flexible. They liked how effortless it was to interact with it and how it was always ready for use. On the social side, participants mentioned that chatting with the bot made them feel less nervous compared to speaking tasks in the classroom. The AI partner was perceived as non-judgmental and emotionally neutral, encouraging freer expression and risk-taking in oral production.

At the same time, several limitations were identified. Technologically, students reported challenges related to repetition, lack of memory retention, and limitations in the chatbot's ability to provide corrective feedback. The quality of the chatbot's replies depended significantly on the input, and some students noticed misunderstandings or sudden shifts in conversation topics. From an educational perspective, learners sometimes found the chatbot's feedback too general or not instructive enough. They wanted a clearer structure, explicit grammar corrections, and more personalized guidance. On the social side, although interacting with the chatbot helped reduce speaking anxiety, it fell short in capturing the emotional depth and empathy typical of human conversations. Some learners even expressed wanting more natural, expressive dialogues that resemble everyday interactions.

Importantly, what stood out was that learners frequently shared ideas for improving their speaking experience with the AI, touching on both pedagogical and social aspects. They suggested things like customizing the chatbot's voice, making it more culturally aware, helping it remember past topics, and giving clearer corrections. These suggestions showed that students were not just looking for a chat buddy. Instead, they were hoping for a smart, flexible learning companion that could adapt to their needs.

**6.2. Implications for English Language Teachers**

The results of this study highlight the importance of providing structured and goal-oriented tasks when integrating GenAI chatbots into speaking practice. Teachers are encouraged to design weekly themes, prompts, and guiding questions that scaffold learners' interaction with the chatbot. As learners in this study indicated, such structure helped them stay focused, generate ideas, and apply classroom input in their conversations. Therefore, AI interaction should not be left entirely unstructured or informal; rather, it should be intentionally embedded into the curriculum as a supplementary tool that supports communicative objectives.

Moreover, teachers should provide learner training on how to use chatbots for effective language development. Some learners in this study seemed uncertain about how to start or keep

up meaningful practice sessions, which points to the need for clear guidance on both task goals and how to use the chatbot effectively.

However, to materialize this goal, teachers should first know what AI technologies are and how they can get the most benefit out of these technologies. An interesting study by Koraishi (2024) highlights this issue by focusing on cultivating AI literacy among in-service EFL teachers through an AI-enhanced language teaching workshop. The study revealed that while teachers entered the workshop with only moderate levels of AI literacy—especially in areas such as pedagogical application and ethics—the intervention led to notable improvements. This underscores the importance of preparing teachers with foundational and pedagogical knowledge of AI before expecting them to guide learners in effectively using tools like chatbots for language development.

Finally, to make AI-based speaking practice more beneficial, teachers might also add reflective elements. For example, weekly diaries can be implemented to help students become more aware of their own speaking habits and development. In the end, even though chatbots allow for independent practice, the teacher still plays a key role in guiding, supporting, and making sense of these experiences for learners.

**6.3. Implications for Chatbot Developers**

Learners in this study shared a number of concerns and suggestions. They aslo highlighted specific areas where chatbot design could be improved. One major takeaway is the need for clearer, more accurate, and context-aware feedback. Many participants said the chatbot often overlooked their grammar mistakes or didn't explain them well enough. To address this, developers should consider adding feedback tools that focus on language form—especially in tricky areas for L2 learners like verb tenses, articles, and pronunciation.

Another substantial concern raised by learners was the chatbot's inability to remember past conversations or maintain a sense of continuity. Even though the chatbot has the ability to retain past conversation to some extent, many students still said they wished the chatbot could recall previous chats and build on them later, which would make the experience feel more personalized and help topics evolve over time. Alongside this, learners also expressed a desire for more natural and emotionally rich interactions. To create a more socially engaging experience, developers should work on making the chatbot's voice more expressive, fine-tuning its response timing, and adding humanlike features that make it feel more empathetic and lifelike.

Finally, learners expressed a desire for customization options, including topic selection, proficiency level adjustment, and cultural relevance. Developers are encouraged to allow

greater user control over conversation parameters and to design chatbots that are responsive to the needs of different learner populations—especially lower-proficiency learners who may struggle with open-ended, unstructured exchanges.

**6.4 Suggestions for Future Research**

Given the limited sample size and context-specific nature of this study, future research is needed to explore the impact of GenAI chatbots on speaking proficiency in broader and more diverse educational settings. Larger-scale studies with longitudinal designs could provide more robust evidence of the long-term effects of AI-mediated speaking practice. Moreover, future research should investigate how different task types, interaction modes (e.g., text-based vs. speech-based), and feedback strategies influence learning outcomes.

There is also a need for further exploration into the socio-emotional dimensions of chatbot interaction. While this study touched on the improvement of speaking skills, future studies could use psychometric tools or observational data to better understand the affective shifts that accompany AI-assisted speaking practice. In addition, upcoming research could benefit from exploring hybrid teaching models that blend chatbot use with teacher support and feedback. This kind of approach may offer clearer insights into how AI tools and human instruction can complement each other to improve language learning outcomes.

All in all, while this study offers a starting point for understanding how learners experience GenAI chatbots in EFL speaking contexts, there still seems to be a lot of room for more detailed and in-depth studies. Future research could explore deeper into how these technologies can be fine-tuned to meet learners' needs not just pedagogically but also technologically and socially.

**6.5. Concluding Remarks**

Overall, this study offers valuable insight into the role of GenAI chatbots in supporting EFL speaking development. While the experimental findings did not yield statistically significant differences between groups, the combination of quantitative trends and qualitative evidence highlights the potential of AI chatbots to function as meaningful tools in EFL instruction—particularly when paired with scaffolded design and learner support. The use of structured prompts, guided topics, and reflective diary tasks in this study helped create a well-rounded learning experience, and it encouraged students to practice and reflect—even beyond the classroom walls.

This research also adds to the ongoing discussion about AI's place in language education. As students become more familiar with using digital tools in their everyday lives, it is crucial that teaching methods evolve to include technologies that promote learner motivation

and engagement. It highlights the need for AI-powered activities that go beyond simple back-and-forth conversations, and activities should also steer learners toward real, meaningful communication skills. As interest in generative AI grows in the field of language learning, it can be vital to keep a strong balance between embracing new tech and staying true to solid teaching principles. That way, technology becomes a true asset in supporting learners' growth.

## REFERENCES


Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning With Applications*, *2*, 100006. https://doi.org/10.1016/j.mlwa.2020.100006

Ai, H. (2017). Providing graduated corrective feedback in an intelligent computer-assisted language learning environment. *ReCALL*, *29*(3), 313–334. https://doi.org/10.1017/s095834401700012x

Al-Kadi, A. (2018). A Review of Technology Integration in ELT: From CALL to MALL. *Language Teaching and Educational Research*, *1*(1), 1-12. https://www.researchgate.net/publication/342447253

Al-Murtadha, M. (2019). Enhancing EFL learners ' willingness to communicate with visualization and goal-setting activities. *Tesol Quarterly, 53*(1), 133–157. https://doi.org/10.1002/tesq.474

Al-Obaydi, L. H., Pikhart, M., & Klimova, B. (2023). ChatGPT and the general concepts of education: Can Artificial Intelligence-Driven Chatbots support the process of language learning? *International Journal of Emerging Technologies in Learning (iJET)*, *18*(21), 39–50. https://doi.org/10.3991/ijet.v18i21.42593

Ali, F., Choy, D., Divaharan, S., Tay, H., & Chen, W. (2023). Supporting self-directed learning and self-assessment using TeacherGAIA, a generative AI chatbot application: Learning approaches and prompt engineering. *Learning: Research and Practice*, 9, 135 - 147. https://doi.org/10.1080/23735082.2023.2258886

Amaral, L. and Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL, 23*(1), 4-24. https://doi.org/10.1017/s0958344010000261

Azizimajd, H. (2023). Investigating the impacts of voice-based student-chatbot interactions in the classroom on EFL learners' oral fluency and foreign language speaking anxiety. *Technology Assisted Language Education, 1*(2), 61–83. https://doi.org/10.22126/tale.2023.2732

Bandura, A. (2002). Social Cognitive Theory in Cultural context. *Applied Psychology*, *51*(2), 269–290. https://doi.org/10.1111/1464-0597.00092

Bozkurt, A. (2023). Unleashing the potential of generative AI, conversational agents and chatbots in educational praxis: A Systematic Review and Bibliometric Analysis of

GenAI in Education. *Open Praxis*, *15*(4), 261–270. https://doi.org/10.55982/openpraxis.15.4.609

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge.

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2303.04226

Chang, D., Lin, M., Hajian, S., & Wang, Q. (2023). Educational Design Principles of Using AI Chatbot That Supports Self-Regulated Learning in Education: Goal Setting, Feedback, and Personalization. *Sustainability, 15*(17), 12921. https://doi.org/10.3390/su151712921

Chen, H. H., Yang, C. T., & Lai, K. K. (2020). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments*, *31*(3), 1335–1350. https://doi.org/10.1080/10494820.2020.1833043

Chen, X., Xie, H., Zou, D., & Hwang, G. (2020). Application and theory gaps during the rise of Artificial Intelligence in Education. *Computers and Education Artificial Intelligence*, *1*, 100002. https://doi.org/10.1016/j.caeai.2020.100002

Chiu, T. K. F. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: a case of ChatGPT. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-024-10366-w

Choi, I. (2014). Efficacy of an ICALL tutoring system and process-oriented corrective feedback. *Computer Assisted Language Learning*, *29*(2), 334–364. https://doi.org/10.1080/09588221.2014.960941

Cogo, A., Patsko, L., & Szoke, J. (2024). Generative artificial intelligence and ELT. *ELT Journal*, *78*(4), 373–377. https://doi.org/10.1093/elt/ccae051

Cohen, G., & Cohen, A. (2024). Motivational GenAI chatbot for promoting SRL in problem solving: Data-driven design principles. *Ubiquity Proceedings*, 25. https://doi.org/10.5334/uproc.147

Colby, K. M., Weber, S., & Hilf, F. D. (1971). Artificial paranoia. *Artificial Intelligence*, *2*(1), 1–25. https://doi.org/10.1016/0004-3702(71)90002-6

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage publications.

Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, *22*(5), 811–817. https://doi.org/10.1017/s1351324916000243

Dewaele, J.-M., & Pavelescu, L. M. (2021). The relationship between incommensurable emotions and willingness to communicate in English as a Foreign Language: A multiple case study. *Innovation in Language Learning and Teaching, 15*(1), 66–80. https://doi.org/10.1080/17501229.2019.1675667

Dizon, G. (2017). Using intelligent personal assistants for second language learning: a case study of Alexa. *TESOL Journal*, *8*(4), 811–830. https://doi.org/10.1002/tesj.353

Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language, Learning and Technology*, *24*(1), 16–26. https://doi.org/10125/44705

Dizon, G. (2021). Affordances and constraints of intelligent personal assistants for second-language learning. *RELC Journal*. Article 003368822110205. https://doi.org/10.1177/00336882211020548

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education Artificial Intelligence*, *6*, 100230. https://doi.org/10.1016/j.caeai.2024.100230

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., . . . Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Eales, F., & Oakes, S. (2015). *SpeakOut intermediate student's book* (2nd ed.). Pearson Education.

Eke, D. O. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology*, *13*, 100060. https://doi.org/10.1016/j.jrt.2023.100060

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, *121*, 103254. https://doi.org/10.1016/j.system.2024.103254

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *ELT Forum Journal of English Language Teaching*, *12*(1), 44–58. https://doi.org/10.15294/elt.v12i1.64069

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066x.34.10.906

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology, 10*(3), 8–14. https://doi.org/10125/44068

Fryer, L.K., Coniam, D., Carpenter, R., & Lăpușneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology*, *24* (2), 8-22. http://hdl.handle.net/10125/44719

García-Peñalvo, F. J. (2023). La percepción de la Inteligencia Artificial en contextos educativos tras el lanzamiento de ChatGPT: disrupción o pánico. *Education in the Knowledge Society (EKS)*, *24*, e31279. https://doi.org/10.14201/eks.31279

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism, 10*(2), 486–489. https://doi.org/10.5812/ijem.3505

Ginns, P., & Leppink, J. (2019). Special Issue on Cognitive Load Theory: Editorial. *Educational Psychology Review*, *31*, 255-259. https://doi.org/10.1007/S10648-019-09474-4

Given, L. M. (2008). Purposive sampling. In *The SAGE encyclopedia of qualitative research methods*. SAGE Publications. https://doi.org/10.4135/9781412963909.n349

Given, L. M., & Saumure, K. (2008). Trustworthiness. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods*. Sage. https://doi.org/10.4135/9781412963909.n470

Godwin-Jones, R. (2021). Big data and language learning: Opportunities and challenges. *Language Learning & Technology, 25* (1), 4-19. http://hdl.handle.net/10125/44747

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective*. Pearson Education International.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). Sage.

Han, D. (2020). The effects of voice-based AI chatbots on Korean EFL middle school students' speaking competence and affective domains. *Asia-pacific Journal of Convergent Research Interchange*, *6*(7), 71–80. https://doi.org/10.47116/apjcri.2020.07.07

Hapsari, I. P., & Wu, T. (2022). AI Chatbots learning model in English speaking skill: alleviating speaking anxiety, boosting enjoyment, and fostering critical thinking. In *Lecture notes in computer science* (pp. 444–453). https://doi.org/10.1007/978-3-031-15273-3_49

Harry, A. (2023). Role of AI in education. *Interdisciplinary Journal of Humanity (INJURITY), 2*(3), 260–268. https://doi.org/10.58631/injurity.v2i3.52

Heller, B., Procter, M., Mah, D., Jewell, L., & Cheung, B. (2005). Freudbot: An investigation of chatbot technology in distance education. In *EdMedia + Innovate Learning* (pp. 3913–3918). Association for the Advancement of Computing in Education (AACE). https://www.researchgate.net/publication/242084006

Hsu, M. H., Chen, P. S., & Yu, C. S. (2023). Proposing a task-oriented chatbot system for EFL learners speaking practice. *Interactive Learning Environments*, *31*(7), 4297-4308. https://doi.org/10.1080/10494820.2021.1960864

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

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, *26*(1), 112–131. https://www.jstor.org/stable/48707971

Hutson, M. (2021). Robo-writers: the rise and risks of language-generating AI. *Nature*, *591*(7848), 22–25. https://doi.org/10.1038/d41586-021-00530-0

Hwang, G., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education Artificial Intelligence*, *1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Hwang, W., Guo, B., Hoang, A., Chang, C., & Wu, N. (2022). Facilitating authentic contextual EFL speaking and conversation with smart mechanisms and investigating

its influence on learning achievements. *Computer Assisted Language Learning*, *37*(7), 1632–1658. https://doi.org/10.1080/09588221.2022.2095406

Ifelebuegu, A. O., Kulume, P., & Cherukut, P. (2023). Chatbots and AI in Education (AIEd) tools: The good, the bad, and the ugly. *Journal of Applied Learning & Teaching*, *6*(2). https://doi.org/10.37074/jalt.2023.6.2.29

Jamshed, M., Allehyani, B., Albedah, F., & Banu, S. (2024). Analyzing ChatGPT's sustainability as a language learning tool: Saudi EFL learners' perspectives, problems and pedagogical prospects. *Theory and Practice in Language Studies, 14*(12), 3836–3844. https://doi.org/10.17507/tpls.1412.18

Jensen, D. (2008). Credibility. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods*. Sage. https://doi.org/10.4135/9781412963909.n77

Jia, J., & Chen, W. (2008). The further development of CSIEC project driven by application and evaluation in English education. *British Journal of Educational Technology*, *40*(5), 901–918. https://doi.org/10.1111/j.1467-8535.2008.00881.x

Jia, J., Chen, W. (2008). Motivate the Learners to Practice English through Playing with Chatbot CSIEC. In: Pan, Z., Zhang, X., El Rhalibi, A., Woo, W., Li, Y. (eds.) Technologies for E-Learning and Digital Entertainment. Edutainment 2008. Lecture Notes in Computer Science, vol 5093. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69736-7_20

Kara, E., Ayaz, A. D., & Dündar, T. (2017). Challenges in EFL speaking classes in Turkish context. *European Journal of Language and Literature*, *3*(2), 66-74. https://doi.org/10.26417/ejls.v8i1.p66-74

Kim, H., Cha, Y. & Kim, N. Y. (2021). Effects of AI chatbots on EFL students' communication skills. *Korean Journal of English Language and Linguistics 21*, 712-734. https://doi.org/10.15738/kjell.21..202108.712

Kim, Na-Young, Cha, Yoonjung, & Kim, Hea-Suk. (2019). Future English learning: Chatbots and artificial intelligence. *Multimedia-Assisted Language Learning, 22*(3), 32-53. https://doi.org/10.15702/mall.2019.22.3.32

Kim, S., Eun, J., Seering, J., & Lee, J. (2021). Moderator Chatbot for deliberative discussion. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–26. https://doi.org/10.1145/3449161

Koç, F. Ş. (2024). *The development of listening and speaking skills in EFL via an artificially intelligent chatbot application: A quasi-experimental design study* (Publication No.

11511/107856) [Doctoral dissertation, Middle East Technical University]. METU. https://hdl.handle.net/11511/107856

Koç, F. Ş., & Savaş, P. (2024). The use of artificially intelligent chatbots in English language learning: A systematic meta-synthesis study of articles published between 2010 and 2024. *ReCALL*, 1–18. https://doi.org/10.1017/s0958344024000168

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Kong, S., & Yang, Y. (2024). A Human-Centered learning and Teaching framework using generative Artificial intelligence for Self-Regulated learning development through domain knowledge learning in K–12 settings. *IEEE Transactions on Learning Technologies*, *17*, 1588–1599. https://doi.org/10.1109/tlt.2024.3392830

Koraishi, O. (2025). *Cultivating AI literacy among educators through an AI-enhanced language teaching workshop* [Master's thesis, Bursa Uludağ University]. Bursa Uludağ University Institutional Repository. https://acikerisim.uludag.edu.tr/entities/publication/da32f065-ba31-4299-af99-f57f0e4400d4

Köse, D. (2004). Yabancı Dil Öğretiminde Ortak Ölçütler ve TÖMER. *Dil Dergisi*, *125*, 24-34. https://doi.org/10.1501/DILDER_0000000018

Kostka, I., & Toncelli, R. (2023). Exploring Applications of ChatGPT to English Language teaching: Opportunities, challenges, and recommendations. *Teaching English as a Second or Foreign Language--TESL-EJ*, *27*(3). https://doi.org/10.55593/ej.27107int

Kulaksız, G. C. (2024). *Artificial intelligence-based language modelling: The effect of ChatGPT application on writing skills in the context of teaching English as a foreign language* (Master's thesis, Bursa Uludağ University). Bursa Uludağ University Institutional Repository**.** https://acikerisim.uludag.edu.tr/entities/publication/68ddad61-a55a-401a-8729-df5fc3c13eb7

Lai, J. (2024). Adapting Self-Regulated Learning in an Age of Generative Artificial Intelligence Chatbots. *Future Internet*, 16, 218. https://doi.org/10.3390/fi16060218

Lee, H., Kang, J., & Yeo, J. (2021). Medical Specialty recommendations by an artificial intelligence chatbot on a smartphone: development and deployment. *Journal of Medical Internet Research*, *23*(5), e27460. https://doi.org/10.2196/27460

Lee, J. H., Yang, H., Shin, D., & Kim, H. (2020). Chatbots. *ELT Journal*. *74*(3), 338–344. https://doi.org/10.1093/elt/ccaa035

Liao, J., Zhong, L., Zhe, L., Xu, H., Liu, M., & Xie, T. (2024). Scaffolding computational thinking with ChatGPT. *IEEE Transactions on Learning Technologies*, *17*, 1668–1682. https://doi.org/10.1109/tlt.2024.3392896

Lin, C.-J., & Mubarok, H. (2021). Learning analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. *Educational Technology & Society, 24*, 16–35. https://www.jstor.org/stable/48629242

Lin, M. P.-C., & Chang, D. (2020). Enhancing Post-secondary Writers' Writing Skills with a Chatbot: A Mixed-Method Classroom Study. *Journal of Educational Technology & Society*, *23*(1), 78–92. https://www.jstor.org/stable/26915408

Liu, M., Zhang, L. J., & Biebricher, C. (2023). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Computers & Education*, *211*, 104977. https://doi.org/10.1016/j.compedu.2023.104977

Mahn, H. (1999). Vygotsky's Methodological contribution to Sociocultural Theory. *Remedial and Special Education*, *20*(6), 341–350. https://doi.org/10.1177/074193259902000607

Marietto, M. D. G. B., Aguiar, R. V., De Oliveira Barbosa, G., Botelho, W. T., Pimentel, E., Franca, R. D. S., & Da Silva, V. L. (2013). Artificial Intelligence Markup Language: A brief tutorial. *International Journal of Computer Science & Engineering Survey*, *4*(3), 1–20. https://doi.org/10.5121/ijcses.2013.4301

Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, *10*(2). https://doi.org/10.1080/2331186x.2023.2236469

McTear, M. (2020). Conversational AI: dialogue systems, conversational agents, and chatbots. *Synthesis Lectures on Human Language Technologies*, *13*(3), 1–251. https://doi.org/10.2200/s01060ed1v01y202010hlt048

Merriënboer, J., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development*, *53*, 5-13. https://doi.org/10.1007/BF02504793

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Mohamed, A.M. (2024). Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: perceptions of EFL Faculty Members. *Educ Inf Technol 29*, 3195–3217. https://doi.org/10.1007/s10639-023-11917-z

Molnar, G., & Szuts, Z. (2018). The Role of Chatbots in Formal Education. *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)*. https://doi.org/10.1109/sisy.2018.8524609

Moussalli, S., & Cardoso, W. (2016). Are commercial 'personal robots' ready for language learning? Focus on second language speech. In S. Papadima-Sophocleous, L. Bradley, & S. Thouësny (Eds.), *CALL communities and culture – short papers from EUROCALL 2016* (pp. 325–329). Research-publishing.net. https://doi.org/10.14705/rpnet.2016.eurocall2016.583

Mungai, B., Omieno, P., Egessa, M., & Manyara, P. (2024). AI chatbots in LMS: A pedagogical review of cognitive, constructivist, and adaptive principles. *Engineering and Technology Journal*. https://doi.org/10.47191/etj/v9i08.15

Ng, D., Tan, C., & Leung, J. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *Br. J. Educ. Technol.*, 55, 1328-1353. https://doi.org/10.1111/bjet.13454

OpenAI. (2024, May 13). *GPT-4o system card*. https://openai.com/index/gpt-4o-system-card/

Özçelik, N. P., & Ekşi, G. Y. (2024). Cultivating writing skills: the role of ChatGPT as a learning assistant—a case study. *Smart Learning Environments*, *11*(1). https://doi.org/10.1186/s40561-024-00296-8

Öztürk, G., & Gürbüz, N. (2014). Speaking anxiety among Turkish EFL learners: The case at a state university. Journal of Language and Linguistic Studies, 10(1), 1-17.

Pack, A., & Maloney, J. (2023). Using Generative Artificial Intelligence for Language Education Research: Insights from Using OpenAI's ChatGPT. *TESOL Quarterly*, *57*(4), 1571–1582. https://doi.org/10.1002/tesq.3253

Peng, J.-E. (2020). Teacher interaction strategies and situated willingness to communicate. *ELT Journal, 74*(3), 307–317. https://doi.org/10.1093/elt/ccaa012

Pintrich, P. R. (2000). The role of Goal orientation in Self-Regulated Learning. In *Elsevier eBooks* (pp. 451–502). https://doi.org/10.1016/b978-012109890-2/50043-3

Qasem, F., Ghaleb, M., Mahdi, H. S., Al Khateeb, A., & Al Fadda, H. (2023). Dialog chatbot as an interactive online tool in enhancing ESP vocabulary learning. *Saudi Journal of Language Studies, 3*(2), 76–86. https://doi.org/10.1108/SJLS-10-2022-0072

Rostami, M., & Abadi, P. (2023). The Impact of Doing Assignments with Chatbots on The Students' Working Memory. *Health Nexus*. https://doi.org/10.61838/hn.1.1.10

Santos, A., Olesova, L., Vicentini, C., & Oliveira, L. (2023, March). *ChatGPT in ELT: Writing affordances and activities* [Conference presentation]. TESOL 2023

International Convention & English Language Expo, Portland, OR, United States. https://www.researchgate.net/publication/376688011_ChatGPT_in_ELT_Writing_Affordances_and_Activities

Schmidt, T., & Strasser, T. (2022). Artificial intelligence in foreign language learning and teaching: a CALL for intelligent practice. Anglistik: International Journal of English Studies, 33(1), 165–184.

Schreiber, J. B. (2008). Pilot study. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 625–626). SAGE Publications. https://doi.org/10.4135/9781412963909.n320

Sharadgah, T. A., & Sa'di, R. A. (2022). A systematic review of research on the use of artificial intelligence in English language teaching and learning (2015-2021): What are the current effects? *Journal of Information Technology Education: Research*, *21*, 337-377. https://doi.org/10.28945/4999

Shooshtari, Z. G., & Mir, F. (2014). ZPD, Tutor; Peer Scaffolding: Sociocultural Theory in Writing Strategies Application. *Procedia - Social and Behavioral Sciences*, *98*, 1771–1776. https://doi.org/10.1016/j.sbspro.2014.03.605

Shumanov, M., & Johnson, L. (2021). Making conversations with chatbots more personalized. *Computers in Human Behavior*, *117*, Article 106627. https://doi.org/10.1016/j.chb.2020.106627

Skulmowski, A., & Xu, K. (2021). Understanding Cognitive Load in Digital and Online Learning: a New Perspective on Extraneous Cognitive Load. *Educational Psychology Review*, *34*, 171 - 196. https://doi.org/10.1007/s10648-021-09624-7

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education*, *151*, 103862. https://doi.org/10.1016/j.compedu.2020.103862

Song, C. & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1260843

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

Tai, T., & Chen, H. H. (2020). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, *31*(3), 1485–1502. https://doi.org/10.1080/10494820.2020.1841801

Tanveer, M. (2007). *Investigation of the factors that cause language anxiety for ESL/EFL learners in learning speaking skills and the influence it casts on communication in the target language* [Master's thesis, University of Glasgow]. https://doi.org/10.13140/RG.2.1.1995.1129

Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science*, *379*(6630), 313. https://doi.org/10.1126/science.adg7879

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*(1). https://doi.org/10.1186/s40561-023-00237-x

Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, *LIX*(236), 433–460. https://doi.org/10.1093/mind/lix.236.433

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

Uria, L., Maritxalar, M., & Zabala, I. (2014). An environment for learner corpus research and error analysis. *International Journal of Computer-Assisted Language Learning and Teaching*, *4*(3), 34–51. https://doi.org/10.4018/ijcallt.2014070103

Van Dis, E. a. M., Bollen, J., Zuidema, W., Van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, *614*(7947), 224–226. https://doi.org/10.1038/d41586-023-00288-7

Wang, C., Wang, Y., & Zou, B. (2024). Revolutionising EFL pedagogy: Innovative strategies for integrating GAI (ChatGPT) into language teaching. *Journal of Language Teaching*, *4*(1), 1–7. https://doi.org/10.54475/jlt.2024.004

Wheeler, S. (2015). *Learning futures: Emerging trends in educational technology* [Keynote presentation]. Institute for Adult Learning, Singapore. https://www.ial.edu.sg/getmedia/1c2658eb-a21f-4e3c-a597-ff9ab49b8055/Keynote5_Steve-Wheeler.pdf

Wu, C., Lam, H., Kong, A., & Wong, W. (2023). The design and evaluation of a digital learning-based English chatbot as an online self-learning method. *International Journal of Engineering Business Management*, *15*. https://doi.org/10.1177/18479790231176372

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, *55*, 10–33. https://doi.org/10.1111/bjet.13334

Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL, 34*(3), 327–343. https://doi.org/10.1017/s0958344022000039

Yin, J., Goh, T., & Hu, Y. (2024). Using a Chatbot to Provide Formative Feedback: A Longitudinal Study of Intrinsic Motivation, Cognitive Load, and Learning Performance. *IEEE Transactions on Learning Technologies*, 17, 1404-1415. https://doi.org/10.1109/TLT.2024.3364015

Yin, J., Goh, T., Yang, B., & Xiaobin, Y. (2020). Conversation Technology with Micro-Learning: The Impact of Chatbot-Based Learning on Students' learning motivation and performance. *Journal of Educational Computing Research*, *59*(1), 154–177. https://doi.org/10.1177/0735633120952067

Yin, Q., & Satar, M. (2020). English as a foreign language learner interaction with chatbots: Negotiation for meaning. *International Online Journal of Education and Teaching (IOJET), 7*(2), 390–410. http://iojet.org/index.php/IOJET/article/view/707

Zadorozhnyy, A., & Lai, W. (2023). ChatGPT and L2 Written Communication: A Game-Changer or Just Another Tool?. *Languages*. https://doi.org/10.3390/languages9010005

Zadrozny, W., Budzikowska, M., Chai, J., Kambhatla, N., Levesque, S., & Nicolov, N. (2000). Natural language dialogue for personalized interaction. *Communications of the ACM*, *43*(8), 116–120. https://doi.org/10.1145/345124.345164

Zhai, C., & Wibowo, S. (2022). A systematic review on cross-culture, humor and empathy dimensions in conversational chatbots: the case of second language acquisition. *Heliyon*, *8*(12), e12056. https://doi.org/10.1016/j.heliyon.2022.e12056

Zhai, X. (2022). ChatGPT User Experience: Implications for Education. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4312418

Zhang, M., & Li, J. (2021). A commentary of GPT-3 in MIT Technology Review 2021. *Fundamental Research*, *1*(6), 831–833. https://doi.org/10.1016/j.fmre.2021.11.011

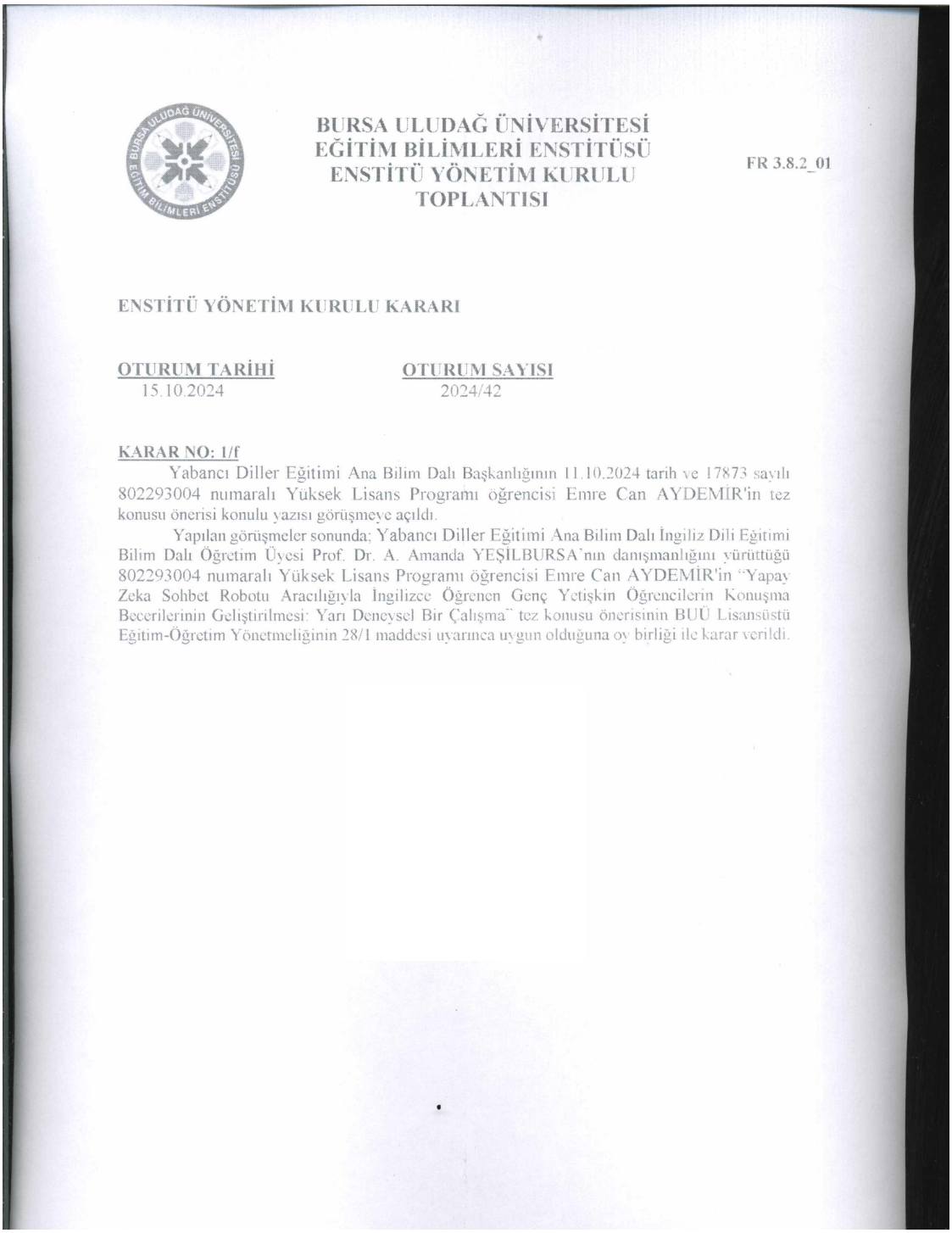
Zhang, Z., & Huang, X. (2024). The impact of chatbots based on large language models on second language vocabulary acquisition. *Heliyon*, *10*(3), e25370. https://doi.org/10.1016/j.heliyon.2024.e25370

Zhi, R., & Wang, Y. (2024). On the relationship between EFL students' attitudes toward artificial intelligence, teachers' immediacy and teacher-student rapport, and their willingness to communicate. *System*, *124*, 103341. https://doi.org/10.1016/j.system.2024.103341

Zimmerman, B. J. (2000). Attaining Self-Regulation. In *Elsevier eBooks* (pp. 13–39). https://doi.org/10.1016/b978-012109890-2/50031-7

# APPENDICES

## APPENDIX 1: Institutional Research Approval Document



BURSA ULUDAĞ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
ENSTİTÜ YÖNETİM KURULU
TOPLANTISI

FR 3.8.2_01

ENSTİTÜ YÖNETİM KURULU KARARI

OTURUM TARİHİ
15.10.2024

OTURUM SAYISI
2024/42

KARAR NO: 1/f

Yabancı Diller Eğitimi Ana Bilim Dalı Başkanlığının 11.10.2024 tarih ve 17873 sayılı 802293004 numaralı Yüksek Lisans Programı öğrencisi Emre Can AYDEMİR'in tez konusu önerisi konulu yazısı görüşmeye açıldı.

Yapılan görüşmeler sonunda; Yabancı Diller Eğitimi Ana Bilim Dalı İngiliz Dili Eğitimi Bilim Dalı Öğretim Üyesi Prof. Dr. A. Amanda YEŞİLBURSA'nın danışmanlığını yürüttüğü 802293004 numaralı Yüksek Lisans Programı öğrencisi Emre Can AYDEMİR'in "Yapay Zeka Sohbet Robotu Aracılığıyla İngilizce Öğrenen Genç Yetişkin Öğrencilerin Konuşma Becerilerinin Geliştirilmesi: Yarı Deneysel Bir Çalışma" tez konusu önerisinin BUÜ Lisansüstü Eğitim-Öğretim Yönetmeliğinin 28/1 maddesi uyarınca uygun olduğuna oy birliği ile karar verildi.

**T.C.**
**BURSA ULUDAĞ ÜNİVERSİTESİ**
**Eğitim Bilimleri Enstitüsü Müdürlüğü**

Sayı: E-20585590-000-17957 18.10.2024
Konu: Tez Konusu Önerisi (Emre Can AYDEMİR)

YABANCI DİLLER EĞİTİMİ ANABİLİM DALI BAŞKANLIĞINA

İlgi : 11.10.2024 tarihli ve 56643073-000-17873 sayılı yazınız.

Anabilim Dalınız 802293004 numaralı Yüksek Lisans Programı öğrencisi Emre Can AYDEMİR'e ait bildirilen tez konusu önerisine ilişkin alınan karar örneği ekte gönderilmiştir.

Bilgilerinizi ve kararın öğrenci ile danışmanı Prof. Dr. A. Amanda YEŞİLBURSA'ya bildirilmesi konusunda gereğini rica ederim.

Prof. Dr. Gökhan ARI
Müdür

Ek:
Karar Örneği

Bu belge, güvenli elektronik imza ile imzalanmıştır.
Belge Doğrulama Kodu:Z_ntQjfpxEmyE1YfedwBog Belge Doğrulama Adresi: https://udos.uludag.edu.tr/Teyit/
BUÜ Eğitim Bilimleri Enstitüsü Görükle Kampüsü 16059 Nilüfer/BURSA
Telefon No: 02242942884 Faks No: 0(224)294 0975
e-Posta: fdemirci@uludag.edu.tr İnternet Adresi: uludag.edu.tr/egitimbilimleri
Kep Adresi: uludag.rektorluk@hs03.kep.tr
Bilgi için: Filiz DEMİRCİ
Memur
Telefon No: 02242942884
Bu belge UDOS ile hazırlanmıştır.

# APPENDIX 2: Ethics Committee Approval Certificate

**T.C.**
**BURSA ULUDAĞ ÜNİVERSİTESİ**
**Eğitim Bilimleri Enstitüsü Müdürlüğü**

Sayı: E-20585590-044-18819 03.01.2025
Konu: Emre Can AYDEMİR'in araştırma izni.

YABANCI DİLLER EĞİTİMİ ANABİLİM DALI BAŞKANLIĞINA

İlgi : 18.10.2024 tarihli ve 56643073-300-17941 sayılı yazınız.

Ana Bilim Dalınız yüksek lisans programı öğrencisi Emre Can AYDEMİR'in tez çalışmasına ilişkin Ankara Üniversitesi Rektörlüğü Türkçe ve Yabancı Dil Uygulama ve Araştırma Merkezi Müdürlüğünden alınan yazı örneği ekte gönderilmiştir.

Bilgilerinizi ve öğrenci ile danışmanı Prof. Dr. Ayşegül Amanda YEŞİLBURSA'ya bildirilmesi konusunda gereğini rica ederim.

Prof. Dr. Gökhan ARI
Müdür

Ek:
Yazı Örneği

Bu belge, güvenli elektronik imza ile imzalanmıştır.
Belge Doğrulama Kodu:twPRZB0JwUGS2eP13o_tvQ Belge Doğrulama Adresi: https://udos.uludag.edu.tr/Teyit/
BUÜ Eğitim Bilimleri Enst. Üniversite 1 Cd. No: 221 Görükle Kampüsü 16059 Nilüfer/BURSA
Telefon No: 0224 29 42 884 Faks No: 0224 29 40 975 Bilgi için: Halil AYDOĞAN
e-Posta: egtbil@uludag.edu.tr İnternet Adresi: uludag.edu.tr/egitimbilimleri Bilgisayar İşletmeni
Kep Adresi: uludag.rektorluk@hs03.kep.tr Telefon No: 0224 29 40 978
Bu belge UDOS ile hazırlanmıştır.

T.C.
ANKARA ÜNİVERSİTESİ REKTÖRLÜĞÜ
Türkçe ve Yabancı Dil Uygulama ve Araştırma Merkezi Müdürlüğü
İngilizce Koordinatörlüğü

Sayı : E-38268124-622.01-1646653
Konu : Emre Can AYDEMİR'in Araştırma İzni

ÖĞRENCİ İŞLERİ DAİRE BAŞKANLIĞINA

İlgi : 07.11.2024 tarihli ve E-94390400-302.08.01-187230 sayılı yazınız.

İlgide kayıtlı yazınıza istinaden İngiliz Dili Eğitimi Yüksek Lisanas Programı öğrencisi Emre Can AYDEMİR'in "Yapay Zeka Sohbet Robotu Aracılığıyla İngilizce Öğrenen Genç Yetişkin Öğrencilerin Konuşma Becerilerinin Geliştirilmesi: Yarı Deneysel Bir Çalışma" konulu tez çalışması Merkezimizce uygun görülmüştür.

Gereği için bilgilerinize saygılarımla arz ederim.

Prof. Dr. Timur GÜLTEKİN
Müdür

Bu belge, güvenli elektronik imza ile imzalanmıştır.
Doğrulama Kodu: EF77851E-FEE9-4850-A78F-34E6A55B9353 Doğrulama Adresi: https://www.turkiye.gov.tr/ankara-universitesi-ebys

Ankara Üniversitesi Beşevler 10. Yıl Yerleşkesi Ord.Prof.Dr. Şevket Aziz Kansu
Binası Emniyet Mahallesi İncitaşı Sokak No :10 Beşevler / ANKARA
Telefon No: 0(312) 214 13 50 Belge Geçer No: 0312 600 01 31 Kep Adresi:
ankunvrek@ankuni.hs01.kep.tr
www.tomer.ankara,edu.tr
KEP Adresi : ankunvrek@ankuni.hs01.kep.tr

Bilgi için:Seher KILIÇ
Sekreter

## APPENDIX 3: Experimental Group Consent and Demographic Information Form

**Bilgilendirilmiş Onay Formu**

Sayın Katılımcı,

Bursa Uludağ Üniversitesi Yabancı Diller Eğitimi Bölümü'nde yüksek lisans öğrencisiyim. "Yapay Zeka Sohbet Robotu Aracılığıyla İngilizce Öğrenen Genç Yetişkin Öğrencilerin Konuşma Becerilerinin Geliştirilmesi: Yarı Deneysel Bir Çalışma" başlıklı yüksek lisans tez çalışmamı yürütüyorum. Bu araştırmanın amacı, İngilizce öğrenen Türk öğrencilerin yapay zeka destekli sohbet botu uygulamaları konusundaki görüşlerini elde etmektir. Sohbet botları, kullanıcıların metin veya ses yoluyla İngilizce dilinde iletişim kurmalarına yardımcı olan Otomatik Konuşma Tanıma ve Doğal Dil İşleme teknolojilerine sahip iletişim asistanlarıdır. Katılımınız, bu dönemin başında ve sonunda bir konuşma sınavına katılmayı içerecektir. Ayrıca bir anket sorularını yanıtlayacak; sekiz hafta boyunca bir yapay zeka sohbet botu ChatGPT kullanımına katılacak ve konuşmalarınızı kaydedeceksiniz. Süreç boyunca deneyimlerinizi yansıtan haftalık yansıtıcı günlükler yazacak ve hepsinin sonunda İngilizce sohbet botu kullanımıyla ilgili bir görüşmeye katılacaksınız.Bu araştırmaya katılımınız gönüllüdür. Kimliğiniz ve paylaştığınız veriler bu çalışmadan elde edilecek herhangi bir yayında açıklanmayacaktır. Ayrıca istediğiniz zaman çalışmadan herhangi bir kayıp yaşamadan çekilme hakkına sahipsiniz ve bu durumda bilgileriniz bu çalışmaya dahil edilmeyecektir. Bu çalışma hakkında herhangi bir sorunuz varsa, lütfen e-posta yoluyla benimle iletişime geçin. İşbirliğiniz için çok teşekkür ederim.

Saygılarımla,

Emre Can Aydemir, Yüksek Lisans Adayı

E-mail:

Bu onay formunu okudum ve sorular sorma fırsatım oldu. Bu çalışmaya katılmayı kabul ediyorum.

Katılımcı/Veli İmzası: Tarih:

Lütfen ilgili kutuyu işaretleyin veya gerekli olduğunda cevabınızı belirtin.

| | |
|---|---|
| 1. Yaş: | 13-17 |
| | 18-21 |
| | 22-26 |
| | 27-31 |

| | |
|---|---|
| | 32-36 |
| | 37+ |
| 2. Şu an eğitim gördüğünüz yer. (lise, üniversite vb.) | |
| 3. Şu anda hangi seviyede eğitim alıyorsunuz? | Birinci Sınıf |
| | İkinci Sınıf |
| | Üçüncü Sınıf |
| | Dördüncü Sınıf |
| 4. Ne kadar süredir İngilizce öğreniyorsunuz? | 0-5 yıl |
| | 5-10 yıl |
| | 10+ yıl ve üstü |
| 5. Cinsiyet: | Erkek |
| | Kadın |
| 6. Sohbet botları hakkında bilgi sahibi misiniz? | Yes |
| | No |
| 7. Daha önce bir yapay zeka sohbet botu kullandınız mı? "Eğer evet ise, lütfen 8. ve 9. soruları da yanıtlayın. Eğer hayır ise, 8. ve 9. soruları atlayın. | Yes |
| | No |
| 8. Şu anda yapay zeka destekli sohbet botlarını (örneğin, Siri, Cortana, Google Asistan, Amazon Echo/Alexa vb.) İngilizce öğreniminde ne sıklıkla kullanıyorsunuz? | Asla |
| | Nadiren |
| | Ara Sıra |
| | Sıklıkla |
| 9. Şimdiye kadar hangi yapay zeka sohbet botlarını kullandınız? | Siri |
| | Alexa |
| | Google Assistant |
| | Cortana/Copilot |
| | Diğer |
| 10. Yapay zeka sohbet botları hakkında ne kadar bilgili olduğunuzu düşünüyorsunuz? | Oldukça İyi Bilgili |
| | İyi Bilgili |
| | Yeterince Bilgili |
| | Yetersiz Bilgili |
| | Bilgisiz |

## APPENDIX 4: Control Group Consent and Demographic Information Form

**Bilgilendirilmiş Onay Formu**

Sayın Katılımcı,

Bursa Uludağ Üniversitesi Yabancı Diller Eğitimi Bölümü'nde lisansüstü öğrencisiyim. Genç Yetişkin İngilizce Öğrencilerinin Konuşma Becerilerini geliştirmesiyle alakalı bir yüksek lisans tez çalışması yürütüyorum. Bu araştırmanın amacı, İngilizce öğrenen Türk öğrencilerin İngilizce konuşma becerilerini ve genel dil öğrenme deneyimlerini etkileyen çeşitli faktörler hakkında fikir edinmektir. Katılımcı olarak katılımınız, 8 haftalık süreçte hem başında hem de sonunda bir konuşma sınavını tamamlamayı içermektedir. Geri bildirimleriniz ve yanıtlarınız, öğrencilerin dil öğrenme deneyimleri ve konuşma becerisi gelişimi hakkında daha derin bir anlayışa katkıda bulunacaktır. Bu çalışmaya katılım tamamen gönüllülük esasına dayanmaktadır ve sağladığınız tüm veriler gizli kalacaktır. Kimliğiniz ve paylaştığınız bilgiler bu çalışmadan kaynaklanan hiçbir yayında kullanılmayacaktır. Ek olarak, herhangi bir sonuç doğurmadan çalışmadan çekilme hakkını saklı tutarsınız ve bunu seçerseniz toplanan hiçbir veri çalışmaya dahil edilmeyecektir. Çalışmayla ilgili herhangi bir sorunuz varsa lütfen e-posta yoluyla benimle iletişime geçmekten çekinmeyin. İş birliğiniz için çok teşekkür ederim.

Emre Can Aydemir, Yüksek Lisans Adayı
E-mail:

Bu onay formunu okudum ve sorular sorma fırsatım oldu. Bu çalışmaya katılmayı kabul ediyorum.

Katılımcı İmzası: Tarih:

Lütfen ilgili kutuyu işaretleyin veya gerekli olduğunda cevabınızı belirtin.

| | |
|---|---|
| 1. Yaş: | 13-17 |
| | 18-21 |
| | 22-26 |
| | 27-31 |

| | |
|---|---|
| | 32-36 |
| | 37+ |
| 2. Şu an eğitim gördüğünüz yer. (lise, üniversite vb.) | |
| 3. Şu anda hangi seviyede eğitim alıyorsunuz? | Birinci Sınıf |
| | İkinci Sınıf |
| | Üçüncü Sınıf |
| | Dördüncü Sınıf |
| 4. Ne kadar süredir İngilizce öğreniyorsunuz? | 0-5 yıl |
| | 5-10 yıl |
| | 10+ yıl ve üstü |
| 5. Cinsiyet: | Erkek |
| | Kadın |

# APPENDIX 5: Speaking Rubrics – Oral Presentation and Spoken Interaction

## Oral Presentation (40)

**ORAL PRESENTATION (40-Point Scale)**
**CRITERIA AND ASSESSMENT RUBRIC**

| Criteria & Weighting | Grammar (8) | Vocabulary (8) | Fluency (8) | Coherence (8) | Pronunciation (8) | Total (40) |
|---|---|---|---|---|---|---|
| STUDENT'S NAME | ☐ *0 points*: The student makes more than half of their sentence structure, word, and suffix choices incorrectly.<br>☐ *4 points*: The student makes some mistakes in suffixes, word usage, and sentence structure, but errors are limited.<br>☐ *8 points*: The student makes almost no grammatical mistakes. | ☐ *0 points*: The student does not use any words expected at their level.<br>☐ *4 points*: The student uses only a limited number of words expected at their level.<br>☐ *8 points*: The student uses almost all of the words expected at their level. | • *0 points*: The student's speech is completely non-fluent.<br>• *4 points*: The student hesitates occasionally, but overall, their speech is fluent.<br>• *8 points*: The student speaks confidently and fluently. | • *0 points*: The student's sentences are not logically connected at all.<br>• *4 points*: The sentences are somewhat connected but lack strong logical structure.<br>• *8 points*: The sentences are logically well-connected and highly relevant to the topic. | • *0 points*: The student mispronounces most words.<br>• *4 points*: The student mispronounces some words.<br>• *8 points*: The student pronounces almost all words correctly. | |
| 1 | | | | | | |
| 2 | | | | | | |
| 3 | | | | | | |
| 4 | | | | | | |
| 5 | | | | | | |
| 6 | | | | | | |
| 7 | | | | | | |
| 8 | | | | | | |
| 9 | | | | | | |
| 10 | | | | | | |
| 11 | | | | | | |
| 12 | | | | | | |
| 13 | | | | | | |
| 14 | | | | | | |

**Spoken Interaction 60**

*SPOKEN INTERACTION*

CRITERIA AND ASSESSMENT RUBRIC (60-Point Scale)

| Criteria & Weighting<br><br>Student's Name | Grammar (8) | Vocabulary (12) | Fluency (12) | Listening Skills (8) | Communication (8) | Sustaining the Dialogue (20) | Total Score (60) |
|---|---|---|---|---|---|---|---|
| 1 | ☐ *0 points:* The student frequently makes mistakes and does not use new grammatical structures.<br>☐ *3 points:* The student makes errors when using new grammar structures, but their speech is fairly understandable.<br>☐ *8 points:* The student makes no significant grammar mistakes. | • *0 points:* The student does not use any new vocabulary.<br>• *4 points:* The student uses a very limited number of new words.<br>• *8 points:* The student uses a sufficient number of new words.<br>• *12 points:* The student uses a variety of new words. | • *0 points:* There is no fluency in the student's speech.<br>• *4 points:* The student speaks slowly and frequently hesitates.<br>• *8 points:* The student hesitates occasionally but speaks naturally.<br>• *12 points:* The student speaks fluently and confidently. | • *0 points:* The listening student does not understand what the speaking student is saying.<br>• *3 points:* The listening student understands only after several repetitions.<br>• *8 points:* The listening student understands questions and statements easily on the first attempt. | • *0 points:* The student is not active during the conversation.<br>• *3 points:* The student tries to be active but is less engaged than their partner.<br>• *8 points:* The student is highly engaged and communicates well with their partner. | • *0 points:* There is little to no discussion or interaction.<br>• *4 points:* The discussion and communication are insufficient.<br>• *8 points:* The discussion and communication are adequate.<br>• *12 points:* The discussion and communication are excellent. | |
| 2 | | | | | | | |
| 3 | | | | | | | |
| 4 | | | | | | | |
| 5 | | | | | | | |
| 6 | | | | | | | |
| 7 | | | | | | | |
| 8 | | | | | | | |
| 9 | | | | | | | |
| 10 | | | | | | | |
| 11 | | | | | | | |
| 12 | | | | | | | |
| 13 | | | | | | | |
| 14 | | | | | | | |
| 15 | | | | | | | |
| 16 | | | | | | | |
| 17 | | | | | | | |

## APPENDIX 6: Student Weekly Reflective Diary Template

**REFLECTIVE DIARY PROMPT**

**Name Surname:**_____ **Date:** _____

**Week No: 5** **Topic of the Week: Future Goals & Career Aspirations**

**Student reflection diaries**

**Instructions:** The topic of discussion for this week is "future goals and career aspirations". By using the prompts you learned in the class, talk to your AI chatbot companion, ChatGPT, for at least 20 minutes. You need to bring your reflection diary to the class by Saturday. Please remember to record your conversation with ChatGPT using voice recorders available on your devices. Please share your recordings and conversation links with your teacher on Whatsapp. Upon the finish of your conversation, write in your weekly diary about the following aspects:

How was your general experience in talking to AI chatbot ChatGPT? Please state atleast three strengths, three challenges, and three suggestions for improvement.

Strengths:

1. _____
2. _____
3. _____

Challenges:

1. _____
2. _____
3. _____

Suggestions for improvement:

1. _____
2. _____
3. _____

How was your experience in taking turns in speech while speaking to ChatGPT thisweek?

Please state at least three strengths, three challenges and three suggestions forimprovement.

Strengths:

1. _____

2. _____

3. _____

Challenges:

1. _____

2. _____

3. _____

Suggestions for improvement:

1. _____

2. _____

3. _____

How would you rate the effectiveness of your experience in interacting with the AIchatbot this week? Please circle one number.

0 1 2 3 4 5 6 7 8 9 10

*Not Effective* *Extremely Effective*

**Week 5: Future Goals & Career Aspirations**

The aim is for students to explore their hopes and plans for the future, discussing what they want to achieve in their personal and professional lives. Here are prompts to guide these conversations with ChatGPT:

1. **Long-Term Goals**
   - "What are your long-term goals? One of mine is to [goal, e.g., travel around the world or learn a new language]."
   - "What's something you hope to achieve in your lifetime? I'd like to accomplish [goal]."
2. **Dream Career**
   - "If you could have any job, what would it be? I'd love to be a [career, e.g., doctor or engineer] someday."
   - "What job do you think would make you happiest? I think being a [career] would be fulfilling for me."
3. **Skills to Develop**
   - "Are there any skills you're working on to help you in the future? I'm trying to improve my [skill, e.g., public speaking]."
   - "What's a skill you think is important for success? I believe [skill] is crucial."
4. **Education & Studies**
   - "What are your academic goals? I want to study [subject] in the future."
   - "Do you know what you want to study in college? I'm considering [field of study]."
5. **Personal Goals**
   - "Besides work or school, do you have any personal goals? One of mine is to become better at [activity]."
   - "What's something you want to accomplish outside of your career? I want to focus on [personal goal]."
6. **Role Models and Inspirations**
   - "Who inspires you to achieve your goals? For me, it's [role model]."
   - "Do you have a role model or mentor who has helped shape your goals? I admire [person]."
7. **Challenges and How to Overcome Them**
   - "What challenges do you think you'll face on the way to your goals? I think my biggest challenge might be [challenge]."
   - "How do you plan to handle obstacles in your path? I try to stay positive by [strategy]."
8. **Balancing Career and Personal Life**
   - "How important is work-life balance to you? I think it's important because [reason]."
   - "What do you think is the best way to balance personal time with career goals?"
9. **Success and Happiness**
   - "What does success mean to you? For me, it's about [definition of success]."
   - "Do you think being successful is the same as being happy? I think [perspective]."
10. **Vision for the Future**

- "Where do you see yourself in ten years? I hope to be [describe future vision]."
- "What kind of life do you want to create for yourself? I want to have a future where I can [describe goal]."

## APPENDIX 7: Semi-Structured Interview Questions

### Yarı Yapılandırılmış Görüşme Soruları

1. ChatGPT adlı yapay zeka sohbet botunu kullanma deneyiminiz nasıldı?
2. İngilizce dil öğreniminizi desteklemek için yapay zeka sohbet botlarının avantajları hakkında ne düşünüyorsunuz?
3. ChatGPT'yi kullanırken İngilizce konuşma becerilerinizi geliştirme açısından size en çok fayda sağlayan özellikler nelerdir?
4. İngilizceyi yabancı dil olarak öğrenirken ChatGPT'yi kullanırken karşılaştığınız zorluklar nelerdir?
5. Yapay zekâ destekli bir İngilizce konuşma asistanı uygulamasının kelime bilgisi, gramer, dinleme, konuşma, telaffuz, tonlama pratiği ve kültür açısından hangi özelliklere sahip olması gerektiğini düşünüyorsunuz?
6. Aşağıdaki özelliklerin bir sohbet botunda bulunmasının önemi hakkında ne düşünüyorsunuz? Her bir özellik için görüşlerinizi ve gelecekteki EFL (Yabancı Dil olarak İngilizce) öğrenimi için sohbet botlarının geliştirilmesiyle ilgili yorumlarınızı belirtiniz:
    - Yazım hatalarını yönetebilme
    - Tam oluşmamış soruları anlayabilme
    - Yanlış kelime sırasına sahip cümleleri işleyebilme
    - Konuşmacılar tarafından bir kerede söylenen birden fazla cümleyi anlayabilme
    - Bir konunun akışını takip edebilme
    - Konuşma sırasında geniş bir kelime dağarcığı kullanabilme
    - Önceki bir yoruma yanıt verebilme
7. İngilizce Dil Öğrenimi için etkili bir sohbet botunun özellikleri nelerdir? Lütfen, en az üç özellik belirterek nedenlerinizi açıklayın.

### Geriye Bakmak ve İleriye Yönelmek

- Bu program sırasında başarı hissi duyduğunuz bir durum var mıdır?
- İngilizcenizi geliştirmeye devam etmek için bir sonraki adımınız nedir?

# CURRICULUM VITAE

| ÖZ GEÇMİŞ | | |
|---|---|---|
| Adı-Soyadı | Emre Can Aydemir | |
| Bildiği Yabancı Diller | İngilizce | |
| Eğitim Durumu | Başlama - Bitirme | Kurum Adı |
| Lise | 2013-2017 | Cavit Çağlar Mesleki Ve Teknik Anadolu Lisesi |
| Lisans | 2017-2021 | Trakya Üniversitesi<br>İngilizce Öğretmenliği Bölümü |
| Yüksek Lisans | 2022-2025 | Bursa Uludağ Üniversitesi Eğitim Bilimleri Enstitüsü Yabancı Diller Ana Bilim Dalı İngiliz Dili Eğitimi Bilim Dalı |
| Doktora | | |
| Çalıştığı Kurum | Başlama - Ayrılma | Çalışılan Kurumun Adı |
| 1. | 2021-2023 | Ukla Academy Yabancı Dil Kursu |
| 2. | 2024- Günümüze | Ankara Üniversitesi Tömer Bursa Şubesi |

Yayınlar:

Aydemir, E. C., & Yeşilbursa, A. (2025, April). Harnessing AI Chatbots for EFL Speaking Practice: Student Perceptions and Recommendations. Paper presented at the 22nd INGED International ELT Conference, Bolu Abant İzzet Baysal University. Published in the Book of Abstracts, p. 46. Retrieved from https://www.inged2025.com