



The Literacy and Attitude of English Language Teaching

Students Towards ChatGPT in Türkiye

Yüksek Lisans Tezi

Umut CEYLAN

Eskişehir 2025

**THE LITERACY AND ATTITUDE OF ENGLISH LANGUAGE TEACHING
STUDENTS TOWARDS CHATGPT IN TRKIYE**

Umut CEYLAN



YKSEK LİSANS TEZİ
Yabancı Diller Ana Bilim Dalı
İngilizce Eğitimi Programı
Danışman: Prof. Dr. Özgr YILDIRIM
Anadolu niversitesi
Lisansst Eğitim Enstits
Mayıs 2025
Eskiřehir

JÜRİ VE ENSTİTÜ ONAYI

Umut CEYLAN'ın "Türkiye'deki İngilizce Öğretmenliği Öğrencilerinin Yapay Zeka Bazlı Büyük Dil Modellerine Karşı Yeterlilikleri ve Tutumları" başlıklı tezi 09/05/2025 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca Yabancı Diller Ana Bilim dalında İngilizce Eğitimi Programında Yüksek Lisans tezi olarak kabul edilmiştir.

	<u>Ünvanı Adı Soyadı</u>	<u>İmza</u>
Prof. Dr. Özgür YILDIRIM (Danışman):		
Prof. Dr. Ali MERÇ:		
Dr. Öğr. Üy. Mustafa SARIOĞLU:		

.....

Prof.Dr. Nafiz Öncü CAN
Lisansüstü Eğitim Enstitüsü Müdürü

FINAL APPROVAL FOR THESIS

This thesis titled “The Literacy and Attitude of English Language Teaching Students Towards ChatGPT in Türkiye” prepared and submitted by Umut CEYLAN in partial fulfillment of the requirements in “Code of Graduate School of Anadolu University” for the Degree of Master of Science in the Department of Foreign Language Education has been examined and approved on 09/05/2025.

	Title, Name Surname	Signature
Prof. Dr. Özgür YILDIRIM (Supervisor):		
Prof. Dr. Ali MERÇ:		
Dr. Öğr. Üy. Mustafa SARIOĞLU:		

.....

Prof.Dr. Nafiz Öncü CAN
Dean of Graduate School

ABSTRACT

THE LITERACY AND ATTITUDE OF ENGLISH LANGUAGE TEACHING STUDENTS TOWARDS CHATGPT IN TÜRKİYE

Umut CEYLAN

Department of Foreign Language Education

Programme in English Language Teaching

Graduate School of Anadolu University, Mayıs 2025

Supervisor: Professor Özgür YILDIRIM

The development of large language models has affected many fields, with one of them being language education. Contemporary technology has always been a part of English learning and teaching, from the invention of the computer to all the way to and beyond the release of ChatGPT. With ChatGPT naturally being integrated into the field, there remain many gaps in literature around many aspects of English education and large language models. To fill one of these gaps, the main aim of this research is to find out the ChatGPT literacy and ChatGPT attitude of ELT students studying in a state university in Western Türkiye, and to find out if there are correlations between the two variables in addition to other variables such as university year, ChatGPT awareness, gender, and so on. To fulfill these aims, adapted ChatGPT Literacy Scale (S. Lee & Park, 2024), adapted TAME-ChatGPT (Abdaljaleel et al., 2024), ChatGPT ELT Attitude Survey was utilized for 141 participants in addition to semi-structured group interviews for 8 participants in total. The results reveal that there are various significant correlations between ChatGPT literacy and attitude, alongside significant correlations between university year, ChatGPT awareness, gender, whether students have taken compulsory ELT courses with the two main variables. It is concluded that ChatGPT literacy should be taught to language students and teachers, and teachers, students alike should be made aware of the positive and negative possibilities of large language models.

Keywords: Large language models in English education, ChatGPT literacy, ChatGPT attitude, Teacher education and large language models, contemporary technological tools in English education

ÖZET

TÜRKİYE'DEKİ İNGİLİZCE ÖĞRETMENLİĞİ ÖĞRENCİLERİNİN YAPAY ZEKÂ BAZLI BÜYÜK DİL MODELLERİNE KARŞI YETERLİLİKLERİ VE TUTUMLARI

Umut CEYLAN

Yabancı Diller Eğitimi Anabilim Dalı

İngilizce Eğitimi Programı

Anadolu Üniversitesi Lisansüstü Eğitim Enstitüsü, May 2025

Danışman: Profesör Özgür YILDIRIM

Dil eğitimi de dahil olmak üzere büyük dil modellerini gelişimi birçok alanı etkilemiştir. Çağdaş teknoloji bilgisayarın icadından ChatGPT'nin geliştirilmesine ve sonrasına kadar her zaman İngilizce öğrenmenin ve öğretmenin parçası olmuştur. ChatGPT'nin doğal olarak alana entegre olmasıyla beraber, İngilizce eğitimi ve büyük dil modelleri arasında hala literatürde birçok boşluk bulunmaktadır. Bu boşluklardan birini kapatmak için, bu tez Batı Türkiye'deki bir devlet üniversitesi okuyan İngilizce öğretmenliği öğrencilerinin ChatGPT okuryazarlığı ve tutumuna odaklanmıştır. Bu araştırmanın ana hedefi öğrencilerin ChatGPT okuryazarlığını ve ChatGPT tutumunu bulmak ve üniversite yılı, ChatGPT farkındalığı, cinsiyet, ve benzer değişkenlerle ilişkileri yanı sıra bu iki ana değişkenin arasında korelasyon olup olmadığını bulmaktır. Bu hedefler için adapte edilmiş ChatGPT okuryazarlık ölçeği (S. Lee & Park, 2024), adapte edilmiş TAME-ChatGPT ölçeği (Abdaljaleel et al., 2024), ChatGPT ELT tutum anketi 141 katılımcı üzerinde uygulanmıştır ve aynı zamanda yarı yapılandırılmış görüşme soruları 8 öğrenci üzerinde kullanılmıştır. Sonuçlar ChatGPT okuryazarlığı ve tutumu arasında birçok farklı anlamı korelasyonun yanı sıra, aynı zamanda üniversite yılı, ChatGPT farkındalığı, cinsiyet, ve öğrencilerin zorunlu İngilizce öğretmenliği dersleri alıp almamaları ile bu iki ana değişken arasında da anlamlı korelasyon olduğunu göstermiştir. Sonuç olarak, dil öğretmenlerine ve öğrencilerine ChatGPT okuryazarlığının öğretilmesi gerektiği ve hem öğretmenlerin, hem de öğrencilerin büyük dil modellerinin olumlu ve olumsuz potansiyelleri hakkında farkındalık kazanmaları gerektiği sonucuna varılmıştır.

Anahtar Sözcükler: İngiliz eğitiminde büyük dil modelleri, ChatGPT okuryazarlığı, ChatGPT tutumu, Öğretmen eğitiminde büyük dil modelleri, İngiliz eğitiminde çağdaş teknolojik araçlar

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to the people without whom this thesis would not be possible.

I would first like to express my deepest gratitude to my supervisor Prof. Dr. Özgür YILDIRIM for his supervision and guidance throughout the various steps of this thesis and study. Without his feedback, guidance, and support, this thesis would not be possible.

I would like to thank my thesis committee members for their time, effort, and feedback.

I would also like to express my gratitude for the lecturers and students at the institution I work in.

Most importantly, I would like to express my deepest of gratitudes and thanks to my mother, Esra YILDIRIM, without whom I would not have the background, discipline, and drive to realize this thesis, and anything else I have done, achieved, and strive to achieve in the course of my life.

I would also like to express my deepest gratitude to my friends, Alperen İÇKE, Efe MERİÇ, Ege YENİİZ, Harun Can ERGİN, and Ubeyde Utku İŞMAN for their support. Without them, this thesis would not have been completed.

Umut CEYLAN

Eskişehir, 2025

STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES

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

Umut Ceylan

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Anadolu Üniversitesi tarafından kullanılan “Bilimsel İntihal Tespit Programı”yla tarandığını ve hiçbir şekilde “intihal içermediğini” beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

Umut Ceylan

ÜRETKEN YAPAY ZEKA KULLANIM BEYANI

Bu tezi hazırlarken üretken yapay zekâ programlarından destek almadığımı beyan ederim. Tezin hazırlığı aşamasında üretken yapay zekâ programlarını araştırma konusu yapay zekâ programları olduğu için araştırma içinde kullanmak için yapay zekâ programlarından örnek metinler kullandım. Bu durumun dışında yapay zekâ programlarını destek için kullanmadım.

Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

Umut Ceylan

.../.../20...

DECLARATION OF GENERATIVE ARTIFICIAL INTELLIGENCE USAGE

I declare that I did not receive support from generative artificial intelligence programs while preparing this thesis. I have used generative artificial intelligence programs to receive output and text to present in the study for the reason that this study is focused on generative artificial intelligence programs. Other than that purpose, I did not use generative artificial intelligence programs for support.

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

Umut Ceylan

TABLE OF CONTENTS

ABSTRACT.....	iii
ÖZET	iv
ACKNOWLEDGEMENTS	v
LIST OF TABLES	xiii
LIST OF FIGURES	xvi
LIST OF ABBREVIATIONS	xvii
1. INTRODUCTION	1
1.1. Statement of the Problem	6
1.2. Aims and Research Questions.....	7
1.3. Significance of the Study	8
2. REVIEW OF LITERATURE.....	11
2.1. Introduction.....	11
2.2. The Emergence of AI	11
2.3. The History of Chatbots	14
2.4. ChatGPT	19
2.5. AI and Chatbots in Education	27
2.5.1. AI and chatbots in language education	32
2.5.2. ChatGPT in language education.....	41
2.5.2.1. ChatGPT literacy in language education.....	50
2.5.2.2. ChatGPT attitude in language education	53
3. METHODOLOGY	60
3.1. Introduction.....	60
3.2. Research Design	60
3.3. Research Context	63
3.4. Participants.....	65
3.5. Instruments.....	68

3.5.1.	ChatGPT Literacy Scale	69
3.5.2.	TAME-ChatGPT	74
3.5.3.	ChatGPT ELT Attitude Survey	79
3.5.4.	Interview.....	80
3.6.	Procedure	82
3.7.	Analysis of the Data	84
3.8.	Ethical Declaration	88
4.	RESULTS AND DISCUSSION	89
4.1.	Introduction.....	89
4.2.	Results	89
4.2.1.	ChatGPT Literacy Scale	90
4.2.2.	TAME-ChatGPT	99
4.2.3.	ChatGPT ELT Attitude Survey	108
4.2.4.	Interviews	115
4.2.4.1.	Teaching language skills.....	116
4.2.4.2.	Evaluation	118
4.2.4.3.	Lesson/material design	120
4.2.4.4.	Professional development	124
4.2.4.5.	Ethical considerations.....	126
4.2.4.6.	Outlook as a teacher	128
4.2.4.7.	Future of AI	130
4.2.5.	Comparison.....	132
4.3.	A Brief Summary of the Results	139
4.4.	Discussion of the Findings	141
5.	CONCLUSIONS AND IMPLICATIONS	153
5.1.	Introduction.....	153
5.2.	An Overview of the Present Study.....	153

5.3. Implications	154
5.4. Limitations of the Study	161
5.5. Suggestions for Further Studies.....	163
REFERENCES	166
APPENDICES.....	1
ÖZ GEÇMİŞ	12



LIST OF TABLES

	<u>Page</u>
Table 2.1. Performance of ChatGPT-3, ChatGPT-4, And Humans Across Problems with Varying Difficulty Levels.....	25
Table 2.2. Content Analysis of the Interview About the Participants' Views.....	47
Table 2.3. Mean and Standard Deviation for the Sub-Groups of the Study.....	52
Table 2.4. One-Way Analysis of Variance (ANOVA) Results Showing the Values of Correlation Between Teachers' Digital Competency and Their Teaching Experience...	54
Table 3.1. Frequency Table of The Participants.....	66
Table 3.2. Table for Which Data Collection Tool Corresponds to Which Research Question.....	87
Table 4.1. Descriptive Statistics of the Adapted ChatGPT Literacy Scale Results.....	91
Table 4.2. Bonferroni Post-hoc Results for University Levels for ChatGPT Literacy Scale Scores.....	95
Table 4.3. Bonferroni Post-hoc Results for University Levels for Technical Proficiency Scores.....	96
Table 4.4. Bonferroni Post-hoc Results for University Levels for Critical Evaluation Scores.....	96
Table 4.5. Bonferroni Post-hoc Results for University Levels for Communication Proficiency Scores.....	97
Table 4.6. Bonferroni Post-hoc Results for University Levels for Creative Application Scores.....	97
Table 4.7. Bonferroni Post-hoc Results for University Levels for Ethical Competence Scores.....	98
Table 4.8. Descriptive Statistics of the Adapted TAME-ChatGPT Results.....	100
Table 4.9. Bonferroni Post-hoc Results for University Levels for the Adapted TAME-ChatGPT Survey Score.....	102

Table 4.10. Bonferroni Post-hoc Results for University Levels for the ‘Heard’ Section Scores.....	103
Table 4.11. Bonferroni Post-hoc Results for University Levels for the ‘Used’ Section Scores.....	104
Table 4.12. Bonferroni Post-hoc Results for University Levels for Perceived Risk Scores.....	104
Table 4.13. Bonferroni Post-hoc Results for University Levels for Anxiety Scores.....	105
Table 4.14. Bonferroni Post-hoc Results for University Levels for Technology/Social Influences.....	106
Table 4.15. Bonferroni Post-hoc Results for University Levels for Perceived Usefulness Scores.....	106
Table 4.16. Bonferroni Post-hoc Results for University Levels for Behavior/Cognitive Factors Scores.....	107
Table 4.17. Bonferroni Post-hoc Results for University Levels for Ease of Use Scores.....	108
Table 4.18. Descriptive Statistics of the ELT Attitude Survey Results.....	109
Table 4.19. Bonferroni Post-hoc Results for University Levels for ELT Attitude Score.....	112
Table 4.20. Bonferroni Post-hoc Results for University Levels for Teaching Score.....	112
Table 4.21. Bonferroni Post-hoc Results for University Levels for Evaluation Score.....	113
Table 4.22. Bonferroni Post-hoc Results for University Levels for Lesson Design/Material Development Score.....	113
Table 4.23. Bonferroni Post-hoc Results for University Levels for Attitude Towards ChatGPT as A Teacher Score.....	114
Table 4.24. Information About the Interviews’ Length, Word Count, And Participants’ Number.....	116

Table 4.25. Coding Table for The Theme ‘Teaching Language Skills’ and Its Sub-Themes.....	117
Table 4.26. Coding Table for The Theme ‘Evaluation’ And Its Sub-Themes.....	119
Table 4.27. Coding Table for The Theme ‘Lesson/Material Design’ And Its Sub-Themes.....	122
Table 4.28. Coding Table for The Theme ‘Professional Development’ And Its Sub-Themes.....	125
Table 4.29. Coding Table for The Theme ‘Ethical Considerations’ and Its Sub-Themes.....	127
Table 4.30. Coding Table for The Theme ‘Outlook as A Teacher’ And Its Sub-Themes.....	129
Table 4.31. Coding Table for The Theme ‘Future Of AI’ and Its Sub-Themes.....	131
Table 4.32. Table for Research Questions and the Related Results.....	140

LIST OF FIGURES

	<u>Page</u>
Figure 2.1. Documents With “Chatbot” in the Article Name From 2010 to 2024 by Subject Area.....	18
Figure 2.2. Documents With the Word “Chatbot” in the Article Name From 2010 to 2024 by Year.....	19
Figure 2.3. The History of Generative AI in CV (Computer Vision), NLP (Natural Language Processing and VL (Vision Language).....	20
Figure 2.4. A Game of Blocks in Progress.....	28
Figure 2.5. The General Concept Of MICALI.....	34
Figure 2.6. Distribution Of Language Skills in Each Stage of Years on Research Concerning AI In Language Education.....	35
Figure 2.7. Published Literature on AI In Language Education In 1990-2020.....	36
Figure 2.8. Distribution Of Participant Educational Level in Each Stage” Of Years on Research Concerning AI In Language Education.....	36
Figure 2.9. Data Sources Used in The Studies.....	49
Figure 2.10. Using ChatGPT for ELT Required Technological Proficiency and Knowledge.....	55
Figure 2.11. Figures Showing the Rates of ChatGPT Usability Score and ChatGPT Performance Rating.....	58
Figure 3.1. Mixed Research Process Model.....	61
Figure 3.2. Concurrent Triangulation Design.....	63
Figure 3.3. Stages of the Delphi Technique.....	70
Figure 3.4. The Seven Steps in the Thematic Analysis.....	86
Figure 4.1. Frequency Graph of the Adapted ChatGPT Literacy Scale Scores.....	92
Figure 4.2. Correlation Coefficient Strength of Spearman Coefficient.....	133

LIST OF ABBREVIATIONS

AI	: Artificial Intelligence
AIML	: Artificial Intelligence Markup Language
AIED	: Artificial Intelligence in Education
ANOVA	: One-way Analysis of Variance
CAL	: Computer-Assisted Learning
CEFR	: Common European Framework of Reference
CSIEC	: Computer Simulation in Educational Communications
DARPA	: Defense Advanced Research Projects Agency
EFL	: English as a Foreign Language
ELT	: English Language Teaching
ESL	: English as a Second Language
FGA	: French Grammar Analyzer
FROG	: French Robust Grammar
GPT	: Generative Pre-trained Transformer
MANOVA	: Multivariate Analysis of Variance
MICALI	: Model for Intelligent Computer-Assisted Language Instruction
MMO	: Massively Multiplayer Online
MUD	: Multi-user Dungeon
SFL	: Systematic Functional Linguistics
YKS	: Yükseköğretim Kurumları Sınavı
YÖK	: Yükseköğretim Kurulu

1. INTRODUCTION

“Much has been said about how ChatGPT will disrupt the education sector” –

“I can only hope that this is the case.”

-Michael Wade (Dwivedi et al., 2023)

Technology has always been a part of education, and the experimentation of how emerging technologies can affect language education has been a source of study throughout its history (Eisenstein et al., 1987). With the widespread use of this new technology, emerges the notion for its use in education in various ways (Dwyer, 1999), therefore does the desire for scientific research over the possibilities and other variables regarding said new technology also. The need for more scientific data and results over these resources with their possibility of use in general education has always been integral to the literature and practical considerations (Dodds & Lefler, 1946).

The importance of technology in education shines particularly brightly and clearly in language education. Due to the importance of input (Krashen, 1989), cultural interaction and communication (Ware & Kramsch, 2005), and the overall cross-cultural and communication-related needs naturally found in teaching and learning a second and/or a foreign language, it is clear to see that the use of technology can be highly beneficial for the learning process and outcomes. In a considerable number of studies, the effects of integration of technological devices and materials into the learning and teaching process proved to be significantly beneficial for the variables controlled. For example, in a study done on 33 Thai English as a foreign language (EFL) undergraduate students regarding the effects of integrated performance assessment modules on their English-speaking ability, the comparison between the pretest and posttest displayed significant increase in speaking ability (Plengkham & Wasanasomsithi, 2023). A study separating the English learners into three different groups of “English teaching assisted by mobile learning technology (Rain Classroom), a social media tool (WeChat) and the traditional multimedia projecting system for one semester” (Z. Yu et al., 2022) with a total of 101 participants found that the posttest revealed significant differences in the three groups, with the Rain classroom providing the highest improvement among the groups, followed by, but still significantly lesser than it, the WeChat group. Similar studies exist that

demonstrate the significant improvement when technology is used in the learning process (Gu et al., 2021; Ma et al., 2024; Shimray, 2023).

Various studies varied in their findings regarding if there is a significant difference between the outcomes of the learners using the technological construct, and the control group. In a study with 120 college student participants, the use of watching captioned movie clips showed significant vocabulary development for the learners, as well as the control group; however, the amount of vocabulary development between the groups was not significantly different (Yüksel & Tanrıverdi, 2009). There have also been studies regarding different technology-based resources, such as e-learning, that showed significant improvement to learning compared to the control group (D. Zhang et al., 2004).

While positive outlook towards technology is seen in teachers, there are also other variables such as the technological competence of the learner and the teacher, appropriate use, classroom management, and various other concepts to look out for when regarding the use of technology in the classroom (Carstens et al., 2021).

However, the use of technology in the classrooms has its own challenges and considerations. Language teachers are seen to have a positive outlook on using technology in the classroom, but there were difficulties such as “technical problems”, “students’ attitudes towards using technology”, “lack of time”, “lack of computers and other devices”, and “lack of technical support” (Alamri, 2021). Similar issues can also be seen in teacher training. “Insufficient training, lack of basic facilities in the practicum schools, and student teachers’ own choices” were found to be the main challenges for pre-service teachers (Merç, 2015).

In addition to these examples, when it comes to the topic of technology use in and around the language learning classroom, one of the most discussed and researched topics of the past few years have been artificial intelligence (AI), especially generative AI chatbots such as ChatGPT (Mohamed, 2024). ChatGPT (Generative Pre-trained Transformer) is a conversational generative AI chatbot developed by OpenAI. It launched in 2022 and was updated in the following two years to improve in its various functionalities, and its development continues.

Even though ChatGPT’s improvement in response quality and overall accuracy compared to older hardware and software garnered attention from the general public,

chatbots as a concept and as functional software are not constructs created by this recent phenomenon. The first chatbot, Eliza, was developed in 1966 by Joseph Weizenbaum and is described as a system “which makes certain kinds of natural language conversation between man and computer possible” (Weizenbaum, 1966). In a simplified way, the program generated responses by looking for keywords, discovering as much context as it could, transform the response to be appropriate, and generate a response. The way ChatGPT works does not differ much from this first incarnation of the construct. ChatGPT uses deep learning and gets “trained on big datasets to generate predictions based on the data” (Haleem et al., 2022). The overall interest to these new era of language models might have come from the fact that ChatGPT could create new things from the prompts it is given. While older incarnations of chatbots could do so as well, the quality of the final product given by ChatGPT seemed to have passed the Turing Test (Turing, 1950) to the human cursory glance, or even that of software in certain cases, due to the ever-improving nature of the new technology (Farrelly & Baker, 2023). ChatGPT’s first prototype, GPT, and the details of how it used the data given to train itself and improve as a result was discussed by Radford, Narasimhan, Salimans, and Sutskever (2018).

With ChatGPT’s perceived ability of creating new material from its database and machine learning algorithm, comes the question of whether it could be used in and around the classroom. There are opportunities to incorporate the use of this large language model to the teachers’ many resources that help them in their teaching process. Possibilities for its use in lesson planning, research and writing, assessment and evaluation, and in most areas of the teaching and the learning process have been discussed (Kasneci et al., 2023). Its use in teaching higher order skills like critical thinking skills in higher education are studied to be beneficial (Y. Guo & Lee, 2023), as well as improvement in various variables in learners as young as elementary education level (Bai et al., 2024).

While the possibilities of this new technology piques the interests of researchers, practitioners, and students alike, there are also certain concerns raised with the use of ChatGPT, its limitations, and its perceived abilities. There is the threat that the information and overall output given by ChatGPT may be biased because of the data it is trained on (Lo et al., 2024). While the amount of data fed to its algorithm is vast, it still does not have the ability to pull information from every dataset available digitally, and

even less so if the information and data it is searching for has not been copied digitally from a physical source.

Furthermore, the issue of plagiarism in general education, and also in academia is also seen as a significant threat to academic integrity. Recent developments suggest that while there are ways of detecting AI-generated text via the use of AI detection and plagiarism detection tools, practitioners should be made aware of this new threat of plagiarism and should be taught the ways to detect when they are used (Cotton et al., 2024). In a study done by generating abstracts by prompting AI based on the title and journal of five real studies, while the ‘GPT-2 output detector’ software used as a detection tool detected the AI-generated abstracts correctly in most cases, when the same was asked to human reviewers, only 68% of the AI-generated abstracts were detected (Gao et al., 2023). While there are threats to originality and threats of plagiarism, ChatGPT also “has considerable power to advance” academia in both “anxiety-provoking and exciting new ways” if “how to use this technology responsibly and ethically” is considered carefully (Lund & Wang, 2023). Therefore, researchers and practitioners should be made well aware of the threats raised by this new technology, in addition to more and careful scrutiny by academic journals are being necessary to make sure of the “integrity of submitted manuscripts” (Shen et al., 2023).

In addition to being dependent on the data it is provided and the possibility of bias, the threats to plagiarism both by students, teachers, and academicians, there is also the threat of false information, and “hallucinations” (Ahmad et al., 2023). Working word after word by algorithmically working to find the best answer to the given prompt, ChatGPT, especially older versions such as GPT-3 can, in some cases, present false information, inexistant references, and made-up data (Alkaissi & McFarlane, 2023). In a study done on ChatGPT-3.5, it was found that “at best, only 75% of the references of an individual researcher were correct” in output generated by the AI (Metze et al., 2023). Therefore, taking ChatGPT as the sole source of information in research and education could lead to using false information and references, and this shortcoming of the chatbot should be taken into consideration when it is used.

The impact of ChatGPT on education and education research is apparent, be them positive or negative (Lo et al., 2024). As a tool to use in teaching English, the potential of ChatGPT is an area currently researched by many, but due to the recency of this

technological development, there are still many areas and aspects of teaching language and their interaction with large language models like ChatGPT yet to be researched, or in need of more data from different contexts. Attitude towards ChatGPT also is a topic with a need for more research. One study, by researching the perceptions of EFL faculty members towards ChatGPT in EFL teaching, displayed that faculty members attitude towards ChatGPT was positive, although there was still need for more research on the topic (Mohamed, 2024). As for EFL teachers, a study done on English language teachers in Thailand illustrate that from the teachers' perspectives, ChatGPT was seen as a "friend, supporter, help, text developer" but also could have and does have issues on the topics of "lack of authenticity, students' overdependence, academic dishonesty, unable to think critically, understanding ChatGPT, and responsible use of ChatGPT" (Ulla et al., 2023). Similar results were seen for university students, with plagiarism, fairness, and false information, and the accuracy of data are given as some of the concerns raised by the participants (Acosta-Enriquez et al., 2024). A different study done on 88 educational technology students in a university setting also showcased an overall positive opinion on ChatGPT (Güner et al., 2024).

While there are up-to-date research done on the attitude of English teachers (Mutammimah et al., 2024), and ChatGPT literacy of language teachers (Q. Ma et al., 2024), the amount of research done on prospective teachers studying to be English teachers are lower, if any; and in similar research, while the reasons for positive or negative outlook towards the use of ChatGPT both as a student and an educator are collected as data, the variables that affect the attitude of the participants are not the focal points of these studies. Furthermore, the perceptions towards the use of ChatGPT "may differ based on demographic variables such as gender, age, and cultural background" (Acosta-Enriquez et al., 2024), which is a valuable consideration while using the findings of the research and generalize it to include the Turkish context. Therefore, there is a need in research for AI-based research in Türkiye for English Language Teaching (ELT) students' attitudes towards ChatGPT use as a student, and as a teacher; in addition to research concerning what variables may significantly affect such attitude, and if ChatGPT literacy is one of them.

Lastly, the latest incarnation of GPT, which is ChatGPT-4 Omni, or GPT-4o is the focal point, and the current latest version of the chatbot of the time of this research; which,

in addition to having additional features like speech recognition and the capacity to speak (Shahriar et al., 2024) , also has seen improvements in its overall capacity to do its tasks. In the topic of hallucinations, it is found out that GPT-4 has shown improvements in reducing them, in addition to reducing the production of harmful content, disinformation, and bias; although, there is still need for further improvement (Koubaa, 2023). Due to the release date of GPT-4o being May 2024, the amount of research on it is overall, especially in ELT, little to zero.

For the aforementioned reasons, the relation between ChatGPT literacy and ChatGPT attitude should be researched in addition to other variables such as age, gender, academic year, and so on, to help students and prospective teachers understand the perceptions about ChatGPT, to find out about the positive and negative aspects of this new technology, in addition to closing a small gap in the research gap of ChatGPT and its use in language education.

1.1. Statement of the Problem

ChatGPT has been shown in various research and studies that it has the potential to improve various aspects of teaching and learning. Significant learning was achieved through the use of ChatGPT in the learning process. For example, significant improvement was seen in L2 writing through the use of ChatGPT (Yan, 2023); however, the same study that exhibited such findings also found out that students' beliefs about ChatGPT showed more concern about its threats than a positive outlook of its application. Similar beliefs, at least the mention of how ChatGPT possesses many risks and threats to the integrity of education and academia is displayed in almost any research in the topic (Ahmad et al., 2023; Gao et al., 2023; Lo et al., 2024; Lund & Wang, 2023; Shoufan, 2023), as it is one of the main discussion points and focus of research concerning ChatGPT. However, from how quickly the technology is improving on the topic, and the increasing widespread use of it for the educators, students, and the public. In addition to the possibilities in application, large language models like ChatGPT can also help with the development of other technological hardware and software and personalize those constructs to fit better with the needs of various contexts (Kasneci et al., 2023). In a prompt given to ChatGPT in a study addressing the concerns about the use of ChatGPT in a university setting, ChatGPT responds to the prompt *“What significant benefits or*

opportunities for higher education can ChatGPT offer?” by listing certain benefits, ending the output with:

“By leveraging the strengths of ChatGPT while addressing its limitations, higher education institutions can harness the benefits and opportunities it offers to enhance the learning experience, promote student success, and foster innovative educational practices.” (Michel-Villarreal et al., 2023)

It can be argued that the research concerning a new emerging technological device is more pressing when there is a possibility that said device may cause more harm than good if it is to be used without proper training, considerations, and awareness. Therefore, it is highly important to understand where different populations stand on the topic, for both their literacy on ChatGPT, and also their attitude towards different aspects of it. Due to the fact that findings are anecdotal on the subject, or are studied in different contexts, with different group of participants, or have not taken the interaction between literacy and attitude as a focus of study, there remains a gap in literature where there is not enough data on why the participants have the attitude that they have on this new emerging technology and its implications and what their attitude on topic is overall.

Through this study, a rough framework can be given to carry out the same study in different populations, whether to see if there is any difference, with little modification to the method and tools; in addition, if a significant relationship between attitude and literacy of the large language model is to be found in this study, there could be implications on further studies to find correlation, research on other variables, and practical consideration in ChatGPT education for students and practitioners.

1.2. Aims and Research Questions

The main aim of this research is to find out the literacy and attitude of university students studying in the field of English language teaching regarding ChatGPT, how often they use this resource on their studies, how they use it, their thoughts and opinions about it, and their thoughts about the positive and negative connotations of the existence and use of this technology, both as a current student and as a prospective teacher; and to find out if there is a correlation between ChatGPT literacy and attitude; ChatGPT literacy and academic year, age, or gender; and ChatGPT attitude and academic year, age, or gender.

- 1) What is the ChatGPT literacy of ELT students?

- 2) What is the ChatGPT attitude of ELT students?
- 3) Is there a significant correlation between ELT students' ChatGPT literacy and ChatGPT attitude?
- 4) Is there a significant correlation between ELT students' ChatGPT literacy and whether if the students have taken compulsory pedagogical lessons in the field?
- 5) Is there a significant correlation between ELT students' ChatGPT attitude and whether if the students have taken compulsory pedagogical lessons in the field?

1.3. Significance of the Study

The precedent that language learning and its different aspects could be reasonably explained in blanket ideas that would apply to all context, teacher, and learner profiles simply cannot be true, due to the overwhelming variety of results garnered from research throughout the world of English education, and how they differ depending on different variables. However, the benefits of appropriate use of technology in the learning process is a mostly agreed net positive (AlTwijri & Alghizzi, 2024; W.-L. Chang & Sun, 2024; D. Chen et al., 2023; Cruz et al., 2023; Jiang et al., 2022; Sylvia Liu et al., 2024; Mulcahy et al., 2024). As technology advances, in education, it presents “more opportunity and more choices than ever before” (Finn, 1960), but it also can provide challenges, and new considerations possibly harmful to the education and academic process; therefore, the phenomenon of ChatGPT in education should be researched from all angles and varieties to find out how to circumnavigate without falling to the pitfalls of the misuse of this new technology, and harness the potential positive for the education process. Because of the productive capabilities and the potential of ChatGPT, the consequences for inadequate overall literacy to the concept of large language models can lead to uncaught plagiarism that can be significantly avoided through the use of the proper resources, although concerns about the accuracy of such tools are still a highly discussed and researched topic as well (Baron, 2024).

Due to the fact that the current study touches open a number of different aspects of student education, student identity, different aspects of being a prospective teacher, teaching identity, and so on, this study can be significant towards both students and

educators garnering information and ideas about their professional development as teachers and/or students. Integration of technology into studying and teaching plays a significant role in these types of identities; and from the idea that ChatGPT and large language models as a whole may play a highly significant part in the topic of technology use in education, ChatGPT literacy and attitude can also play a highly important role in student/teacher identifying towards technology use in education. For this reason, this study can help the reader identify their own thoughts, opinions, and biases towards ChatGPT, understand how much they understand of this new technology, and may shape their literacy and beliefs towards the subject going forward.

Additionally, in terms of curriculum design and the implementation of large language models in language education, it is believed that this thesis may help the policymakers and curriculum designers in understanding the different perspectives and literacy levels around this new technology; which may affect bring attention to the perceived need of AI education in both students and teachers. These may be used as insights by the curriculum designers in employing AI education into the education process; and more focused on language education, may design parts of the curriculum, or the curriculum as a whole, with the new technology of large language models and AI in general in mind, which may have a possibility of changing various aspects in language education in the future.

This study and its structure have not been repeated in any context yet. This fact reveals that this study and its findings have a high chance of being unique in a local, and also global context. In addition to its uniqueness, which helps the study paint forward new ideas and discussion points in literature and also in practical application of the topic by teachers and students, and also may be used for the purposes of finding out the practical implementation and acquirement of new technology for ELT students who are also prospective teachers who will be using this new technology if they are to decide on acquiring it for their studying and teaching arsenal of technological tools and devices. This can be used to see how new technology is perceived in the space, and how long does it take for students to gain a certain level of literacy towards the new technology, and how they approach it regarding their attitude towards it.

Lastly, this study has been designed to be scalable in the future. The overall research design can be used, with academical permission from the original creators of the

ChatGPT literacy scale and TAME-ChatGPT, in different contexts, different types of students, can be adapted towards its use for practicing teachers in different context, can be used for different contexts in different parts of the world, and can be scaled up to include more participants, more institutions, and multiple contexts which can be compared and contrasted as well. For this reason, as a blueprint for the particular study aims that this thesis presents, this study may be used as a part of construction for a similar study in the future, which presents the possibility of this study being a positive addition to the literature.

Overall, the significance of this study stems from the objective of finding out where university-level students studying to be English teachers stand on the topic of ChatGPT from different angles concerning education and academia, and considering that as future teachers, they may possess the potential to influence their future students on the topic of education (Baron, 2024; Cohen-Azaria & Zamir, 2021), understanding what the students' attitudes towards ChatGPT would prove to be useful in various facets. In addition, the inclusion of the comparison between the attitude and literacy of the participants can initiate the interest over the reasons behind the students' attitude and literacy towards ChatGPT, and what could be practically considered to be implemented in student or teacher training that could help the overall more appropriate use of ChatGPT and awareness of the concerns and considerations about the technology, as well as its potentials to improve language education.

2. REVIEW OF LITERATURE

2.1. Introduction

The aim of this research is to find out if there is a correlation between the attitude towards ChatGPT, and literacy of ChatGPT, among other sub-aims that include other variables such as age, gender, academic year concerning university students studying in the Turkish context to be English teachers.

For this aim, a modified version of TAME-ChatGPT was used to find out the students' attitude (Abdaljaleel et al., 2024) and a modified version of "ChatGPT literacy scale" to find out the students' ChatGPT literacy (S. Lee & Park, 2024). To contextualize which area of literature these aims are designed to provide new findings to, this section will analyze the history of literature covering different, increasingly specific areas of research and academia, starting from the emergence of the concept of AI in literature, types of research and findings given in a variety of different specializations, in addition to a brief explanation in literature of how AI functions, and what are the mechanizations of how we understand AI in its current and past form.

Continuing, the history of chatbots and their improvement over the years will be reviewed until the modern era of large language models and ChatGPT, and in the current era, research on chatbot literacy and chatbot attitude will be examined in this chapter, especially regarding ChatGPT in detail. In the last section, literature around AI and chatbots will be reviewed in detail regarding education, ending with the review of AI and chatbots in language education specifically, and research around ChatGPT in language education.

2.2. The Emergence of AI

The first mention of what would be called a robot in current language comes from Greek mythology, from the idea of the bronze giant Talos invented by Hephaestus, the god of inventions (Mayor, 2018). The idea of the automata spanning back to the Iliad (Rossi & Russo, 2017) illustrates how far back the human ponderance about making a copy of itself, making life as it knows it that thinks and acts like it does goes back.

Going forward more than two dozen centuries, a proposal paper written by McCarthy, Minsky, Rochester, and Shannon (2006) expresses that, originally written in 1955, "The speeds and memory capacities of present computers may be insufficient to simulate many

of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.” and through this proposal, it is written “An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.” The hypothesized advancements of artificial intelligence would take longer than a summer, and it can be argued that AI in its desired form is still in its infancy, the developments done starting with this proposal, which is the source of the term ‘artificial intelligence’ being coined, the foundation of AI research to come in the future is built.

From the first incarnations of AI research, the desire to shorten the process in which AI tries to find a solution for a problem by testing every possible answer is apparent (Minsky, 1961). From the world of mythology to science fiction, to systematic research of a concept, while far away in its dreamt potential, AI becomes a possibility with the help of every research paper.

During the same era, a code-breaking machine constructed for the British government in the era of World War 2 is to be regarded as the first computer (Haenlein & Kaplan, 2019). Through the recency of both developments in computational technology and in the field of AI, one can surmise that AI is not a concept developed after the advancement of computers closing on what is known of them in recent times, but a concept growing with the technology it is based and is connected to. Alan Turing, working on this machine, named The Bombe, later went on to publish material regarding AI, and intelligent machines, and coined the term “Turing Test” (Turing, 1950). In simple terms, the Turing Test can be described as the idea that if a machine could communicate with a person, in what is described in the article as a teleprinter, but could apply for other technological devices today where face-to-face interaction is not necessary for communication, and throughout the conversation, if the person cannot distinguish that they are not communicating with a real human being, the machine is thought to ‘pass’ the Turing Test. While the non-simplified version of the Turing Test, Imitation Game, has different aspects like the existence of third person to the equation called the interrogator, and the idea of the two communicators trying to decipher the other side’s gender through general

conversation; however, throughout the years, the concept has been simplified to the idea that if the person does not know that they are communicating with a machine, and therefore there is no noticeable difference between the human and the AI, the AI succeeds in imitating real intelligence (Saygin et al., 2000).

There were theories against the Turing Test and its potential for pin-pointing the existence of intelligence in machines. The Chinese Room thought experiment by John R. Searle (1980) posits that the machine giving the output does not understand or internalize the input, it only follows basic instructions to create imitations of answers to the questions given, while not even understanding the language or perhaps even the concept of language. Keith Gunderson points out almost two decades earlier, the idea that the machine does not have to actively think like the human does:

“But one might well contend that machines can't think, for they do much better than that. We could forever deny that a machine could think through a mathematical problem, and still claim that in many respects the achievement of machines was on a higher level than that attained by thinking beings, since machines can almost instantaneously and infallibly produce accurate and sometimes original answers to many complex and difficult mathematical problems with which they are presented. They do not need to "think out" the answers.” (Gunderson, 1964)

The philosophy of what is artificial intelligence changes from source to source. The Oxford Dictionary defines artificial intelligence as “the study and development of computer systems that can copy intelligent human behaviour.” There are similar thought experiments (Block, 2022; Łupkowski, 2006); in addition to arguments that the Turing Test is difficult, perhaps even too difficult as a benchmark (Saygin, 1999). However, in short, it is pointed out that no chatbots have been able to pass the Turing Test so far (Saygin et al., 2000). This demonstrates that recent chatbot developments and the overall developments in AI have been highly recent and as of the year 2000, the Turing Test was still seen as a benchmark still to be beaten by AI. In 2024, AI does pass the Turing Test more than its previous incarnations and is getting better with every version (Biever, 2023). In a study done by generating medical advice using ChatGPT and constructing a Turing Test inspired methodology, it was found that, on average, chatbot responses were identified as AI by the participants 65.5% of the cases, and human responses were correctly guessed 65.1% of the time (Nov et al., 2023).

After the conception of AI field of research, there have been many research articles written about different aspects of the concept; however, after the first boom and the initial developments, AI research hit a roadblock despite the general stance towards AI development being optimistic at the time. While the exact date of the ‘AI winter’ changes from source to source, it is usually agreed to be around the 80s (Haenlein & Kaplan, 2019). After the 80s, AI saw new developments through the advancement of technology. Some examples include Deep Blue, a chess machine that defeated then-World Chess Champion Garry Kasparov in 1997 (Campbell et al., 2002), solving the game of Checkers with the use of AI to analyze every possible position the game can be played (Schaeffer et al., 2007), the development of Siri, a virtual assistant built into the iPhone 4S developed by Apple (Knight, 2012).

The current state of AI, starting from the 2010s, is concerned with Deep Learning (LeCun et al., 2015), Big Data (Boyd & Crawford, 2012), AI image-creation tools (Saharia et al., 2022), and other similar concepts. With the advancement in AI technology, questions such as the ethical properties of AI and the responsibilities in using it have arisen (Barredo Arrieta et al., 2020). One of the more popular AI tools discussed with these topics is ChatGPT; however, chatbots as a concept and tool have not been created with the development of ChatGPT; but more than half a century before.

2.3. The History of Chatbots

While the name “chatbot” or “chatterbot” did not see use in literature until the year 2000 (Zdravkova, 2000), the first iteration of what would be called a chatterbot, a term later shortened to chatbot, not to be confused with a chatbot named chatterbot (Nuruzzaman & Hussain, 2020), came to be in 1966. Joseph Weizenbaum developed a program for the IBM 7094, a computer developed in 1962 by IBM as the upgraded version of the IBM 7090 that was developed earlier in 1959, and this program was called the “ELIZA Program” (Weizenbaum, 1966). As it is explained, it was a “program which makes natural language conversation with a computer possible”. The program worked by analyzing what was written by the user to find a keyword, and if the keyword is found, the sentence is transformed with the rules given around that keyword, with varying other rules for the software to function working in tandem with the given input. Therefore, it can be said that the program manipulates the user’s input rather than creating a response from scratch. Here is a small excerpt from the conversation given by Weizenbaum:

“Men are all alike

IN WHAT WAY

They’re always bugging us about something or other.

CAN YOU THINK OF A SPECIFIC EXAMPLE

Well my boyfriend made me come here.

YOUR BOYFRIEND MADE YOU COME HERE

He says I’m depressed much of the time.

I AM SORRY TO HEAR YOU ARE DEPRESSED

It’s true. I am unhappy.

DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY?”

(Weizenbaum, 1966)

As it can be seen, the responses make sense and are relevant to the topic and context given by the user. Even though an immense leap in AI technology, it had its issues, pointed out by Weizenbaum. The program had issues figuring out the most relevant keyword, and therefore the main context of the input. It also had problems with the identification of the context of the keyword, transforming the input properly, problems responding when no keyword was found, and so on. It had smaller issues like not being able to use punctuation, and the user not being able to use the question mark. The technology was rudimentary, but the steppingstones that would lead to current technology was being laid out by it.

In 1972, a chatbot was created by Kenneth Mark Colby called PARRY. it was used to function as a paranoid schizophrenic patient and rather than ELIZA’s mirroring approach where the statement given by the user was reflected back at them by constructing the structure into a sentence, directly in some cases, PARRY tries to “provoke controversies” to lead the user into giving more elaborate answers. It was used to train young psychiatrists and garner research into psychiatry, especially the topic of schizophrenia (Zemčík, 2019).

With the construction of Jabberwacky in 1988, and its subsequent versions developed, it significantly focused on dialogical patterns, and constructing such patterns

through the users interfacing with it. It was developed with the main reason being entertainment (Černý, 2022), although considerations on its use in language education have been made for its later incarnations as well (Fryer, 2006). The way the chatbot was written and functioned paved the way for the road towards the advancement of chatbots (Adamopoulou & Moussiades, 2020). It was first made available to use online in 1997, and a later version of it called Cleverbot was released in 2008 (Al-Amin et al., 2024).

Starting from 1991, the Loebner Prize Competition promised a \$100,000 award that would be increasing every year until a computer was proven to pass the Turing Test (Epstein, 1992). This could have caused an effect on chatbot development, since the 90s saw an uptick in AI research and the construction of new chatbots. Around two years earlier, Jim Aspnes revealed TINYMUD, a chatbot that also had the ability to create “textual scenery simulating physical spaces”, and “user extensibility”, since the aim of this software was to help the users create a multi-user dungeon (MUD), or what will later evolve to Massively Multiplayer Online (MMO) games (Bartle, 2015). Aimed to create a world for people to interact and communicate in, computer-based characters alongside characters that people were controlling was also introduced into the system, one of these computer-based characters was connected to ELIZA which people could interact and communicate with (Mauldin, 1994). Ultimately, it was decided that TINYMUD could not pass the Turing Test, and therefore did not win the proclaimed prize; however, its development did not only push AI research further, but it also helped online gaming and digital communication develop as well.

With the advances in technology, the foundation of pattern-matching algorithm (Adamopoulou & Moussiades, 2020) laid out by ELIZA could be better realized in ALICE, Artificial Linguistic Internet Computer Entity (Wallace, 2009), developed by Richard Wallace and launched in 1995 (Černý, 2022). As a chatbot, “ALICE used sophisticated pattern matching rules to have natural-sounding dialogs” (Al-Amin et al., 2024). While an improvement on ELIZA, ALICE still had drawbacks, such as not having a reasoning functionality, grammatical and sentiment analysis, and a tendency to be repetitive in its responses with the same inputs (Pardeshi et al., 2020). Nevertheless, ALICE was able to win the Loebner Prize in 2000, 2001, and 2004 (Bradeško & Mladenović, 2012).

In 2001, a chatbot named Smarter Child was created by Active Body and saw mass appeal and popularity in its use (Al-Amin et al., 2024). With its ability to actively converse and provide information depending on what the user is typing, Smarter Child demonstrated the “potential for chatbots to serve as virtual assistants” (Pillare et al., 2023) which would later be realized in Siri and other similar virtual assistants, since the feasibility for a chatbot to be more than a conversation bot, but also be a source of information was seen with Smarter Child (Al-Amin et al., 2024).

A similar advancement, alongside an advancement in what was steadily taking shape to become the concept of machine learning took place with the development of CALO, Cognitive Assistant that Learns and Organizes, developed with the funding of the Defense Advanced Research Projects Agency (DARPA) through a 5 year period, the main focus of the chatbot was to be an assistant that was self-teaching and self-improving through use (X. Huang, 2021), paving the way even more towards the modern personal assistants like Siri, Cortana, and other similar products.

Based on Artificial Intelligence Markup Language (AIML) technology, which incorporates pattern recognition and pattern matching, used also for ALICE, Mitsuku was developed by Steve Worswick in 2005. Developed to be a chatbot socially interactable by the user (Croes & Antheunis, 2020), it is described as “quick-witted, occasionally confusing, and strangely engaging” (Mone, 2016). Because of its social and conversational nature, Mitsuku has also seen interest with its potential as a second language conversation partner, and how it offers L2 learners “substantial opportunities as a conversation partner” (Shin et al., 2021). It has also won the Loebner Prize in 2013, 2016, 2017, 2018, and 2019, being the last Loebner Prize winner before the contest was discontinued (Jacquet & Baratgin, 2021).

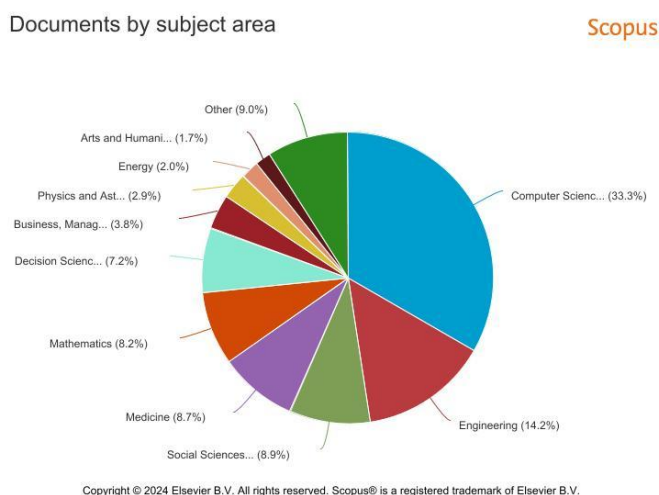
The first chatbot to incorporate auditory speech was Dr. Sbaitso, short for SoundBlaster Acting Intelligent Text-to-Speech Operator. It was created by Creative Labs in 1991 to behave similarly to ELIZA and also act like a psychotherapist (Deryugina, 2010). Its ability to auditorily output its answers was the first of its kind. In the following decades, the technology around the idea would grow to become the technology behind personal assistant, with the first widespread example being Siri. Developed by Apple, Siri was released as a standalone application in 2010 and was incorporated into iOS in 2011 (Hoy, 2018). The main aim of Siri was to help and guide the users like a human assistant

would, with natural, human-like language, and with apt responses to the input given to it. It is able to receive input auditory or written and can respond in both ways as well. The AI can help the user look up facts, use navigation, manage devices, and so on (Černý, 2022); however, because of its vast access to user data, there have been concerns about the ethical and privacy-related side of improving AI via user data through self-learning algorithms (Sharif & Tenbergen, 2020). With the success of Siri, other major corporations followed suit like Microsoft's Cortana, Amazon's Alexa, and Google's Assistant (Hoy, 2018).

Throughout the 2010s, AI research gained more traction every year. Interest from corporations and services towards the subject increased as well due to its potential for revenue (Dale, 2016). AI's ability to have natural conversation with the user became the norm even if the developed AI had specializations towards different subjects and fields. Throughout the same decade, using the vast amount of data available due to the internet and interactional user-based platforms used by a significant portion of the human population to train AI (Roh et al., 2021), and algorithms where AI would train itself and improve itself through the data it is fed (Helm et al., 2020) became concepts highly sought after in a variety of fields concerning chatbot research (Figure 2.1).

Figure 2.1

Documents with “chatbot” in the article name from 2010 to 2024 by subject area

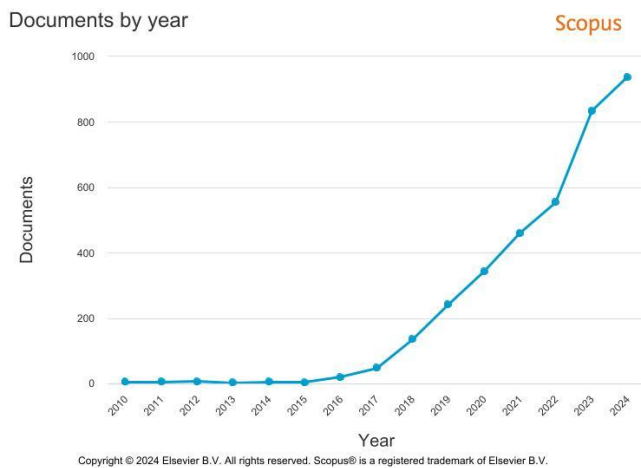


Source: Helm et al. (2020)

In addition to the interest by corporations and the general public, chatbot research and literature also has shown increasing interest every year in the 2010s, and even more so in the 2020s (Figure 2.2).

Figure 2.2

Documents with the word “chatbot” in the article name from 2010 to 2024 by year



Source: Helm et al. (2020)

From a total of 3593 documents published in Scopus from 2010 to 2024 that include the word “chatbot” in its title, 2325 of them were published in the years 2022-2024. The interest in AI, while tumultuous and wavering between its first conception in ELIZA in 1966 all the way to ChatGPT-4o in 2024, has shown an intensely increasing interest in research in the recent years, in addition to public interest to chatbots and AI in general (Qi et al., 2024). This interest, looking at ChatGPT specifically, can be seen to be even more significant. Since the first use of the word “ChatGPT” in the name of an article published in a journal on 7 January 2022 (Haque, 2023), to 18 November 2024, 6635 documents can be found including the word “ChatGPT” in their title.

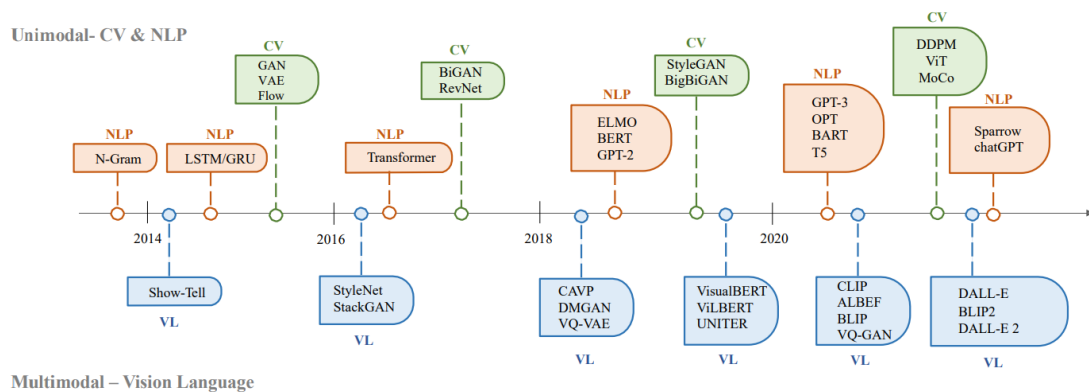
2.4. ChatGPT

OpenAI was established in 2015 from a group of AI researchers (Rajaraman, 2023), a year before the release of Google Assistant. While the company worked on various generative artificial intelligence, which can be summarized as a type of artificial intelligence capable of generating arguably original content (Banh & Strobel, 2023), the first version released to the public, ChatGPT-3.5, was released in November 2022, garnering significant interest (Samir Abdelhafiz et al., 2024). The roadmap and history of

generative AI related algorithms, concepts, and programs can be seen below in Figure 2.3.

Figure 2.3

The history of generative AI in CV (Computer Vision), NLP (Natural Language Processing and VL (Vision Language) (Y. Cao et al., 2023).



Source: Y. Cao et al. (2023)

As can be seen, ChatGPT as an application does not exist in a vacuum, nor is it separate from the current and past AI research and developments. Using known and tested machine learning concepts as a source, it is trained on large-scale data with the result being “an intelligent chatting robot which is able to provide a detailed response according to an instruction in a prompt.” (Wu et al., 2023). ChatGPT, arguably the natural development of personal assistants, separates itself from the chatbots in the past by being able to have natural conversations with the user with variety and relevancy, present available information, and, most notably, form content and material with the input, or prompt, in mind. It showed the use of the Pretrained Foundation Model, which is the model of AI being trained on large amount of data. In addition to its pretraining, it is also actively training itself on the new data being sent to it via users prompting the AI and using the application (Zhou et al., 2024). The way the output is generated by ChatGPT is by the large language model using its pre-training in terms of training through a massive number of texts, and through what is called the autoregressive model, it predicts the most likely word to be used in its output, and the next word after, and so on (Kocoń et al., 2023).

ChatGPT, since its public release, has been put to test in different fields for its effectiveness in answering questions, giving relevant answers, and so on. One study put ChatGPT through four final exams designed for law school. While ChatGPT passed all the exams, and if its success was consistent throughout the entire law school, it would be able to graduate and get a law degree; however, it scored near the bottom of each class and averaged a C+. Lastly, it seemed to perform better in essay type exams compared to multiple choice questions (Choi et al., 2021). It is also relevant to point out that this study was done using GPT-3.5, and a follow-up study using GPT-4o and later, yet to be released versions could prove to be illuminating in ChatGPT's development in whether if it would score significantly higher or not.

In another study, ChatGPT was given six Turkish neurosurgical society proficiency board exams. The results of the tests showcased that ChatGPT did not only pass the exams, but also ranked first in three of the six exams it was given. Overall, it answered "78.77 % of the questions correctly, while the candidates answered 62.04 % of the questions correctly" (Sahin et al., 2024). It is also pointed out in the study that ChatGPT is highly sensitive to the structure of the question it is given as an input, and if there is an error in the question pattern, while the student can "intuit what is being asked", the AI has difficulty understanding the question.

Concerning another study, in a computer science exam on a course titled "Algorithms and Data Structures", ChatGPT 3.5 scored 20.5 out of 40, while ChatGPT 4 scored 24; while the mean of 200 students taking the exam was 23.9 (Bordt & von Luxburg, 2023). Similarly, ChatGPT 3.5, on an English language comprehension test for high school level students, scored a mean grade of 7.3 on a scale of 1 to 3 compared to the 6.99 mean average of students (de Winter, 2024). In medical exams, ChatGPT 4's performance significantly increased compared to its previous version of ChatGPT 3.5 (Kipp, 2024).

Concerning the improvement between ChatGPT 3.5 and ChatGPT 4, it can be seen that over the course of all exams in different fields, ChatGPT 4 scores similarly, or in most cases, higher scores in the simulated examinations compared to ChatGPT 3.5 (Achiam et al., 2023). In a systematic review about studies testing ChatGPT concerning medical exams, from the 16 exams, it received at least a passing grade in ten of them, failed four of them, while in two of the studies, it was not mentioned whether if ChatGPT

failed or not. In the same studies reviewed, strengths of ChatGPT concerning exams were given as, from most mentioned to least, with at least two separate articles mentioning them, “explanation, accuracy, memorization, problem solving/reasoning, concordance, insight”. In terms of academic limitations, the order was “cannot understand images and words, unreliable/limited data, robustness, lacks insight, plausibility, and lacks critical thinking” (Sumbal et al., 2024). In summary, it can be said that, in its current state, ChatGPT’s ability to provide answers to all manners of questions is relatively high, with it at least providing a passing grade for a significant majority of exams it takes; and compared to its predecessors as chatbots, ChatGPT is a leap forward in AI technology in a lot of fields (Tyagi, 2024) (Gilson et al., 2023).

Because of the newly found tool that is ChatGPT capable of constructing new material depending on specific input in the hands of people from different specializations, industries, professions, and interests, ChatGPT in the short amount of time it is available found itself to be the topic of many research.

In terms of ChatGPT being tested for its effectiveness in varying fields, one study aimed to answer the question of “What can ChatGPT do” by analyzing 233,914 tweets from the social media website Twitter. The results showed that ChatGPT has five main functions: “creative writing, essay writing, prompt writing, code writing, answering questions” (Taecharungroj, 2023). For the purposes of giving a small number of examples to display the capabilities of ChatGPT, the aforementioned five main functions will be used to categorize the function of ChatGPT.

In terms of creative writing, one study prompted ChatGPT with creative writing questions relating to “ideas, characters, plots, and stylistic elements such as metaphors and without within various genres” (Fiialka et al., 2024). The output given by the large language model, ChatGPT, when assessed, was regarded to be at a level where even with limitations and shortcomings, signify that ChatGPT can be a useful tool to help human writers “not only within the narrative, but also in poetry and drama” (Fiialka et al., 2024). These limitations were explained to be mostly concerning repetitive and biased results. One article demonstrated a dialogue with ChatGPT where the user prompted it to write a story and prompting it to change various aspects of the story as it went on. The results showed that ChatGPT succeeded in creating a coherent story to the user’s specifications (Raiola, 2023). In another study done around creative writing course in Chinese

universities, using ChatGPT as a way to construct an AI-assisted process for the learners, 93% of the students expressed that they were satisfied with the process (Chen & Zhao, 2024). A fine-tuned version of GPT-2 was able to write 50 Turkish poems which were later revised to be a book titled “Gerçekliğin İçinde”, which was published with ChatGPT as the author (U. Kurt & Çayır, 2023). Lastly, one study showcased the success of ChatGPT in retelling a story; however, the results, when compared to people retelling stories, concluded that “spontaneous retelling by humans involves ongoing creativity, anchored by emotions, beyond the default probabilistic wording of large language models such as ChatGPT” (Breithaupt et al., 2024). To conclude, in terms of creative writing, ChatGPT has been found to be a highly valuable source whether used as the main tool for the production of creative writing products, or as a supplementary tool to help the user achieve their goals in creative writing; however, the output generated by AI and humans, when compared, points to the idea that ChatGPT is not yet at human-level when giving prompts involving creative writing.

ChatGPT has been researched to see its effectiveness as a tool for essay writing as well. In one study, 16 student essays written by or with the help of AI found by two AI detectors were compared to student-written essays. The results revealed that while how the language used noticeably different between essays written by ChatGPT and by humans, the AI-written essays were still analyzed to be successful in their goals as being essays about “personal experiences and opinions regarding real-life ethical dilemmas” (Kaliterna et al., 2024). As a study done with the Turing Test in mind, 24 student-written philosophical essays and AI-written essays were given to ten prospective teachers to assess (Bohlmann & Berger, 2024). Overall, 78.7 percent of the results correctly identified what essay was produced by the AI or the students. Because of the advancements in AI technology and the abilities of ChatGPT to construct essays with correct structure and flow, the article theorized that, relating to the prospective teachers who assessed the essays:

“Interestingly, however, they used criteria like argumentative and logical flawlessness and neutrality. We concluded from this that they are using an inverted Turing test and are no longer looking for rationality in machines but for irrationality in humans.”(Bohlmann & Berger, 2024)

A similar study had 69 high school teachers and 140 high school students take an AI Identification Test where they read pairs of essays and guess which essay was the AI generated one. Results were that teachers accurately guessed 70%, and students accurately guessed 62% of the pairs of essays they were given in the test (Waltzer et al., 2023). To summarize, in terms of essay writing, ChatGPT is highly capable of providing an output as an essay according to the specifications of the prompt it is given; however, similarly to creative writing examples, the difference between the human-written and AI-written essays can still mostly be recognized, with ChatGPT lacking the human element in writing.

In terms of code writing, ChatGPT has varying results and success. In a study done with the act of prompting ChatGPT to give answers to coding exercises in a Java programming course in an undergraduate level, the results showed that when the instruction given to the chatbot was clear, the answers given were effective; however, ChatGPT has also been observed to give incorrect or incomplete answers to the exercises if the prompt was more complex. In the end, it was discussed that ChatGPT can be a “effective tool for students to learn asynchronously” (Ouh et al., 2023). In another study comparing the code explanations of students with ChatGPT’s output, it was seen that ChatGPT produced explanations that were easier to understand and were more accurate overall to the given code (Leinonen et al., 2023).

In another research, ChatGPT was given coding challenges for the Python programming language, with problems around algorithms, data structures, and strings. The same problems were given to ChatGPT-3, ChatGPT-4, and humans. The results revealed that the same types of problems given to the large language models to train on the data, when asked again, ChatGPT-3 got the percentage of correct solutions more or the same with the human answers; and ChatGPT-4’s correct solution percentage is always higher than humans. When given a test set that does not include the same problems given in the training data, ChatGPT-3 gives out incorrect solutions more than humans regardless of the difficulty; and while ChatGPT-4’s correct solutions are higher on occasion than humans, as the problems have increased difficulty, it too falls under the percentage of correct solutions given by humans (Arefin et al., 2024).

Table 2.1

Performance of ChatGPT-3, ChatGPT-4, and humans across problems with varying difficulty levels (Arefin et al., 2024).

Difficulty Level	Topics	Correct Solutions %					
		Train Set			Test Set		
		GPT-3	GPT-4	Humans	GPT-3	GPT-4	Humans
Easy	Algorithms	96.42	96.42	58.88	33.34	58.34	55.95
	Data Structures	72.72	81.81	68	33.34	66.67	55.64
	Strings	100	83.33	52.4	62.5	75	69.81
Medium	Algorithms	58.15	75.17	51.17	14.6	30.34	44.93
	Data Structures	64.04	64.04	54.5	8.1	35.13	45.7
	Strings	54.54	72.72	46.89	28.57	28.57	38.47
Hard	Algorithms	29.03	45.16	43.83	0	5.08	40.22
	Data Structures	27.45	49.01	46.64	6.45	10.34	40.31
	Strings	29.41	52.94	44.31	0	0	33.62

Source: Arefin et al. (2024)

On the topic of fixing pre-existing codes, ChatGPT was given five different code snippets with vulnerabilities embedded in them in terms of security. When ChatGPT was asked to identify the problem with the codes it was given, it managed to find the major problems with the code three times out of five. In one code, while it found smaller problems within the code, it missed the existence of the major problem. In the last example, it missed the problem and was unable to fix the problem when prompted several times without changing the semantics of the function within the code, therefore changing a main property of it (Gasiba et al., 2023).

For the last functionality of ChatGPT given by Taecharungroj (2023), answering questions, various examples of ChatGPT prompted to answer exam questions, and its various degrees of success in doing so was discussed beforehand; however, aside from the assessment angle, there are also various other research showcasing ChatGPT's ability to answer questions. In a study where both ChatGPT and Google Bard, a similar software to ChatGPT developed by Google as a AI-powered chatbot, were compared in their ability to answer questions regarding ophthalmology, and their responses were found to be overall sufficient and correct. Google Bard was also given fundus images to try to diagnose patients; in which the AI correctly recognized the image's context but gave information regarding a different condition than the one present in the picture (Waisberg et al., 2024) In a similar study comparing the two AI models, regarding answering questions related to the history of America, it was seen that ChatGPT-4 outperformed Google Bard and "demonstrated a high degree of accuracy and ability to create solid arguments" (Rodríguez, 2024).

In the field of oncology, especially regarding head and neck cancer, ChatGPT gave comprehensive/correct answers 86.4% of the time, gave incomplete/partially correct answers 11% of the time, and gave completely inaccurate/irrelevant responses only 2.6% of the time (Kuşcu et al., 2023). Regarding the hospitality industry, one study demonstrated the effectiveness of ChatGPT as a support system for customers used by hoteliers (Singh & Singh, 2023).

Overall, ChatGPT has been found to be proficient and highly effective in various industries and fields; even though it still has its limitations. As a new era for AI research and use seems to have begun, it could prove to be useful to look at how far the technology has improved from its first iteration to current time. When ChatGPT is prompted by the same sentences given to ELIZA as an example of the chatbot's capabilities (Weizenbaum, 1966) GPT-4o, to be exact, the following is given:

“Men are all alike

It sounds like you've had some experiences that make you feel this way. Would you like to talk about it? Sometimes patterns can stand out on us, especially if we've been through similar situations more than once.

They're always bugging us about something or other.

It sounds frustrating! Sometimes it can feel like there's a constant demand or expectation coming from others, especially if they're not fully understanding your needs or boundaries. What's been on your mind lately? Talking it out might help relieve some of the tension.

Well my boyfriend made me come here.

Ah, I see! It sounds like maybe he suggested talking things out, hoping it would help. Relationships can sometimes bring us into situations we're not sure we need or want, especially if we're being nudged into them. What made him think coming here could be useful for you?" (OpenAI, 2025)

Compared to the answers given by ELIZA as showcased in the 1966 paper (Weizenbaum), ChatGPT's responses are much more detailed, comprehensive, and apt in finding nuance and comprehension in the input it is given. When ChatGPT is asked about how it differs from ELIZA, it listed and explained seven categories which were purpose and functionality, technology and model, context awareness, language understanding,

knowledge base, customization and memory, and complexity and creativity. At the end of the output, ChatGPT summarized the topic by saying “In summary, while ELIZA was a groundbreaking system for its time, I represent the culmination of decades of progress in AI, capable of understanding and generating far more sophisticated and meaningful interactions” (OpenAI, 2025).

2.5. AI and Chatbots in Education

A chatbot called Flexi-OLM was constructed to be a learner model that would ask questions to learners to find out their proficiency in the C programming language (Mabbott & Bull, 2006). A modified version of this model was used to see if it could give arguments about the learners’ beliefs about the programming language and the aspects around it to change or alter the belief of the learner. The results revealed that the approach was regarded as a positive by the learners, in addition to some arguments being able to change the belief of the learner (Kerly & Bull, 2006). As an experiment to see if negotiated learning could be done with the interaction being done by a chatbot, the study displayed positive results on the side of such a learning being feasible with the rudimentary chatbot technology in 2006 compared to current GPT models. As one of the first studies to incorporate chatbots directly into the learning process in an educational setting (Kerly et al., 2007), the number of research involving chatbots in education would increase as the technology improved over the years.

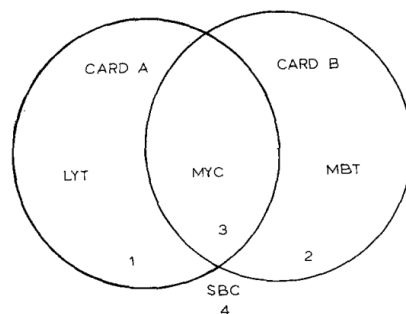
Even though the use of chatbots with their modern definitions and capabilities is a recent topic in educational research, the idea of using AI for the purposes of education is a few decades older concept than it. A programming language was created in the year 1967 by Bolt Beranek and New, Inc. with the aim of it being an educational programming language aimed to teach programming reasoning, logic, and artificial intelligence as a concept (Feurzeig & Lukas, 1972). The idea of teaching mathematics through this programming language, LOGO, was a tool researched through the experiments on prompting a digital turtle with different commands to guide it through different obstacles (Howe & O’Shea, 1978), looking to the idea of programming a rudimentary script that could be considered the most basic form of artificial intelligence for the sake of education where the learners learn through experimenting and analyzing their mistakes through the use of this programming language (Priest, 1981). After the first mention of AI in the field of education (Priest, 1981), the first study in education to incorporate AI as a problem-

solving tool incorporated a game called Blocks (Gallagher, 1981). The game is shortly described as having attribute blocks with three different shapes, colors, and sizes; and these blocks are placed around two circles forming a “Venn diagram” (Qualter, 1959). The aim of the player is to determine the attributes written on the blocks by asking which blocks can be placed on which loops, where different areas of the diagram have random, hidden rules on what attributes that could be placed in them (Figure 2.4). The objective is to place the fewest number of blocks on the correct places according to the rules (Brown & Burton, 1978).

AI Computer-Assisted Instruction was used to help the learner with the various stages of the game and answer questions given by the learner for the purposes of successfully completing the game. The only significant difference found was the difference between the control group and computer-assisted group, where the students that have received assistance from the AI performed more poorly on a transfer task compared to the students performing the task without the use of the program (Gallagher, 1981).

Figure 2.4

A game of Blocks in progress (Gallagher, 1981)



Key to stimuli: L = large, Y = yellow, T = triangle, M = medium, B = blue, C = circle,
S = small, Card A = Y and Card B = M.

Source: Gallagher (1981)

Using a computer-assisted program to teach how to solve mathematical problems to elementary school children, the “open sentence” mathematical questions were solved by different groups and while the computer-assisted groups showed significant increase in their ability to solve said problems, the control group who did not receive the help of the program also showed significant increase with more, for the lack of a better word, traditional approach to said mathematical problems (Barnard & Sandberg, 1988). The

difference of improvement between the groups were not found to be significant, signifying the result that while computer-assisted teaching in this experiment was successful, and did not hinder the children's ability to learn the subject, it was not significantly better than the ones learning without it.

Built on the improvements made by LOGO, the turtle programming program and language that aimed to teach students about programming through a rudimentary AI system, the research field of Artificial Intelligence in Education (AIED) was established. In Computer-Assisted Learning (CAL), some researchers believed the way forward to be using AI when the computer was involved in the learning process, and saw non-AIED CAL that was researched before the improvements in AI technology and the first vestiges of AI in education “educationally impoverished and rigid” (Cumming, 2003). Even in the first stages of AI research, the idea of using it in education was considered; in addition to the examples given by LOGO (Howe & O'Shea, 1978) and the game of Blocks (Gallagher, 1981), there are also other examples of software developed, some earlier than the two given examples. In 1970, SCHOLAR was first exhibited in a research article where its capabilities and working systems were explained. SCHOLAR was a software that could present information about certain geography, generate text, questions, and answer the given questions, with the possibility of “mixed-initiative dialogue between student and computer” (Carbonell, 1970). SCHOLAR could be considered the starting point of AI-based software constructed for the purposes of education and learning (Wenger, 1987).

From the history of AI development, it could be seen that even though some AI-based software were not designed with the main purpose of education, they would later be considered for the purposes of learning anyhow. MYCIN as a software was developed with the main objective was diagnosing bacterial infections as a rule-based expert system, even though the designer of the program also had the potential of the program that could educate physicians, practicing or studying, in mind (Shortliffe, 1976). In the same source where the capabilities of the artificial intelligence are discussed, six design considerations of what an artificial intelligence program needs to achieve to be considered if it is to be accepted by the user “the use of the computer and make use of its knowledge”. These principles were:

- “Program should be useful.

- Program should be educational when appropriate.
- Program should be able to explain its advice.
- Program should be able to understand questions.
- Program should be able to acquire new knowledge.
- Program knowledge-base should be easily modified.” (Shortliffe, 1976)

The given characteristics, especially the fifth property where it is discussed that the program should be able to acquire new knowledge points to the idea of machine learning decades before it would see any implementation in AI research. The second property of the educational value of the program was more realized with the design of GUIDON. This program was built on the properties and technical capabilities laid out by MYCIN, and was designed to be a software capable of dialogue to have “teaching expertise for helping a student solve a case” as a “case method tutor” with “problem-solving and tutorial dialogue capabilities” (Clancey, 1979). It is also discussed when describing what MYCIN is able to achieve that the AI program should not be the sole source and tool used in teaching, but a tool in addition to other tools, methods, and sources, which is an idea that will be repeated in the era of ChatGPT (C.-Y. Chang & Tang, 2024).

Self (1974) divides the knowledge of a teaching program to three categories, which are: “(a) knowledge of how to teach, (b) knowledge of what is being taught, (c) knowledge of who is being taught.” Using these demarcations, a student model to be used with the use of AI-based technologies in education is set where type (c) knowledge has to include the consideration of type (b) knowledge; so, the students’ current knowledge would be included and be the basis of the software that would later be useful considering the tutoring potential of the program. The student model presented would be used in the consideration of educational AI models in the future.

The potential for AI to improve in a way where it would be an immensely helpful to the learning process was speculated way before the capabilities of software such as ChatGPT were ever known. Birk (1992) writes “future computer software will involve expert systems and intelligent tutoring systems”, speculating that “the program will be a tool and advisor for the student- students will interact with the program, asking questions,

asking for help or advice, asking for examples. The program will respond intelligently- “ (Birk, 1992).

AI capabilities would improve as the technology improved. After the turn of the millennia, with the advancing technology in computing, considerations about AI use in social sciences would further increase as well (J. Lee & Friedman, 2009). In the 2000s, AI saw application in special education with attention deficit hyperactivity disorder and attention deficit disorder, learners with dyslexia (Drigas & Dourou, 2013), reading, writing, spelling difficulties, learners with autistic spectrum disorders, and sensory and/or physical impairments (Drigas & Ioannidou, 2012). As was mostly seen to be used around teaching computer programming and concept of AI and AI design itself, the advancements in using AI-based software to teaching programming continued as well, with the use of an AI-based software to teach a C programming course in a university level displayed the significant increase in student success in the course compared to the control group that did not use the developed AI tool (Tüfekçi & Köse, 2013).

With AI's improving capabilities, managing the AI to be shaped by learner characteristics and goals also became a possibility. One developed software was able to construct profiles for the students and personalize assistance based on said profiles (Samarakou et al., 2014). In addition to the personalization of output by the artificial intelligence, evaluation of student success could also be relayed through the inner mechanisms of the software, making evaluation via AI more of a possibility that needed to be considered as the technology moved forward. AI-based feedback would also see more of a research-based interest, especially for environments where autonomous learning via AI were slowly being designed only a few years before the public release of ChatGPT (Ge et al., 2018).

With the release of ChatGPT, research around AI use in education would increase substantially. Literature would be filled with research involving the advantages, disadvantages, and other considerations of the various uses of ChatGPT in chemistry education (Leon & Vidhani, 2023), health education (Hsu, 2023), programming (Bringula, 2024), travel and tourism (K. Zhang et al., 2024), geography (Redican et al., 2025), to only name a few; in higher education (C. Yu et al., 2024), in K-12 education (Hays et al., 2024); and mostly importantly for the purposes and objectives of this thesis, in language education, especially ELT (Mabuan, 2024).

2.5.1. AI and chatbots in language education

Wilson & Bates (1981) wrote about the potential of the use of Artificial Intelligence in language instruction concerning aspects like vocabulary building, improving listening, reading, teaching of higher-order skill, managing instruction, assessment, and so on. The article recognizes the potential of AI in aspects of education not considered or thought to be feasible at the time due to the state of the technology of AI at the time that the article was written and published.

It would take years before the theories presented would turn into practice; even then, the first incarnations of AI use in the classroom would be shortsighted in their ability to see the potential of computers and AI (Higgins, 1987a) as it supposes that AI will give the same exact output to the same input every single time, which is not true in current AI technology, and will be predictably even less so in the future. Even without the foresight of current technology, the article was criticized in its recognition of seeing the computer as a static drill machine (Coleman, 1987) and a “big dumb calculating machine” (Higgins, 1987b).

Intelligent Tutoring Systems (ITS) have been developed for learners to be presented with questions and problems that would impact the learning process, and for these systems to shape themselves according to the learners’ needs and learner profile have been ideas presented and experimented since the late 70s (Sleeman & Brown, 1979). The first documented ITS for language learning, XTRA-TE, was revealed in 1989 (L. Chen & Kurtz, 1989). The state of technology during the time showed potential, but it was still a long way from the day when language learners can interact communicatively with a computer in anything approaching the way they do with a teacher or other students (Obermeier, 1989). An ITS was developed, titled ET (English Tutor), that had three different capabilities. These capabilities were named the Tutor, the Student Modeler, and the Domain. The Tutor was able present exercises with varying topics to the student. The topics would also be given to the Domain section of the AI, which then would solve the exercise. These exercises would consist of a few sentences in English with the verbs in the sentences given in their infinitive form, which the AI would instruct student to show the correct form of the verb while also solving the exercise itself as well. The Domain would also display the reasoning behind which form was chosen regarding the exercise. The last section, Student Modeler, would compare the students’ answers to the exercises

with the answers given by the Domain for the purposes of identifying errors and trying to figure out the reason behind the errors. The next exercise given would be adapted by the machine in regard to the students' previous answers and errors (Fum et al., 1989).

As the possibility of the use of AI in the language classroom becomes a question asked by researchers, perspectives on the capabilities and place of AI in the setting is given. The majority belief at the time through the research published displays that AI was thought to have potential, but through perspectives such as the groundwork laid out by CALI-AI (computer-assisted language instruction incorporating artificial intelligence techniques) which looks at the possibility of AI in the language classroom, and how language teachers can construct frameworks and concepts to be incorporated into AI for it to be a better tool for the classroom (Bailin, 1988), also signifying that AI is not "intelligent" on its own, but purely a mechanical tool working off of the framework laid out by the professionals in the area it is being used in. It is theorized that through this framework, an AI replicating a language teacher successfully could be created. Although this is given as the ultimate goal of AI research specifically for language education among other professions by Bailin (1988), it is also pointed out that the current technology at the time the particular research was done was far from the ultimate goal; a sentiment that can be seen repeated at the current technology regarding ChatGPT.

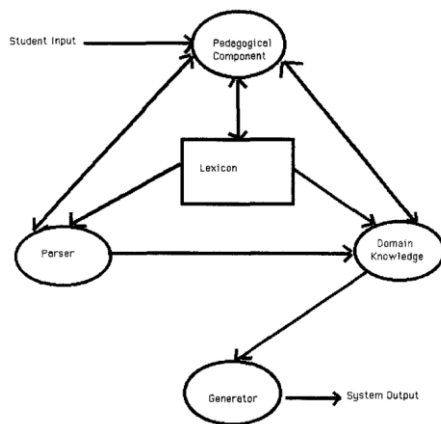
Even though the technology was not advanced enough in the area, through research and forming of patterns and framework for the AI in terms of natural language processing, semantics and pragmatics, morphology, language generation, language learning, and problem solving, AI was theorized to have the potential to reach its ultimate goal of replicating the teacher and be a successful tool in helping the learners reach their language goals (Bailin, 1988).

Another model for the aims of AI in language education was Model for Intelligent Computer-Assisted Language Instruction (MICALI). The model's proposed AI was "highly interactive and communicative", and it could also "initiate conversation with a student or respond to questions on a previously defined domain of knowledge" (Farghaly, 1989). The model focused particularly on three components. The first component MICALI focused on, also was seen as the most important, was the pedagogical component, with the second being the linguistic component, and the third being knowledge representation. The proposed model was said to be used as a way to create

English as a Second Language (ESL) software powerful enough to achieve the goals of MICALI (Farghaly, 1989).

Figure 2.5

The general concept of MICALI



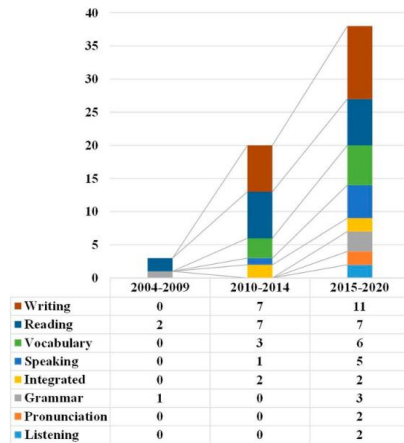
Source: Farghaly (1989)

Another perspective in the era shows that a section of researchers, and potentially practitioners, due to the state of AI and CALL as a whole, did not see the effectiveness of using intelligent computers in the classroom (Salaberry, 1996) due to their limitations and availability; and stated that the medium of teaching different aspects of language were more suited to already existent tools such as the “blockboard, the video, printed matter or the tape recorder” and rarely is the computer a better medium than the stated tools (Nyns, 1989). While AI-based CALL was not regarded as impossible, the models and systems present in the literature was not seen to be satisfactory, with the physical symbol system hypothesis shown in 1976 given as an example (Newel & Simon, 1976).

The technology in its first forms was researched with written input, and in the use of teaching writing for the reason that listening and especially speaking were much harder to be implemented in the software due to the speech recognition and synthesis technologies, and the integration of other skills in the AI language education literature would take decades to start (Figure 2.6).

Figure 2.6

Distribution of language skills in each stage of years on research concerning AI in language education (Liang et al., 2023)

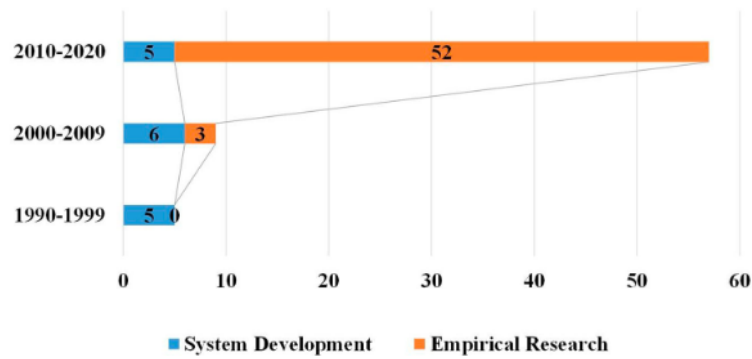


Source: Liang et al. (2023)

In 1985, French Robust Grammar (FROG) checker was developed (Imlah & du Boulay, 1985). Developed for French, it was able to correct the grammar of the input is given by the user, with its main difference being that it was able to do so in input given freely by the user, and not through the sentences it presents to be corrected to the user first. Built upon FROG, French Grammar Analyzer (FGA) aimed to correct the grammar given by the user much like FROG, but also with the addition of a better structure behind it, and with the ability to give error messages about what was corrected in the input, and why, for the purposes of teaching (Barchan et al., 1986). Its limitations in different languages, especially languages not in the same language family as French was discussed, in addition to its limitations in error messages, and the need for the development of the system. The system developed later was researched into its ability to transfer into another language and to be used in the education of that language. In the case of Spanish, its transfer was found to be sufficient (Yazdani & Uren, 1988). These studies into error correction in AI, while not the very first, are important stepstones into practical use of AI as an error corrective, and pedagogically adept manner of teaching through practice and error correction.

Figure 2.7

“Published literature on AI in language education in 1990-2020.” (Liang et al., 2023)

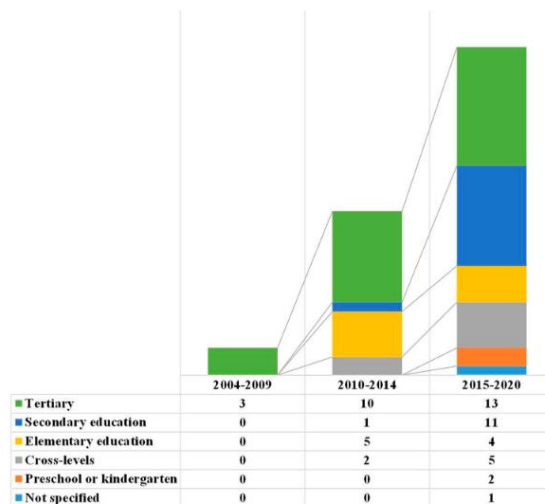


Source: Liang et al. (2023)

The state of AI would improve; and as a result, the pedagogical implications in language education would increase as well. No empirical research into the effectiveness of AI in language education could be found in the literature for a long time (Figure 2.7). The same could also be said for the compartmentalization of different stages of education and the use of AI in those education levels (Figure 2.8).

Figure 2.8

“Distribution of participant educational level in each stage” of years on research concerning AI in language education (Liang et al., 2023)



Source: Liang et al. (2023)

Other systematic reviews on the topic of AI in language education show similar results in research trajectory (X. Chen et al., 2020; Tang et al., 2023). As the technology

improved, other facets of language learning started to be researched in terms of AI potential.

While error correction was one of the first facets incorporated into AI use, processing of language input of the students was later adapted as well with the use of Japanese Computer Assisted Intelligent Computer-Assisted Language Instruction program, or Nihongo-CALI for short (Nagata, 1996). This study was also an example of a study researching the effectiveness of CALI compared to traditional means of education. The developed Nihongo-CALI was able to process “more sophisticated error analysis and feedback” compared to “conventional CALI programs” (Nagata, 1996). The results of the study displayed, through the self-reporting of students using the CALI and their traditional workbooks, that the software was much more effective and useful. The grammar notes given were found to be significantly better than the ones in the workbook, in addition to the exercises, the amount of interest raised by the students, self-study effectiveness, learning effectiveness, and so on.

After the effectiveness of using CALI in grammar and error correction was showcased, as the years pass, the focus on AI in language education would mainly fall to chatbots. Through the use of chatbots, other aspects of the language learning process were able to be studied through the lens of AI.

One of the properties that could not be incorporated into AI-assisted learning in the past but was later introduced as a material and tool to be used by the practitioner in English teaching as well as a source of research is the teaching of speaking. A study into the effects of artificial intelligence regarding speaking performance and speaking anxiety in students learning English (El Shazly, 2021) found out that the 48 EFL learners as participants using AI in the speaking environment improved their speaking ability significantly, although no control group to compare existed in the study. In terms of speaking anxiety, the results revealed that the students became more anxious with the intervention using AI, although not by a significant margin. A similar study was done on 90 EFL students from a state university in Turkey (Çakmak, 2022). Using the chatbot Replika, the students were instructed to use the chatbot outside of the classroom for targeted subjects around speaking for twelve weeks. The results displayed that the students improved their speaking performance significantly more compared to the

traditional face-to-face peer interaction in speaking practice. The same results were also seen with speaking anxiety.

Another study aimed to explore the effect of AI chatbots on speaking performance by dividing 110 university students taking English courses into three groups which were face-to-face chatting, AI text-chatting, and AI voice-chatting groups. From their performance, it was found that the face-to-face group improved significantly in terms of “responding to questions” and “expressing an opinion” parts of speaking performance, while no significant difference between the other two groups was found. At the end, through comparative analysis, it was found out that AI voice-chatting pairs significantly outperformed the other groups in terms of speaking performance tasks concerning the act of expressing an opinion. While there were significant and observable differences between the groups, no standouts between the groups on which was significantly better in improving speaking ability was found (Kim, 2021).

In a different study, EFL teachers were interviewed on the effects of AI chatbots in English speaking, including motivation, anxiety, student enjoyment, and critical thinking, four teachers interviewed believed that the AI chatbots were “both entertaining and effective teaching tools that are ideal for use in speaking classes” with positive outlook towards the “process of speaking learning, promoting students’ critical thinking, reducing speaking anxiety, and boosting students’ enjoyment” (Hapsari & Wu, 2022)

One chatbot that was designed by the researchers had the aim of having the capabilities of conversing with learners and to engage in meaningful conversation that would help the learner practice and improve their speaking abilities, as well as listening. Designed a few years before the public release of ChatGPT, it was titled BookBuddy (Ruan et al., 2019). BookBuddy could be considered as an example of the need for interactive chatbots with the potential to be used dynamically in language education. It, as a software, is able to be input with reading material, and in response, it can design and voice an “interactive conversation-based English”. The pilot regarding the impression of the chatbot showed positive results, at least in terms of the students’ enjoyment and engagement, with the participants being five 6-year-old native Chinese-speaking children learning English; however, the effectiveness of the lessons themselves given by the chatbot can be questioned to be addressed in a follow-up study regarding BookBuddy.

On listening skills, the amount of research done regarding AI chatbots is lower than all other skills of language (Liang et al., 2023). In studies done on the topic, it can be seen that a number of studies does not have sole focus on listening, instead aiming to research language ability concerning various skills. For example, a chatbot named Computer Simulation in Educational Communications (CSIEC) was developed in 2008 for the purposes of helping learning practice English (Jia & Chen, 2008). The chatbot was researched and put into practice and the results displayed that the use of chatbots helped the participants, high school level students, in increasing interest in language learning, confidence in English use, and in addition, also improved their listening skills (Jia & Ruan, 2008; H. Yang et al., 2022). To summarize, while there are studies that point towards listening as a facet in learning that potentially can benefit with the inclusion of chatbots, the gap in literature regarding listening and chatbot is clearly observable before the public release of ChatGPT, and after it as well.

In terms of reading skills, a chatbot, Storybot, was designed to focus on and improve learners' reading skills through digital storytelling (Bailey et al., 2021). The study surrounding the chatbot, a one-sample exploratory study where the participation rates and students' perceptions regarding the chatbot was aimed to be studied, showed that the chatbot required a high level of pre-requisite reading level to be engaged with, and revealed "explicit benefits to reading comprehension" that was shown through the answers given to the questions delivered by the chatbot.

The use of AI in the classroom and around it could be seen to lean towards more reading and writing based research and implications throughout the years (Liang et al., 2023). This could be due to the limitations of the technology, or the implications around the implementation of AI towards different skills, and the proposed difficulty around creating a chatbot around more active skills of the language. One example is iSTART, a web-based application developed in 2004, not described as a chatbot in name, that was designed to be a software aiming to train reading comprehension (McNamara et al., 2004). The software was first developed as an interactive model that had the ability to adapt output depending on the answers and input given by the learner and simulate a "classroom setting with interaction between three animated characters -an instructor character and two student characters – and the human trainee". It was first developed to train learners' reading comprehension and their reading strategy use regarding the

comprehension of science text; however, its interactive and malleable nature later spawned the use and research in language education (Dascalu et al., 2017; K. S. McCarthy et al., 2020).

On writing, similar studies can be found using chatbots. One study aimed to use a chatbot as an instructional supplement for the aims of teaching writing (Lin & Chang, 2020). Rather than a chatbot that was designed to be used by a student on their own, it was a design that was specifically constructed to help with instruction and feedback regarding writing. In the study, the students were showcased to show high interest to the chatbot; however, this interest was hypothesized to have the possibility to be related to the novel nature of the product, and the researchers discussed the possibility of a longer research duration for that effect to be considered as well, as this study was centered around a two-week period. While this study did not come to a conclusion for writing proficiency increase with the conjunction of whether it was related to the chatbot with other possible variables controlled; however, the testimonials given by the students were mostly positive, and an overall improvement in essay outline performance was displayed, which was the main focus of the study in regards to writing and the use of the chatbot.

In 2014, a study aimed to research the effectiveness of a writing-based chatbot named the Writing Pal Intelligent Tutoring System (Roscoe et al., 2014). W-Pal for short, the software had capabilities for “explicit strategy instruction, game-based practice, essay writing practice, and automated formative feedback”. A Writing Pal version 2 was also developed through the findings of this study, changing what students found too difficult, too easy, or overall disliked the nature, graphics, and tone of the material and chatbot. While students found the tool to be easy to use, they thought that reading the feedback and revising essays were difficult. The material through these findings were shaped with the consideration of student feedback by the developers of the chatbot. The adaptation of material and its difficulty and comprehensibility would later be a feat achievable by the chatbots themselves with the era of ChatGPT, and its ability to shape itself with input in real time.

Overall, most chatbots relating to language are developed with eclectic purposes in mind, with very few programs having a sole focus on a singular language skill to teach and/or help practice. One example is Storybot which held reading comprehension as the main focus, but researchers pointed out that the next step in evolution for this particular

chatbot was giving writing prompts to the user focusing the teaching of the skill of writing as well (Bailey et al., 2021).

One study aimed to research the effectiveness of using chatbots for “enhancing students’ performance and engagement” in language education. With 100 participants as the control group and 100 for the experimental group, the study saw positive benefits of using AI chatbots for the students’ learning process, with most students perceiving AI chatbots as a “essential part of their learning process” (Nghie et al., 2019). A research article written eleven years before concluded that “while substantial progress has been made in terms of chatbots’ language-handling, a robust ESL “conversation practice machine” (Atwell, 1999) is still some way off being a reality” (Coniam, 2008). As a more positive outlook towards the feasibility of chatbot use in language education can be hypothesized to be evident as the years get closer to current date and the technology improves, the use of chatbots in other sections of language education, like vocabulary teaching (Rodosthenous et al., 2020), immersion (Y. F. Wang et al., 2017), formative assessment (Benotti et al., 2017), and so on, was also researched.

With the initial date of release for ChatGPT, November 30, 2022, the landscape for AI-based chatbot research and development would change for language education, as well as for other fields.

2.5.2. ChatGPT in language education

ChatGPT, starting from its release, has become an area of academical and therefore practical interest to many fields, with language education being one of them. The interests raised, whether one believes it to be justified or not, has the potential to shape new ideas into the field. The study done by de Winter (2024) ten months after the release of ChatGPT-3.5, and only six months after the public release of ChatGPT-4 aimed to showcase the capabilities of the AI regarding its English language comprehension. The results of the study revealed that ChatGPT-3.5 was able to achieve 6.5 out of a 1 to 10 scale for the national high school exams in the Netherlands for English reading comprehension, and ChatGPT-4 was able to achieve 8.3. The mean of the students overall was 6.99. It can be argued that this demonstrated, at least in the form of reading, that ChatGPT possessed a high value of English language comprehension and could present that through given prompts, whether it can bestow that comprehension to the learner

effectively is another question entirely that has been studied, is being studied, and will most likely continue to be studied in the future.

From the research released since the release of ChatGPT, it can be seen that a high number of the studies focus on writing in particular. It can be theorized that this is because of ChatGPT's apparent ability in terms of writing, especially typing longer than a few sentences, is highly improved compared to the chatbots of the past. It could be from this that a high number of research has been focused around ChatGPT's ability to write, teach writing, and assessing writing.

A study aimed to see the effect of ChatGPT in an EFL writing classroom employed 19 participants who were Korean students majoring in English Language Education (Y.-J. Lee, 2024). The participants were separated into four groups and took part in think-aloud procedures and interviews for data collection. Through the collaborative writing process with ChatGPT with the aim of exploring the potential of ChatGPT in writing classes in addition to human scaffolding, the results showcased that the inclusion of ChatGPT was overall beneficial to the process, and it was noted that the students engaged with ChatGPT to plan, translate, and review during their writing process.

Another study separated 76 undergraduate students into an experimental and control group to see the effect of using ChatGPT-4 in students' writing (Boudouaia et al., 2024). It was found out that the experimental group, the group that used ChatGPT-4 in their writing classes, significantly outperformed the control group in their post-tests. Aside from writing ability, it was also seen that the experimental group improved their views of perceived usefulness, attitude, behavioral intention, and perceived ease of use regarding ChatGPT. The students also showed overall interest in using the chatbot. A similar experiment was aimed for the goal of finding out the effects of using ChatGPT-3.5 for university-level students' writing (Tseng & Lin, 2024). The results revealed significant improvement for the topic of accelerating students writing process through its feedback and generation of ideas, ensuring cohesive organization for the writing that the students are doing, and it was also found to be a "reliable substitution" for peer review.

In second language academic writing, especially in the university level, there is a growing concern towards using ChatGPT for help writing assignments, exams, and so on, where it is argued by some researchers that large language models like ChatGPT can be

a danger to academic writing, and even other software capable of fixing mistakes, grammar, produce ideas, help with the process overall with a technological edge could be a consideration in writing as an outside source helping the student, and whether their use should be indicated as a note to the reader and/or reviewer for the writing material, whatever material it is (Dillon, 2024). It is also important to point out while this is a subject that is widely discussed, it is not specific to EFL or ESL and are discussed in almost all other fields as well, similar to other issues and subjects raised by the existence of ChatGPT. Nonetheless, it is highly important in the subject of second language academic writing, since not only the writing facets of quality writing are reviewed in the courses where academic writing takes place, but also the learners' ability in their second language; therefore, the importance could be considered to be twofold, since ChatGPT can potentially help with both of these aspects in writing.

A study also aimed at academic writing sought out to assess the writing of university students compared to ChatGPT (Saarna, 2024). In this study, 47 instructors were told to rate 8 different essays, 4 written by ChatGPT, 4 written by students. Overall, the results exhibited a 54.25% accuracy in identifying whether the essay was written by a student or ChatGPT, only slightly better than 50% random selection. In this study, ChatGPT was also instructed to write essays mimicking the writing errors and mistakes the students have made to blend in with the other essays and it showcased the ability to successfully do so in the context of this study.

In terms of ChatGPT's usefulness in assisting with academic writing skills, a study done with ESL undergraduate students displayed a positive impact of ChatGPT on writing and "students' perceptions of the impact were also overwhelmingly positive" (Mahapatra, 2024). The data for this study was collected using three tests to assess students' writing ability and focus group discussions. The study also mentions that there are not many studies focusing on the smaller aspects and sub-skills in writing, and usually takes writing ability as a whole in recent ChatGPT-related studies.

The studies focusing on using ChatGPT as a tool for writing assessment is even greater than the number of studies focusing on using ChatGPT as a learning mechanism and/or tool for student growth in writing ability and other writing related skills. One study done in the Turkish context aimed to research the use of ChatGPT as a feedback tool for writing (G. Kurt & Kurt, 2024). The participants for this study were pre-service English

teachers, and data was collected from an academic writing course that they have attended. The results showcased that ChatGPT is a “valuable feedback tool for L2 writing”. The results of another study revealed that in terms of written corrective feedback, the perceived value of feedback from the students were similar in terms of teacher-written feedback and AI-generated feedback, leading to the implication that while AI has not proven itself to be better at written feedback for writing skills than the educator, but it could be a valuable tool or a sufficient replacement, at least in parts (Zeevy-Solovey, 2024).

A study done on 18 EFL students in Indonesia researched the use of ChatGPT alongside Grammarly and Quillbot, both applications aiming to help with the writing of the user, be it grammar, spelling, or other areas of writing like coherence, cohesion, and so on (Gozali et al., 2024).

A research article presented two studies done on “English as a new language” students in a “six-week long repeated measures quasi experimental design” where feedback by ChatGPT-4 was given to the experimental group, while the control group was given feedback by the teacher (Escalante et al., 2023). Study 1 had 48 university level students, while student 2 had 43. The second study was slightly different in the sense that the participants received feedback both from ChatGPT and also from their teacher. The results showed that there was no significant difference between the two groups in terms of learning outcomes in study 1. The preference for feedback was a “near even split” for study 2 in terms of preference for feedback from ChatGPT or human-generated feedback. The study concludes with the implication that using ChatGPT or AI in general as a feedback tool for writing has advantages; however, it should be used in conjunction with the human-generated feedback in a “blended approach”.

Similar to the study done by Escalante, Pack & Barrett (2023), another study published the same year looked at the effectiveness of AI rating for academic writing in the Turkish context (Geckin et al., 2023). The study, focusing more on rating the writing work submitted by 43 first-year university students, compared the ratings given by AI and human raters. A significant correlation between the scores given by human raters and ChatGPT-3.5 was observed for two of the raters and the AI. Similar to related articles, the study suggests that ChatGPT is a useful tool, but should be used to assist human raters, rather than be the sole rater of work.

Looking at the potential to support teacher feedback, a study was done on 50 Chinese undergraduate students where the written work of the students would be evaluated and given feedback by five Chinese EFL teachers and ChatGPT (K. Guo & Wang, 2024). The results of the study presented that while the human feedback was focused on content and language-related issues, the attention was divided equally three ways by AI to “content, organization, and language”; and overall, the tendency to use different feedback types were seen in teachers’ and ChatGPT’s feedback. The collaboration between teachers and AI for feedback on student writing was given as a viable implication from this study.

Bigger datasets could exhibit a more streamlined and “closer-to-the-truth” results that showcase the effectiveness of AI-generated feedback in different aspects. One study took 100 writing assignment by B1-level students in the Egyptian context (Shabara et al., 2024). The writing assignments were evaluated by 11 trained teachers and twice by ChatGPT-3.5. The findings revealed that in terms of scoring, the correlation between ChatGPT and the raters were weak-to-moderate. In terms of reliability, ChatGPT has been found to exhibit “moderate, yet questioned, level of inter-rater reliability”. The study concludes that “ChatGPT is not sufficiently consistent nor accurate in assessing writing quality and accordingly, it is not yet viable to substitute teachers” however, the study also adds that “ChatGPT can support teachers by providing feedback to students on their writing, and consequently, it can save time for them”. A study done on 82 compositions from the “Written English Corpus of Chinese Learners” highlighted that between ChatGPT and human raters, similarly, but somewhat higher, to the study done by Shabara, ElEbyary & Boraie (2024), the inter-rater reliability of ChatGPT was found to be moderately significant; however, in addition, the intra-rater reliability of ChatGPT was found to “not high enough to be acceptable” and the biggest rift in unreliability came from ChatGPT’s evaluation of the organization of the writings. The limitations of the AI are highlighted in the study, and a call for improvement of the large language model was discussed before its full integration into the evaluation process (Y. Yang, 2024).

Writing, and in conjunction and similarity, reading, are the most researched topics in the current era of ChatGPT research. In reading, there are different aspects being researched around reading ease, teaching and learning reading, reading skills, material creation related to reading, and so on.

One aspect of ChatGPT perceived by the general public and is being actively researched and discussed in academia is the ease of material creation for reading. A study on the subject has aimed to research the effectiveness of AI simplification of authentic materials (Çelik et al., 2024). Set in the Turkish EFL context, 105 undergraduate EFL students engaged with authentic texts and ChatGPT-simplified texts which were originally authentic blog texts. The reading comprehension scores and inferencing scores of the students were seen to be significantly higher for the ChatGPT-simplified texts. While reading comprehension and inferencing scores were seen to be significant, no significant impact on reading anxiety was seen.

Another study focusing on authentic materials aimed to investigate the “readability level of reading texts through the framework of Systematic Functional Linguistics (SFL) issued by ChatGPT” with a focus on whether if the complexity level of the material matched the standard of Common European Framework of Reference (CEFR) (Ramadhani et al., 2023). The results revealed that the authentic material generated by ChatGPT was not in keeping with the CEFR guidelines, and ChatGPT seemingly generated the complexity of the material with number of words in a text in mind rather than “lexical density, grammatical intricacy, and lexical variation indexes”. It is also another question entirely whether material generated by ChatGPT could be considered as authentic material.

A study (X. Wang et al., 2024), based on ChatGPT, designed a ChatGPT-based personalized reading comprehension support (ChatPRCS) system based on the proximal development theory (Vygotsky et al., 1978). ChatPRCS employs methods such as skill prediction, question generation, automatic evaluation, and so on. The results of the study done on the use of ChatPRCS highlights that “it provides learners with high-quality reading comprehension questions that are broadly aligned with expert-crafted questions at a statistical level” (X. Wang et al., 2024).

In terms of listening, the number of studies is even lower, and a gap in the literature relating to listening can be seen in ChatGPT EFL research. One article aiming to find out the advancements in AI and how it can relate to listening education looked at cases, listening education, and how ChatGPT functioned and is being used, and concluded that with the addition of using ChatGPT and Convolutional Neural Network model, which

is a “deep learning model commonly used in image and speech recognition tasks”, can help the learner heighten their English listening skills (Xing, 2023).

In another article, while it is discussed that learners can use ChatGPT to generate listening materials and tasks, the listening material created by ChatGPT can be one-dimensional, and learners would need additional exposure to “more voices and accents outside ChatGPT” (Mompean, 2024). It is also discussed in the same article that pronunciation-based activities could have limited results and is a weaker point of generations done by ChatGPT in terms of listening.

It can be seen that the amount of listening-based ChatGPT studies, especially empirical ones, are rare to none. In terms of speaking, while not as many as writing and reading, there are studies done on the various aspects of ChatGPT and speaking in the language education context. A case study employing observation, open-ended questions, and semi-structured interviews on 4 language learners studying in a School of Foreign Languages of a state university in Türkiye displayed the benefits, drawbacks, and the importance of prompts in terms of speaking (Üstünbaş, 2024) (Table 2.2).

Table 2.2

Content analysis of the interviews about the participants’ views (Üstünbaş, 2024)

Benefits	Drawbacks
Providing opportunity to practice speaking in English	Incorrect responses
Scaffolding with detailed explanations and constructive feedback	Long responses
The importance of prompts	The importance of prompts
Flow in speaking	Misunderstanding leading to disruption in the chat

Source: Üstünbaş (2024)

Overall, the study exhibited that regardless of the proficiency level of the participant, they were able to effectively use ChatGPT for speaking practice; however, a degree of digital literacy, and guidance from educators may be required in terms of effective use of ChatGPT for learning a language. It was also pointed out that the way

ChatGPT was able to shape its output and feedback depending on the context and the level of the student was observed to be a benefit for the learner as well.

A study done 65 participants separated to a control group and an experimental group aimed to see the effects of ChatGPT integration in terms of enhancing speaking self-efficacy (Yıldız, 2024). The study found out that the group using ChatGPT in their classes, the experimental group, showcased “elevated levels of self-assurance in verbal communication, enhanced ability to manage stress, improved proficiency in speaking, and increased pleasure in engaging in conversations” when the results were compared with the control group. The study concludes that ChatGPT was a positive impact for the students’ speaking self-efficacy and suggests that teacher should incorporate ChatGPT to help students gain opportunities “to practice speaking in a low-stress setting” which ChatGPT has found to provide in the confines of this study alongside its helpfulness in helping students overcome their speaking anxiety and increase confidence in their speaking abilities.

Lastly, a study done with the aim of investigating the effects of ChatGPT on learners’ speaking skills, well-being, autonomy, and academic buoyancy included 28 EFL students from an Ethiopian University (Sayed et al., 2024). The students were given the Michigan Language Proficiency Test to evaluate their English ability before the study and was also introduced to TOEFL iBT speaking section. Statistically significant results were found in students’ improvement in speaking ability, psychological well-being, learning autonomy, and academic buoyancy. Similar to the study done by Yıldız (2024), this research also concluded that “ChatGPT may create a supportive learning environment that lowers anxiety and increases confidence, thus improving speaking skills and overall psychological well-being” (Sayed et al., 2024) among other positive benefits that researchers and practitioners should give thought and attention to.

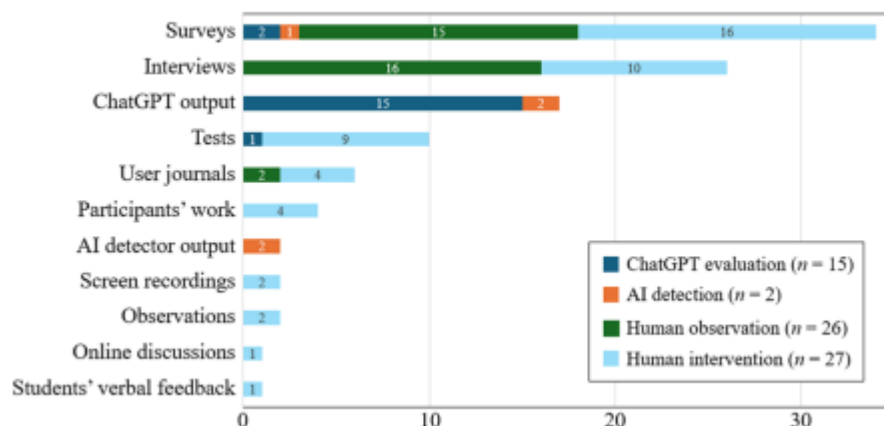
Aside from language skills and other language learning concepts and areas attached to such skills, there are studies that focus on ChatGPT and its various potentials and impacts on language learning as a whole. A systematic review incorporating empirical studies done on the application of ChatGPT in both ESL and EFL educational contexts showcases an overview of literature available on the subject (Lo et al., 2024). Analyzing 70 empirical studies, the systematic review found that the research aim of 30 studies was around writing, 14 on teacher perspectives, 11 student perspectives, 5 on

speaking, and 10 on other research issues. Overall, the studies were separated into 4 categories which were ChatGPT evaluation, 15 in total, AI detection, 2 in total, human observation, 26 in total, and human intervention, 27 in total. It can be seen in Figure 2.9 that a wide array of data collection methods was employed, but the overall data was collected heavily from survey, interviews, and the ChatGPT output given.

The overall conclusions of the systematic review that was collected from 70 empirical studies published in the 1.5-year period from the release of ChatGPT while that a major number of studies are centered around writing, “the effect of ChatGPT use in writing courses remains under-evaluated” and there is a low number of comparison studies. From the numbers, it can be seen that aside from writing, the other language skills are under-researched. It can also be seen that a good number of studies have focused on teachers’ general perspectives, 14 in total, and students’ general perspectives, 11 in total. From the list of the articles in the systematic review, there are several contexts that can be observed to be the focus of these studies; however, no studies done in the Turkish context have been signified in this systematic review; and aside from the systematic review, the Turkish context, especially in regards to student perspective remains, is an under-researched area of ChatGPT research, and an observable gap in literature.

Figure 2.9

Data sources used in the studies (Lo et al., 2024)



Source: Lo et al. (2024)

2.5.2.1. ChatGPT literacy in language education

The word literacy pertaining to AI and digital literacy in conjunction with the word ChatGPT on the ERIC database for second language teaching occurs whether in title or abstract 5 times (Alexander et al., 2023; Alm & Watanabe, 2023; Gozali et al., 2024; McCallum, 2024; Muñoz-Basols et al., 2023). The cultivation of the articles collected via the same filters from other databases only come up with a few more studies excluding the studies listed above. From this fact alone, it can be seen that ChatGPT literacy is an area of literature under the wider umbrella of ChatGPT research in language education is a lacking area in research and can be characterized to have gaps in literature.

Gozali (2024) expresses that more note and support should be given to students' digital literacy, equal access to technology, and ethical use of AI in the classroom. In another paper, for the purposes of looking at ChatGPT in the language learning process through a Freirean perspective, looking at Paula Freire, a humanist educator and philosopher whose work on critical pedagogy is studied often, Alm & Watanabe (2023) expresses that a number of strategies should be used “to ensure that its integration serves goals of equity, social justice and the development of critical consciousness, not just efficiency” of the potential of ChatGPT, one of these strategies being the strategy of developing students' AI literacy.

McCallum (2024) similarly points out the importance AI literacy for the success of telecollaboration classroom, pointing to AI literacy as a “key competence” not only for the student, but also for the teacher as well. Another advocate of teacher training for digital literacy is Alexander (2023), as it is considered a need for ESL lecturers alongside “targeted professional development, the use of advanced detection tools and a review of assessment policies and practices in relation to AI”. In the area of applied translation, Muñoz-Basols (2023) discusses the importance of applied translation and how its integration can help improve digital literacy and critical thinking in the language learning classrooms in the era where “AI-based chatbots, such as ChatGPT, has led to calls for a revision of traditional teaching methods to prioritize reflective reasoning and critical thinking.”

One discussion article aimed to see the potential between digital literacy and ChatGPT from the lens of English education (Bender, 2024). It is argued in the article that components which are laid to be useful when engaging with ChatGPT and similar

large language models like writing proper prompts for AI to give proper and relevant answers, articulating such ideas and creativity into prompts are argued to join and “become part of what constitutes digital literacy” (Bender, 2023). Furthermore, the author theorizes that AI will be a permanent, or at least long term component in education and therefore, digital literacy education and the importance of digital literacy to co-exist and co-work with AI (Baidoo-anu & Owusu Ansah, 2023) is also highly important. Because of the nature of a discussion article in this context, there is no experiment done on students in English classes or other people, be them students or educators, that are in or around the English learning process concerning their ChatGPT literacy.

Majority of researchers agree on the importance of digital literacy and AI literacy, and the importance of upskilling our digital literacy competencies to effectively benefit from the new chatbot technology is highlighted (Bozkurt et al., 2023). AI literacy can be considered to be based on Bloom’s Taxonomy (Bloom et al., 1956) due to the competencies it carries within itself which are “to know and understand, use and apply, as well as evaluate and create AI” (Ng et al., 2021). Meniado (2023) puts forward the idea that “Generative AI literacy programs should be conducted at the beginning of the school year or semester to orient the students on the proper use of ChatGPT and the risks of misusing/abusing it” (Meniado, 2023). The ideas put forward for the purposes of pointing out the importance of AI literacy and digital literacy is not only related to students, but also teachers as well. Research has shown that teachers tend to not be aware of the characteristics and other properties surrounding ChatGPT and were also lacking in terms of fact-checking the content generated by or with the help of software and programs like ChatGPT (Alexander et al., 2023; Lo et al., 2024).

Fitting to the keywords and context, only a few studies exist that are experimental with the objective of findings the ChatGPT literacy of participants who are involved in the study. One of these studies had a total of 492 questionnaires submitted for a questionnaire constructed specifically for the purposes of this study that is separated into six distinct categories of benefits, limitations, prompts, evaluation, assessment, and ethics (Q. Ma et al., 2024). The participants were separated into three groups: university teachers, 281 of them in total, student teachers and others, 141 in total, and school teachers, 71 in total.

Table 2.3*Mean and standard deviation for the sub-groups of the study*

	Benefits	Limitations	Prompts	Evaluation	Assessment	Ethics
	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD
University teachers (N = 281)	4.70/0.95	4.42/0.88	4.14/1.01	4.48/0.87	3.95/1.14	4.59/1.02
Student teachers & others (N = 141)	4.77/0.81	4.43/0.78	4.12/1.06	4.41/0.88	4.02/1.03	4.58/0.95
School teachers (N = 71)	4.64/0.91	4.30/0.97	4.10/1.10	4.38/0.83	3.82/1.18	4.55/0.96

Source: (Q. Ma et al., 2024).

From Table 2.3, it can be seen that the perceived literacy of teachers is high overall, with university teachers and student teachers perceived literacy higher than school teachers, but not enough to be statistically significant. Overall, the student teachers in this study only make up 22% of the participant group, their number being 108 in total. Furthermore, the study does not delve deeper into the sub-group of participants which is the student teacher group, and their involvement in data analysis is usually done with them being grouped together with university teachers. This can point out another gap in literature for student teachers, even more so for students studying to become English teachers altogether.

Going away from the subject of EFL and ESL, the number of ChatGPT literacy studies in education as a whole is still relatively low. The findings that are available point out to the importance of incorporating ChatGPT literacy to the curriculum (Casey, 2024). Two more studies outside the confines of English education research were done to develop a ChatGPT literacy scale the adapted version of which will be used for the purposes of this thesis (S. Lee & Park, 2023, 2024). The experiment explained in both of these studies is the experiment that was done using the developed scale by the authors.

The scale developed was used to collect data from 822 university students actively studying in a South Korean university. The students, with a mean age of 22.7 and an almost even split of male and female participants were sent the online questionnaire consisting of 25 items. The results of which showcased that there was a significant positive association between ChatGPT literacy and usage satisfaction, ChatGPT literacy and information and knowledge acquisition motive, and between information and knowledge acquisition motive and ChatGPT usage satisfaction. Motive was also a partially mediator for the relationship between ChatGPT literacy and ChatGPT usage satisfaction. Lastly, there was also a significant positive correlation between entertainment and leisure motive and ChatGPT literacy. Overall, the study found ChatGPT to be a positive associating with user satisfaction, “which is not only a key

predictor of user intention but also of learning effectiveness (S. S. Kim & Son, 2009; S. Lee & Park, 2023). While a general statement in education and software use, the same statement can be found within English education and online education as well (Ozfidan et al., 2021; Paul & Blau, 2009).

In summary, the number of published experiments and studies done on the ChatGPT literacy of English students can be considered to be nonexistent. While this can be attributed to the recency of ChatGPT's release, a high number of studies have been done regarding other aspects of ChatGPT and language education, and as many studies have pointed out, digital literacy and AI literacy are immensely important aspects to using large language models as the concept gets integrated into the learning and teaching process in education as a whole, and as a competency and literacy, should be given more attention both academically and also in the learning process itself (Alexander et al., 2023; Alm & Watanabe, 2023; Gozali et al., 2024; McCallum, 2024; Muñoz-Basols et al., 2023).

2.5.2.2. ChatGPT attitude in language education

The attitude around ChatGPT regarding both students and teachers is a more researched area in the literature compared to ChatGPT literacy. While the term 'attitude' is used indistinguishably in different research with terms such as 'perspectives' 'opinions' 'stance' and so on, the main focus of research interest can be described as the aim of seeing how students and/or teachers look at ChatGPT and its use in different domains like in the classroom, outside of the classroom, or in summary, around the progression of the teaching and learning process of English language education.

In terms of teacher attitude towards ChatGPT, one study done on the subject had 543 participants who were all teachers from Saudi Arabia (Allehyani & Alghamdi, 2023). The teachers were described as early childhood teachers teaching English as a second language. They were administered an online questionnaire to see their digital competence, beliefs, and perceptions. The questionnaire used was developed by the researchers. The results of the study revealed that the teachers who participated in the study put importance on the idea that ChatGPT was becoming a part of language education, and expressed a need to "improve their skills to encounter the massive educational pedagogical changes related to AI" although there were also concerns raised about the responsible use of ChatGPT and the ethical concerns around the topic of using ChatGPT in language

education as well. The statistical analysis for the questionnaire was also done with the aim of seeing whether if digital competency was correlated in a significant manner with the teaching experience of the teachers. As shown in Table 2.4, there was no statistically significant correlation between digital competence and teaching experience in all four categories the questionnaire was separated into: social awareness, technological proficiency, pedagogical compatibility, and the questionnaire as a whole overall.

A similar study focused on 30 EFL teachers active in L2 research and practice teaching in state universities in Iran (Derakhshan & Ghiasvand, 2024). Using semi-structured interviews, the study displayed a list of positive and negative outcomes from the perspective of the teachers of which the positives were listed as “augmenting learner, autonomy, providing personalized learning, reducing teachers’ teaching workload, designing assessment rubrics, and summarizing lengthy papers” and negatives as “that ChatGPT might kill creativity and academic integrity, encourage cheating in online exams, spread fake and misinformation into the world of research, and cherish high-tech plagiarism” (Derakhshan & Ghiasvand, 2024).

Table 2.4

One-way analysis of variance (ANOVA) results showing the values of correlation between teachers’ digital competency and their teaching experience

		Sum of squares	df	Mean square	F	Sig.
Social awareness	Between groups	3.949	2	1.975	1.062	.346
	Within groups	1003.841	540	1.859		
	Total	1007.790	542			
Technological proficiency	Between groups	1.024	2	0.512	0.283	.754
	Within groups	977.977	540	1.811		
	Total	979.001	542			
Pedagogical compatibility	Between groups	4.224	2	2.112	1.166	.312
	Within groups	978.210	540	1.812		
	Total	982.434	542			
Overall	Between groups	2.853	2	1.427	0.817	.442
	Within groups	943.151	540	1.747		
	Total	946.004	542			

Source: Allehyani & Alghamdi (2023)

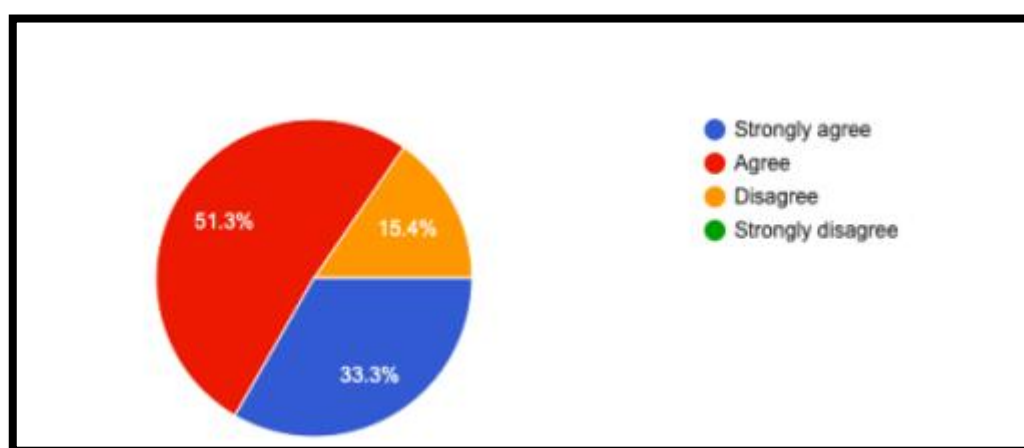
Another study aiming to explore the perspective of teachers regarding ChatGPT and ELT (Mabuan, 2024). The benefits and pitfalls of ChatGPT, when asked to 115

English language teachers in Philippines was highly similar to the ones listed by Derakhshan & Ghiasvand (2024). Furthermore, the study also points out the importance of ChatGPT education for both teachers and learners on how to properly and effectively use ChatGPT, therefore gaining the benefits of ChatGPT without having negative outcomes from the potential pitfalls of it. One section of the questionnaire designed for this study asked the teachers about the importance of technological proficiency and knowledge and if these attributes were required to use ChatGPT and “effectively integrate it into their teaching practices” (Mabuan, 2024). The results displayed that the majority of the teachers agree that technological proficiency and knowledge is required with 84.6% of the participants agreeing or strongly agreeing to the notion (Figure 2.10).

With 10 EFL faculty members from a university in Saudi Arabia, a study was conducted on the perceptions of the participants in terms of ChatGPT and EFL teaching (Mohamed, 2024). The results show that the teachers believe that ChatGPT is a valuable tool for EFL learning but because of the potential risks associated with its use, teachers should receive education on how to incorporate ChatGPT into the learning process. The efficiency to EFL teachers and how ChatGPT can provide support for it has been pointed out by the participants and the research as well.

Figure 2.10

“Using ChatGPT for ELT Required Technological Proficiency and Knowledge”



Source: Mabuan (2024)

Similar results from studies with teacher participant groups can be seen in experiments where the participants are students. To compare the two, 70 pre-service EFL

teachers and 30 in-service EFL instructors all studying or teaching in Türkiye were studied to find out about their perceptions regarding ChatGPT and its use in the EFL environment (Ulum, 2024). The results showcase that the pre-service teachers have a more positive attitude towards ChatGPT while in-service teachers seem to focus on the potential downsides to the use of AI in the EFL context, such as plagiarism, academic integrity, dangers to critical thinking, and so on. The study concludes by underscoring “the importance of responsible AI usage in education” and additionally, “the need for ongoing research and development to address the identified challenges” (Ulum, 2024).

Another study with 189 students enrolled in academic writing courses on their first year of university was conducted to find out about students’ thoughts about using ChatGPT in the context of academic writing (Ozfidan et al., 2024). Overall, the answers to the questions given to students were positive, with a push towards the idea that the advantages of AI tools were high in regard to improving “idea development, organization, syntax, - clarity”, and so on. The students gave positives responses for both grammar and vocabulary in writing for aspects such as saving time and helping them write better altogether.

Studies show that the perception from students and teachers is that both of these groups need to increase their AI literacy and learn how to effectively use tools such as ChatGPT, and a study done on 91 English learners from a University in the Czech Republic comes to the same conclusion (Klimova et al., 2024).

A quantitative study done on 405 EFL learners with Chinese EFL background who have used ChatGPT for their informal language learning were studied for the purposes of finding out about their perceptions and leverages regarding ChatGPT and EFL (G. Liu & and Ma, 2024). Using five constructs of “perceived ease of use, perceived usefulness, attitude, behavioral intention, and actual use”, a questionnaire was constructed. Using this questionnaire, the collected data exhibited that ChatGPT “has the potential to be the guardian angel” to assist learners’ engagement with informal digital learning of English activities creatively and effectively” (G. Liu & and Ma, 2024). After statistical analysis was done, it was found through comparing the constructs that perceived ease of use has a positive and significant effect on perceived usefulness, perceived usefulness can significantly affect attitude, attitude “plays an extremely significant role in predicting” behavioral intention; and lastly, behavioral intention has a

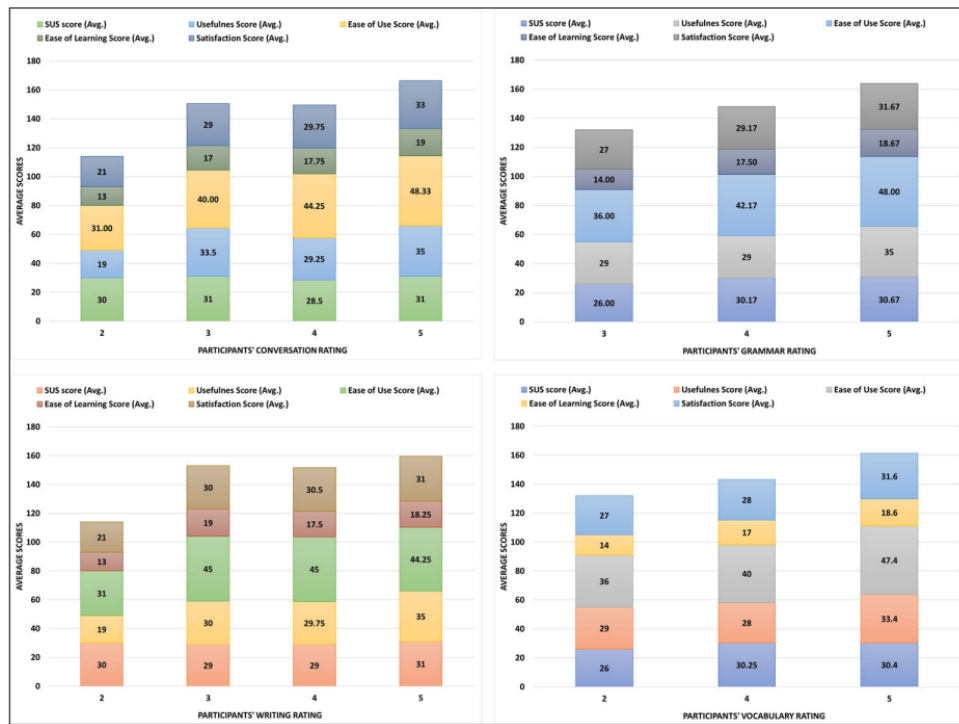
strongly significant and positive impact on actual use. Overall, the study displays the tight relationship between different constructs in ChatGPT attitude for learning English.

A similar study done on 867 EFL learners studying in the Chinese context also aimed to see the students' perceptions towards software such as ChatGPT in regard to informal digital learning of English (G. L. Liu et al., 2024). Similar to other studies (Boudouaia et al., 2024; G. Liu & Ma, 2024), this experiment is also done with a questionnaire design with the principles around technology acceptance model and an adapted technology acceptance model questionnaire which can be seen with the five constructs centered around technology use, ChatGPT in this case, which are "perceived ease of use, perceived usefulness, attitude, intention to use, and actual use" (Davis, 1989). In this study, only four constructs are explicitly used, leaving attitude as a construct out. The results of the study reveal that a significant number of the participants believe that informal digital learning of English occurs with hands-on experimentation with tools like ChatGPT; "intention to use increases as learners negotiate chatbot affordances and constraints"; and the importance of using ChatGPT as "tutors or conversation partners" for the context of informal digital learning of English.

One study that has done the process of the study differently added a step of the participants using ChatGPT within the confines of the study before giving their opinions and thoughts (Shaikh et al., 2023). The study's main aim is to explore the usability of ChatGPT. The results of the study demonstrate that ChatGPT is an effective tool in language education, and relevant to the aims and objectives of this thesis, the relationship between the participants' ratings in regard to their preliminary ChatGPT tasks, which were designed to be around conversation, grammar, writing, and vocabulary, were also discussed. The results of which have not been shown to be statistically significant or not; however, the numbers indicate that the overall positive opinion increases as the score increase; although the information is missing on whether the difference is significant, and two of the four tasks does not show direct correlation with rating and ChatGPT usability as shown in Figure 2.11.

Figure 2.11

Figures showing the rates of ChatGPT usability score and ChatGPT performance rating



Source: Shaikh et al. (2023)

Another study that presents data relevant to the purposes of this thesis has set out to investigate student opinions regarding ChatGPT for 369 university-level EFL students (Vo & Nguyễn, 2024). The study points out that while there are some gender-related differences between the five categories of ChatGPT student perceptions, which are given as “perceived usefulness, ease of use, positive attitudes, negative attitudes, and behavior intention”, the difference between the genders were not statistically significant, and the researcher mentions that form the participant group, 7.8% of the participants were male while 92.2% of the participants were female, making the statistical gender comparison invalid in a statistical sense.

The study also set out to see the difference between school year of the participants and their ChatGPT perception. A multivariate analysis of variance (MANOVA) test was done on the participants who were second, third, and fourth-year students, and the results highlight that there is a .002 significance between class levels and ChatGPT perception. This study is one of the rare examples of looking at the statistical correlation between gender and ChatGPT perception, and between ChatGPT perception and class level of the university students. Currently, no similar study exists for the Turkish context.

In a study with an experimental design, 76 undergraduate students from a private school in Algeria were separated into an experimental and control group for the purposes of testing ChatGPT-4 as a tool to teach and learn English writing (Boudouaia et al., 2024). The results display that the experimental group outperformed their pre-tests, and the control group. Furthermore, on the topic of student perception, the students in the experimental group's "views of perceived usefulness, perceived ease of use, attitudes, and behavioral intention" have seen "substantial improvements".

Overall, examining the studies done on student perception regarding ChatGPT revealed that while there are studies that cover the wide areas of the topic, and also a few of the smaller sections and the relationship between ChatGPT perception and other variables; however, the number of the studies and the array of sections on the topic and other surrounding aspects are limited, and an apparent number of gaps in literature can be seen for both different sections contained within the topic, different research designs, a lack of experimental designs as an example, different contexts, and different participants types, background, and other similar aspects.

3. METHODOLOGY

3.1. Introduction

This chapter will describe the aspects and materials around the methodology of this thesis in detail.

The selected research design and justification for how it was selected and used will be discussed. After, detailed descriptions of where the data was collected, how it was collected, and any information around the background of data, location, and participants will be given. Participants will also be described in detail with numerical data for descriptive statistics like gender, school level, age, and so on. After the participants are explained in detail, the instruments of the study will be examined one by one, with justifications for their use, the origin of the surveys with explanations on how the surveys were created in the first place by their original creators, the adaptations made to the current surveys, reliability and validity analysis, and statistical data for such reliability and validity testing.

Subsequently, after the three used surveys were explained in detail alongside the interview questions and planning, the procedure of the whole data collection stage of the study will be explained, followed by how data was analyzed.

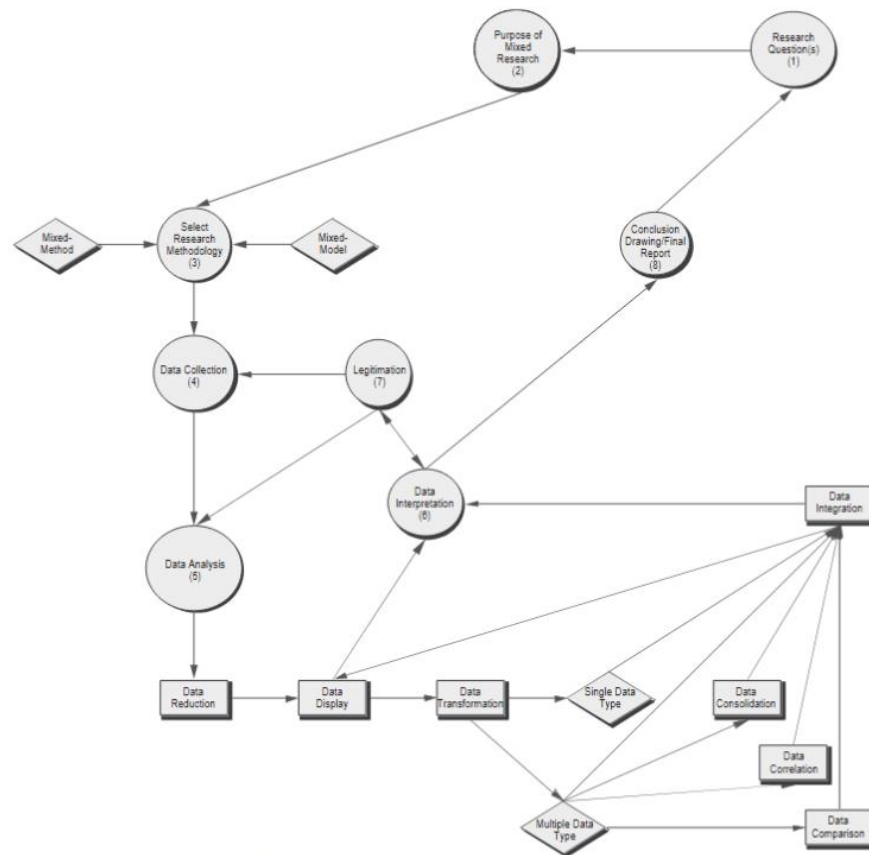
3.2. Research Design

The study for this thesis is based on a mixed-method research design, chosen for the purposes of this study and the aim of answering the research questions and hypotheses set by it. The reason is that while both quantitative data and qualitative data collection and analysis have benefits for the purposes of a study, they also have limitations as well; therefore, in theory, combining both of the methods will combine the strengths of the methods and “overcome the limitations of each” (J. W. Creswell & Creswell, 2017). Furthermore, mixed-method design has been praised in social sciences for the idea that both data collection methods can be integrated and used in ways, steps, and orders where it can provide more structural integrity into a study and its findings. Qualitative data can be used to reinforce quantitative data, vice versa, used in conjunction, one can be used to shape the data collection phase of the other, and so on (J. Creswell et al., 2003). For the purposes of this study, the mixed research process model (R. Burke Johnson & Onwuegbuzie, 2004) was used for the design of the study and the steps that was taken in

order for the study design, application, data collection and analysis, and results (Figure 3.1). No matter the specific type of mixed-method design, it is important for a research design to follow specific and laid-out steps that are thought-out and planned beforehand with the aim of achieving the objectives of the study beforehand the process starts, as can be seen with many examples of mixed-method studies in the literature (R. B. Johnson et al., 2007); therefore, the mixed research process model was followed as an outline for the steps of the mixed-method study design.

Figure 3.1

Mixed Research Process Model



Note. Circles represent steps (1–8) in the mixed research process; rectangles represent steps in the mixed data analysis process; diamonds represent components.

Source: R. B. Johnson et al. (2007)

After careful consideration and planning, the specific method of mixed-method study design chosen for this study was concurrent triangulation design. One justification for not using one of the two data types collected, quantitative or qualitative, to shape the other's data collection tool, method, type, and other aspects surrounding the process is the idea that "the independence of the two study components preserved the assumptive

and methodological integrity of each, thus maximizing the intended value of each set of results” (Greene & McClintock, 1985).

Triangulation of data has been described as the “convergence, corroboration, correspondence or results from different methods” (Bryman, 2006) adding that the emphasis was particularly on corroboration between the two data types, quantitative and qualitative, when both data was simultaneously coded with triangulation in mind. It is also kept in mind that in social sciences, because of the reason that what is being researched is the human and everything that surrounding the person and/or people, the triangulation of qualitative and quantitative data has the potential of exploring and explaining “complex human behavior” more succinctly and effectively (Noble & Heale, 2019).

As per the guide stones of concurrent research design, the qualitative and quantitative data was collected separately, without influencing each other in the data collection phase of the study. Creswell explains that the stage of integration of the data in concurrent triangulation design can occur on the interpretation or analysis phase of the study (J. Creswell et al., 2003). For this thesis, the data was incorporated in the interpretation phase. The reason for this was the objective of looking at the research questions and what could be used from the data to give answers to the research questions from both a quantitative and qualitative lens separately at first, and later adapting both data together to look at the questions from an integrated angle. The separate data analysis is mentioned with the added step of comparing the data results of the quantitative data analysis and qualitative data analysis by Creswell (Figure 3.2) (2009).

As mentioned, the quantitative and qualitative phases of the study are distinct from each other to a certain point, which can also be explained as the use of the two approaches “concurrently, although separately, and integrated – in drawing conclusions” (Maxwell & Loomis, 2003). Concurrent data has also been explained as a way to use quantitative data to inform the aims of the study and the variables in addition to using qualitative data to contextualize the results (Ilic et al., 2013).

Figure 3.2

Concurrent Triangulation Design (Creswell John & Creswell, 2009).



Source: Creswell John & Creswell (2009)

3.3. Research Context

The main aim of the research is to understand the proficiency and attitudes of university students majoring in English Language Teaching towards artificial intelligence language models like ChatGPT. It aims to investigate how often students use these models in their academic lives, how they use them, their opinions about them, and their thoughts on the future of this particular technology. This study focuses on students in Turkey and is conducted within the framework of English Language Teaching education. For this purpose, the research context is set in a state university located in Turkey. Specifically, the university is located in Western Turkey, and from the list of state universities accepting students for the program of English Language Teaching, it is located in the top 10 of universities in terms of the placement points from the state-wide university placement test, Yükseköğretim Kurumları Sınavı (YKS), or Higher Education Institutions Exam in English.

The collection of data took place in the autumn semester of 2024-2025 academic year in the English Language Teaching department of this university, at the starting weeks of the semester. The program requires of total of 240 credits to successfully graduate in addition to a GPA of at least 2.00 out of 4.00, or 50 out of a 100. These credits are provided by lessons who range from elective courses touching on subjects around English teaching and education overall, theoretical courses for foreign language education, principles of education, English literature, linguistics, and teaching practicum, among other lessons.

The planned program takes 8 semesters, or 4 full years of study to complete. The evaluation process changes from lesson to lesson, and is made up of midterms, finals, in-lesson performance, attendance, assignments, portfolios, homework, projects, activities, and so on. Overall, it is a full-time training program with a number of professors, associate professors, instructors, and research assistants. The learning objectives and aims are provided for every lesson and the lesson plans, and the overall program is designed by the principles of education and guides for the English Language Teaching program designated by Yükseköğretim Kurulu (YÖK), or The Council of Higher Education in English.

The data collection process for both the qualitative and quantitative data took place at the same of the autumn semester. For this reason, the participants were in their 1st, 3rd, 5th, and 7th semesters. The reasoning for this decision is to differentiate between students who have taken major courses in English Language Teaching, especially methodology courses, in their 2nd and 3rd year of university to see if there is a significant difference between students who have taken methodology courses concerning their ChatGPT literacy and/or ChatGPT attitude. Because of the timing, 1st and 2nd year students have not taken any English Language Teaching specific courses that the 3rd and 4th year students have taken. This timing also prevents a major distinction between 3rd year and 4th year students in terms of 4th year practicum courses. In the event of data collection being done in the following spring semester, the 4th year students would have completed half of their practicum responsibilities that would differentiate themselves from 3rd year students without finishing their practicum, making it difficult to classify them as studying or in-practice student teachers. The same course of thought can also be seen in 2nd year students. If data is to be collected on the spring semester, 2nd year students would have taken some methodology courses, but considering the number of these courses and the fact that most of these courses are followed by their second part in the following semester, Teaching English to Young Learners I and Teaching English to Young Learners II, for example, it was decided that the best course of action and timing for the data collection to happen at the start of the autumn semester. That is how the justification for the timing of the research process was done for this particular study and thesis.

Because of the objective of keeping the specificity of the university in question private, other information that could be used to pinpoint the identity of the university will not be addressed. However, the information provided so far presents enough information to paint the picture of the context of the research.

3.4. Participants

The total number of potential participants from the English Language Teaching department of the state university the data for this study spans over 200; however, because of constraints and student who did not volunteer to be a part of the study, and a number of questionnaire data taken out of the data analysis due to the inconsistencies and incompleteness of surveys and the questionnaire as a whole, the number of valid participants for the quantitative part of the study was finalized to be 141 students in total. For the interview part of the questionnaire, 8 participants volunteered to be interviewed: 2 of the participants were 2nd year student, 4 of the participants were 3rd year students, and 2 of the participants were 4th year students.

The participants were chosen via non-random sampling, or convenience sampling. Convenience sampling is the act choosing a sample from a population group depending on the availability of the participants, volunteering, resources and the reach of the researcher and/or the researching party. Overall, there are inherent downsides to convenience sampling compared to random sampling, like the risk of inducing bias, the potential for a change in behavior by the participants like the halo effect (Wells, 1907), and so on, although a section of methodology research also points out that the results of research may not change in a significant manner using convenience sampling and representative population-based sampling is done when the principles in appropriate sampling is followed (Mullinix et al., 2015). Even so, the potential for bias alongside other appropriate data analysis methods have been taken into consideration in this study that will be explained in detail in the section for the procedure and analysis of data in this study.

In Table 3.1, the frequency statistics for the participants can be seen. As a sub-objective for this thesis, data analysis has been done between the participants' gender and other data collected, and the findings will be shown and discussed later on.

Table 3.1*Frequency table of the participants*

Characteristic	Participants	
	<i>n</i>	%
Gender		
Male	47	33,3
Female	91	64,5
Other	3	2,1
Age		
18	24	17,0
19	25	17,7
20	29	20,6
21	27	19,1
22	6	4,3
23	5	3,5
24-28	3	2,1
Missing	22	15,6
University Level		
Freshman (1 st year)	39	27,7
Sophomore (2 nd year)	36	25,5
Junior (3 rd year)	34	24,1
Senior (4 th year)	32	22,7
Branch Courses		
Courses Taken	66	46,8
Courses Not Taken	75	53,2
ChatGPT Awareness		
Heard and Used	121	85,8
Heard Only	20	14,2
Total	141	100,0

One limitation that can be seen on the literature for English Language Teaching is the lopsided nature of gender in participant groups and the fact that a significant number of participant groups for research concerning students studying to be English teachers are female focused. While no systematic review was done on the topic, many examples of studies where the gender is significantly unequal in number can be observed, even when the main focus of the study is around gender (Ofiaz, 2019; Tercanlioglu, 2004; Yılmaz, 2010). Similarly, in this study, the number of female participants is 91 and male participants is 47, with 3 participants choosing other as their gender option. Additionally, the 8 participants volunteering to be interviewed for this study were all female participants.

The age statistics of the participants can also be seen in table 3.1. From the 141 participants, 22 participants have opted out of giving an answer for their age, leaving the section blank. The participants were informed of this option beforehand, in written form on the consent form of the questionnaire, and also verbally by the researcher. The age of the 191 participants who have written down their age comes to a mean of 19,98 rounded up to 20 years of age. Only 3 participants are older than 24, and a majority of the participants are within the ages of 18 to 21. This number and percentages are overall in line with the usual university student age in Türkiye.

In terms of university year, while the most amount of data was collected from 1st year students compared to the other years singularly, the differences between the participant numbers in the categories are insignificantly different from one another, ensuring equal resulting data where the differences between different university years can be analyzed in regard to questionnaire and interview results.

For the branch courses category, while this category can be labelled as 1st and 2nd year students compared to 3rd and 4th year students, there have been a number of exceptions where students who have transferred from different universities, taken courses in an uncommon order, and other variables have slightly changed the data compared to the university year category; therefore, it is also paramount to analyze the data and analyze the population depending on whether they have taken branch courses for English Language Teaching pedagogy.

Lastly, the last section of the table displays the ChatGPT awareness of the students. The questionnaire was only given to students who have at least heard of the large language model and additionally to those who have used it before as well. Therefore, no questionnaire was administered to a student who was unaware of the existence of ChatGPT. As a note, only one student reported to be completely unaware of what ChatGPT was, and a vast majority of the students have not only heard of ChatGPT before, but used it as well, as the numbers in the table being that 121 participants have used ChatGPT before, and only 20 of them have only heard of ChatGPT.

The number of participants for the interviews is numerically low, being 8 in total. While the number of the participants who took part in the interviews were low, the output of data from the interviews totaling longer than an hour of recording can be considered to be high compared to the participant number not only in terms of duration, but also in terms of data in the form of themes and sub-themes gained from it, in addition to many excerpts and dialogue which were highly useful in terms of improving the results of this study. The resulting participant number was therefore seen as adequate for the purposes of this study in terms of qualitative data collection and analysis, and comparison of the qualitative data with the quantitative data.

Lastly, through the course of the thesis, the participants are addressed both as “students” and also “prospective teachers”. These two terms are interchangeably used for the purpose of signifying which of their roles needs to be highlighted for the purposes of the section the participants are referred. For example, if the participants are discussed through their answers given to questions like “would you use ChatGPT as a teacher?”, they may be referred to as prospective teachers. However, there are no separate student and prospective teachers, and both terms refer to the same group of participants, being the total number of them.

3.5.Instruments

The instruments used for the purposes of the study are three different surveys constructed into a single questionnaire given to the participants, and questions for a semi-structured interview designed using the facets of the quantitative instruments.

The first survey is an adapted version of the ChatGPT Literacy Scale designed by Lee & Park (2024). The scale has been put through validity and reliability testing to see if there was any change and/or adaptation needed and for the adaptations done for the purposes of this study and the context surrounding it.

The second survey is an adapted version of the Technology Acceptance Model Edited to Assess ChatGPT Adoption (TAME-ChatGPT) (Abdaljaleel et al., 2024). The same testing has been done to this scale as well to adapt it to fit the needs of this study.

The third survey, ChatGPT ELT Attitude survey, was constructed using the aforementioned testing, and through item adaptation from TAME-ChatGPT and the overall literature around the topic of ChatGPT in EFL and ESL to construct a section of the questionnaire where questions would be specifically about learning and teaching the English language.

For the qualitative data collection, semi-structured interview questions were developed around students' proficiency and attitude regarding ChatGPT. The interview questions were adapted from the items changed to fit the context from the three surveys.

The details of the original scales, their validity and reliability testing, the adaptation done, and the detailed testing of the adapted versions of the scales are discussed in the following sections.

3.5.1. ChatGPT Literacy Scale

ChatGPT Literacy Scale, created with the intention of preparing a reliable ChatGPT literacy scale to use for research and practice among the ever-growing interest towards the new technology, was constructed by Lee & Park (2024). The scale was developed with previous tenets of digital literacy used in research measurement and scale development in mind. The scale was separated into 5 main factors, these factors, which will also be the same ones for the adapted survey used for this thesis, are:

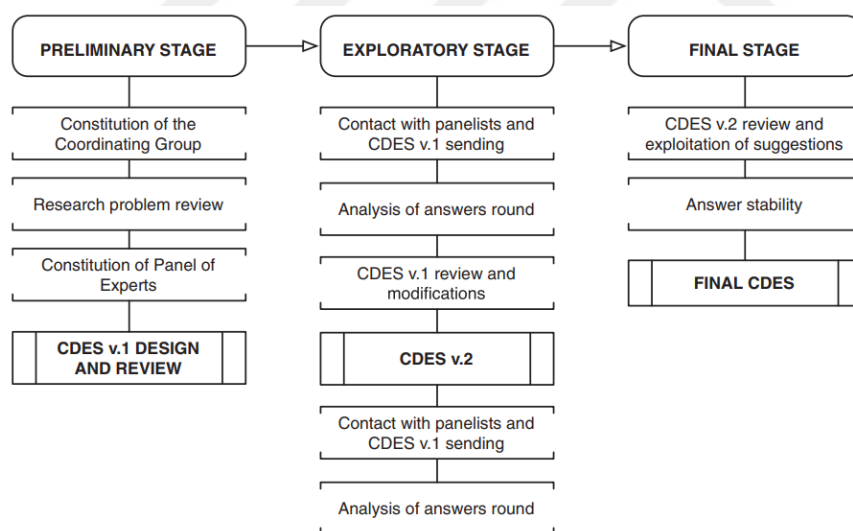
- Technical Proficiency
- Critical Evaluation
- Communication Proficiency
- Creative Application
- Ethical Competence

When the items were developed first for this scale, several steps were followed starting with an extensive literature review on digital, media, and AI literacy. After, other related scales were reviewed, followed by the collection of relevant items from those scale and their adaptation for the purposes of the ChatGPT literacy scale. After a focus group of five experts in AI, natural language process, digital and media literacy, a brainstorming session was done through the five factors established for the scale. After everything was completed, the scale had 34 items.

Next, Delphi validation was done to the items. The Delphi method is the act of developing, validating in this case, a construct through a consensus view across experts who are knowledgeable experts in the subject area (Mengual-Andrés et al., 2016). While the exact steps and focus can be changed depending on the context, the overall outline of the stages of the Delphi technique can be seen in Figure 3.3.

Figure 3.3

Stages of the Delphi Technique



Source: Mengual-Andrés et al. (2016)

After the Delphi technique was used to receive feedback from 10 experts who were “professionals from diverse fields such as computer education, instructional technologies, communication research, and psychometrics.” After the items were rated by these experts, further development was done to delete some items, adapt some items, which was followed by evaluating the content validity of the items. After analyzing the content validity ratios of the items, six items were removed, decreasing the number of

items to 28. A pilot test was conducted with the 28-item scale, with 10 volunteering university students. They were administered the tests and were asked about the face validity of the items regarding their “phrasing, clarity, and logical organization”. After the pilot test was done and the results were deemed adequate, a larger scale test was done with 822 participants. The participants were 49.5% female and 50.5% male and the mean value for their age was 22.7 years with a 2.64 standard deviation. The results were first used to analyze the construct validity of the questionnaire. Using exploratory and confirmatory factor analyses, additional changes were made.

Firstly, the exploratory factor analysis was done on the scale using Kaiser-Meyer-Olkin analysis and Bartlett’s Test of Sphericity on the sample. For the Kaiser-Meyer-Olkin analysis, the result was 0.91, above the recommended limit of 0.60. Bartlett’s Test of Sphericity turned out to be statistically significant, concluding that the exploratory factor analysis of the test was appropriate.

After, the confirmatory factor analysis was done using standardized root mean square residuals, root mean square error of approximation, goodness-of-fit index, comparative fit index, and Tucker-Lewis index. The results of the data revealed that the scale had adequate construct validity after 3 additional items were deleted, leaving 25 final items.

Consequently, the analysis of the data was followed by convergent and discriminant validity testing, both turning out to be adequate. After common method bias analysis was done, it was explained that “common method bias did not pose a significant issue in this research” (S. Lee & Park, 2024).

A systematic review done on different AI literacy scales (Lintner, 2024) has also shown the ChatGPT literacy scale as the one of the “scales with the most robust quality evidence” that involved “good evidence for structural validity, internal consistency, and construct validity.”

For the purposes of this study, ChatGPT Literacy Scale has gone through several stages of adaptation. Firstly, similar to the first construction of this scale, the items of the scale were given to 3 experts in the field of ELT, in addition to areas involved in using technology to teach and learn English, in the form of expert opinion forms.

The results of the form revealed that there were a number of changes needed to be done to the scale to fit the appropriate needs and context of this study. Firstly, to help participants not have any misunderstandings or frustrations with terms, the first item of “I can train and fine-tune ChatGPT for specific purposes or applications” have been modified to include “adjust prompts according to your needs” in parentheses after the word fine-tune to make sure students do not have difficulty understanding the item, which was especially an important aspect kept in mind in the adaptation process due to the fact that the questionnaire was presented to senior-level students all the way down to freshman-level students as well.

Fifth item of “I can understand how ChatGPT generates responses” was deleted from the questionnaire to not confuse students and also due to the fact that the item was covered with a wider item that was the sixth item “I can understand how ChatGPT works”. Ninth item “I can evaluate the reliability of ChatGPT’s responses” was also deleted for the same reason. Twelfth item “I can recognize and explain bias in ChatGPT’s responses” had a change where the word “explain” was deleted to make sure the item did not ask two separate questions to the participant at the same time. Item 15 “I can use technical terms in conversations with ChatGPT” was lengthened to include more clarity, and was adapted to “I can use technical terms related to giving prompts to ChatGPT in conversations with ChatGPT.”

Additionally, because of the unclarity of “rich vocabulary”, item 16 “I can ask accurate questions to ChatGPT using rich vocabulary” was shortened to “I can ask accurate questions to ChatGPT”. A small distinction was added to “I can do creative writing or storytelling using ChatGPT” which was item 19 that changed to “or” to “and/or”. The same was also done to item 21 “I can improve my creativity or innovation skills using ChatGPT”

Because of the possibility of students not being acquainted with terms such as large dataset trends and insights, item 20 “I can use ChatGPT to generate insights and trends on large datasets” was changed to “I can use ChatGPT to analyze large sets of data”.

Item 22 “I can identify potential ethical issues associated with the use of ChatGPT” was clarified with examples and changed to “I can identify potential ethical issues (e.g., plagiarism, copyright) associated with the use of ChatGPT”. Item 23 “I can

explore legal or ethical considerations related to the use of ChatGPT” was deleted due to redundancy with the item 22. The same situation occurred with item 25.

At the end of the process, a number of items were adapted or deleted. No additional items were added for this survey. The adapted version of the scale had 21 items in total: 5 items for factor 1, technical proficiency; 5 items for factor 2, critical evaluation; 5 items for factor 3, communication proficiency; 4 items for factor 4, creative application; and lastly, 2 items for factor 5, ethical competence.

Using the answers to the ChatGPT literacy scale given by the participants, the adapted version of the study was analyzed for the reliability and validity of the survey, and the data gained from it.

Firstly, statistical analysis was done to see the reliability of the questionnaire. Based on reliability statistics, the Cronbach’s Alpha coefficient of the items in the adapted ChatGPT literacy scale came to ,920. Cronbach’s Alpha can be described as the “estimate of the correlation between two random test items” (Cronbach, 1951). A reliability coefficient “demonstrates whether the test designer was correct in expecting a certain collection of items to yield interpretable statements” (Kelley, 1942). The resulting Cronbach’s Alpha value of ,920 for this adapted questionnaire is a high enough number to signify the reliability of this survey.

For factor analysis, the first analysis done to the survey was the exploratory factor analysis for both Kaiser-Meyer-Olkin analysis and Bartlett’s Test of Sphericity. The results have shown that the Kaiser-Meyer-Olkin Measure of Sampling Adequacy was ,873, Kaiser (1974) and Kaiser & Rice (1974) mention the levels of factorial simplicity as follows:

- .90s, marvelous
- .80s, meritorious
- .70s, middling
- .60s, mediocre
- .50s miserable
- Below .50, unacceptable (“Commentaries on the Ten Most Highly Cited Psychometrika Articles from 1936 to the Present,” 2016)

From these sources, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy of ,873 for this scale is in the high section of the meritorious category.

Additionally, the significance value found resulting from the Bartlett's Test of Sphericity ($\chi^2=1453.665$, $df=210$, $p<.001$) is adequate enough to describe the scale's construct validity (Field, 2009) as high in addition to findings by the Kaiser-Meyer-Olkin analysis.

Overall, the adapted scale was developed and analyzed for its validity and reliability values and adapted towards the use for university-level students from freshman-year to senior-year studying to become English teachers, studying in a state university in Türkiye, Western Anatolia. The final survey administered had 21 items, all having the same 5-point Likert Scale spanning from "5=Very Much", "4=Considerably", "3=Neutral", "2=Slightly", "1=Not at all". All the items were positively coded, and there were no reverse items. The survey was given to the participants in its original English format, although a Turkish-translated version of the survey was at hand if any participants deemed the English difficult to understand on the survey; however, no inquiries were made by the participants regarding this topic.

3.5.2. TAME-ChatGPT

The second survey used in the questionnaire is the adapted version of TAME-ChatGPT. The questionnaire, similar to the first questionnaire, was developed similarly for the purposes of collecting data about ChatGPT; for this scale's case, students' attitudes. Because of the reason that the questionnaire was already developed with university students in mind, a lesser number of changes to the overall wording of the scale was made.

Tame-ChatGPT, developed by Abdaljaleel et al. (2024) developed with the aim of helping enrich the literature by being a scale that can be used for attitude research regarding ChatGPT, was used to be able to collect data about ChatGPT attitude for the purposes of researching the correlation and/or relationship between other variables related to ChatGPT, AI, education, and so on; and to "unravel the multifaceted aspects driving the adoption of ChatGPT among university students for educational purposes". The scale, like many scales developed around technology-related education tools, is based around the technology acceptance model (Marangunić & Granić, 2015).

To test the scale, a study involving 2240 participants from the countries of Egypt, Iraq, Jordan, Kuwait, Lebanon was done. The mean age of the participants were 22.25+4.58 years, and 72.1% of the participants were female, pointing out the lopsided nature of gender in numerical data from education; however, the participant group does fit the gender statistics of the study for thesis as well. The scale is developed in a way where there are two sections to the questionnaire. The first section would be done by the participants who have only heard of ChatGPT, and the second part would be for students who have both heard and used ChatGPT before the collection of the data.

The survey is separated into 7 sections which are perceived usefulness, behavior, perceived risk of use, perceived ease of use, anxiety, technology social influence, and perceived risk. The first four factors were given for the usage scale while the last three factors were for the section of “heard of ChatGPT but did not use it”.

Overall, perceived usefulness section had 6, behavior/cognitive factors included 3, perceived risk of use included 3, perceived ease of use included 2, perceived risk included 5, anxiety included 3, technology/ social influence included 5 items.”. From the findings, confirmatory factor analysis was done. Factor analysis was done separately for the two sections of the questionnaire. The Bartlett’s Test of Sphericity and Kaiser-Meyer-Olkin Measure of Sampling Adequacy for the section answered by the participants who have only heard of ChatGPT before included the results $\chi^2/df=300.20/71=4.23$, $p<0.001$. Furthermore, the standardized root mean square residual was 0.050. The standardized root mean square residual is described to be better closer to 0 it is with an acceptable threshold between 0.05 and 0.1 (Tatar, 2015). The root mean square error of approximation of the section was 0.077, which is also within the adequate range, in addition to the comparative fit index being 0.923 and Tucker-Lewis index being 0.901.

For the section of the questionnaire answered by participants who have used ChatGPT before Bartlett’s Test of Sphericity ($\chi^2/df=436.67/62=7.04$, $p<0.001$) was adequate for the results, so was the root mean square error of approximation (0.076), standardized root mean square residual (0.038), comparative fit index (0.942), and Tucker-Lewis index (0.927).

From the maximum scoring of sections, lines were analyzed to be found where the overall attitude can be described as generalizable to the sample. For the first section of perceived usefulness, from a maximum score of 30, a score of 18 indicated neutrality

and a score equal or below 6 indicated disagreements that ChatGPT is useful. For behavioral/cognitive factors, neutral score was 9 out of a maximum of 15 and 3 indicated negativity. For perceived risk of use, neutral was 9 out of 15, and 3 indicated negativity. For perceived ease of use, 6 out of 10 was neutral and 2 was disagreement. Fifth section was classified to signify 15 out of 25 as neutral, and below 5 as low risk. For anxiety, 9 out of 15 was neutral and 3 was low anxiety. Lastly, attitude to technology and social influence was a maximum of 25 and while neutral was 15 points, negative attitude was 5. The third factor “perceived risk of use”, fifth factor “general perceived risks”, and sixth factor “anxiety” was reverse coded.

For this study of which the main aim was for developing the questionnaire, the results revealed that the participants were in agreement with the perceived usefulness of ChatGPT, saw ChatGPT as a positive influence, had a low perceived risk, agreed with the perceived ease of use, general perceived risk was low as well as anxiety, and saw ChatGPT as a positive influence in terms of technology/social influence.

For the purposes of this thesis, an adapted version of the questionnaire was developed. Similarly to the adaptation process for the ChatGPT literacy scale (S. Lee & Park, 2024), the scale was presented to three experts in ELT and technology use in ELT for the purposes of adapting the questionnaire depending on the opinions of the experts. From the results, some items were removed or changed.

Firstly, before the change occurred in any items, the two sections were changed places. For the original scale, the first section the participants would see was the section for participants who have heard of ChatGPT and used it, while the second section was for participants who have heard of ChatGPT but did not use it. These two sections were changed places to better explain which section the participants were meant to answer depending on their experience with ChatGPT.

For the items, the third item’s wording was changed from “I recommend ChatGPT to my colleagues to facilitate their academic duties” to “I recommend ChatGPT to my friends to facilitate their academic duties”. For item 4, the sources of information mentioned in “ChatGPT is more useful than other sources of information that I have used previously” was clarified with “(Google, social media, etc.)” given in parentheses after the word information. The fifth item was removed due to redundancy with the second item. The sixth item “I believe that using ChatGPT can save time and effort in my

university assignments and duties” was shortened to remove the part “-in duties” from the statement. The eighth item “I spontaneously find myself using ChatGPT when I need information for my university assignments and duties” was changed to “ChatGPT is the first source I use when I need information for my university assignments” for clarification and to combat any potential of confusion from the participant. The ninth item was removed due to redundancy with the previous item. The tenth item “I am concerned that using ChatGPT would get me accused of plagiarism” was deleted for redundancy within items apparent in the following items in perceived risk category of the questionnaire as well as the items 11 “I am concerned about the potential security risks of using ChatGPT” and 12 “I think that relying on technology like ChatGPT can disrupt my critical thinking skills”. Because of this, the sections for the perceived risk of use and perceived risk sections were made into one for the adapted version of the scale with the consideration that the participants who have used ChatGPT will answer both sections, while the ones who did not use ChatGPT but have heard of it will only answer the one section; therefore, deleting a redundant section from the usage scale was appropriate with the opinions of the experts.

For the second section of the questionnaire, a number of adaptations have been made as well. Firstly, item 1 was clarified from “I am concerned about the reliability of the information provided by ChatGPT” to “I am concerned about the reliability/accuracy/completeness of the information provided by ChatGPT”. From item 4 “I am afraid that the use of ChatGPT would be a violation of academic and university policies”, the word “afraid” was changed to “concerned” to fit in line with the surrounding items’ wording. Item 5 was deleted for the reason of redundancy with item 2. From item 8 “I am afraid that using ChatGPT would result in a lack of originality in my university assignments and duties” the part of “-and duties” was deleted for the same reason it was deleted for the sixth item in the first section of the scale.

For the final questionnaire, a total of 22 items remained from the original 27 and 6 sections instead of 7. For the heard of ChatGPT but did not use it section, the factor of perceived risk had 5 items, anxiety had 3 items, and technology/social influence had 5 items for the adapted version of the questionnaire. For the heard of ChatGPT and used it section, perceived usefulness factor had 5, behavior/cognitive factors had 2, perceived risk of use was deleted as a whole, and perceived ease of use had 2 items.

Using the data gained by the adapted TAME-ChatGPT scale, the adapted scale was analyzed to find out the reliability and validity of the survey as well as the data itself. Firstly, the first 8 items for the first two sections, which were perceived risk and anxiety, were reverse coded adequately. Then, the reliability of the first 13 items in terms of Cronbach's Alpha, or the section for participants who have heard of ChatGPT but did not use it came out to be adequate ($\alpha=.767$). The results of the Bartlett's Test of Sphericity ($\chi^2=549,409$, $df=78$, $p<.001$) were analyzed to be adequate for the purposes of the exploratory factor analysis in addition to a Kaiser-Meyer-Olkin Measure of Sampling Adequacy of ,721.

For the second section of the scale answered by participants who have heard of ChatGPT and used it, the same analyses were done. Firstly, the Cronbach's Alpha value of ,848 was found, deeming the scale adequately reliable. Additionally, the Bartlett's Test of Sphericity ($\chi^2=435,428$, $df=36$, $p<.001$) displayed similar results for the factor analysis of the second section of the adapted TAME-ChatGPT. Finally, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy was found to be ,812, showcasing the validity of the second section of the scale.

Overall, the two sections separately were found to be reliable for the purposes of this study. Lastly, the two sections were analyzed together for the overall reliability of the adapted scale. The findings revealed that with a Cronbach's Alpha of ,861 which mean the scale had strong reliability, and adequate values for the Bartlett's Test of Sphericity ($\chi^2=950,122$, $df=231$, $p<.001$) and Kaiser-Meyer-Olkin Measure of Sampling Adequacy score of ,725 meant that the entirety of the scale overall had adequate validity as well.

Overall, the adapted scale was created and analyzed for its reliability and validity to properly adapt for the purposes of using it to collect data from university-level students from freshman-year to senior-year studying to become English teachers, studying in a state university in Türkiye, Western Anatolia. The final survey had 22 Likert-scale items spanning from "5=Strongly Agree", "4=Agree", "3=Neutral / No Opinion", "2=Disagree", "1=Strongly Disagree". The participants who used ChatGPT were informed to fill out the entire scale while the participants who only heard of ChatGPT were informed to only fill out the first 13 forms. The first eight items were reverse coded in data analysis, the rest of the 14 items were positively coded. The survey was given in

English to the students not unlike the rest of the questionnaire with the option to opt-out for a Turkish one, although no request was made by the participants.

3.5.3. ChatGPT ELT Attitude Survey

The third questionnaire, ELT Literacy survey, was designed to with the aim of addressing the need to investigate the participants' attitude towards ChatGPT from an educational and purely ELT-based angle. The questionnaire was designed with the help of TAME-ChatGPT and the way the items are used in the questionnaire to elicit relevant information about the attitude of the participants regarding ChatGPT alongside the items themselves. The questionnaire was designed with the act of writing down a high number of items thought out to be relevant to the aims of the questionnaire using similar studies done in the field of English education and the overall facets they emphasize around the topic (Jamshed et al., 2024; Namaziandost & Nasri, 2019; Phosa, 2024; Pokrivcakova, 2024; Zhi & Wang, 2024) alongside the scales adapted for this thesis, which were then lowered in number with the cooperation of the researcher and the thesis supervisor. After, the remaining items were examined carefully to ensure the understandability of the items, wording, order, length, and overall face validity alongside the relevancy of the questions for the scale overall as well as the subsection it is made up of.

The resulting 37 item questionnaire is made up of 4 different sections. These sections are teaching, evaluation, lesson design/material design, and attitude towards ChatGPT as a teacher. The items are all positively coded and follow a 5-point Likert scale of "5=Strongly Agree", "4=Agree", "3=Neutral / No Opinion", "2=Disagree", "1=Strongly Disagree". The first 8 items are in the category of teaching, the following 13 items are in the section of evaluation. The next 5 items are for the lesson design/material development part of the survey, with the last 11 items being in the attitude towards ChatGPT as a teacher category.

The collected data from the study was used to statistically analyze the validity and reliability of the ChatGPT ELT survey.

Firstly, to test the reliability of the questionnaire. Cronbach's Alpha was used. Cronbach's Alpha was calculated for the entirety of the survey, as well as its subsections. The resulting values for the sections of teaching ($\alpha = .889$), evaluation ($\alpha = .949$), lesson design/material development ($\alpha = .913$), attitude towards ChatGPT as a teacher ($\alpha = .948$),

and for the overall ChatGPT ELT survey ($\alpha = .972$) showcased a high value of reliability which was more than adequate for the purposes of this study and thesis.

The results of the following Kaiser-Meyer-Olkin Measure of Sampling Adequacy was shown to be a high value of ,908, in addition to significant data shown by the Bartlett's Test of Sphericity ($\chi^2=4315,621$, $df=666$ $p<.001$). Additionally, from total variances analysis with the initial Eigenvalues of the items, it can be seen that the first 6 items are a cumulative of %75,12 predictor for the rest of the questionnaire, showing strong correlation in the survey items overall.

From the results of the following tests, the validity and reliability of the survey and the data collected via its use was deemed to be adequate for the purposes of this study. The survey was used to collect data alongside the other surveys, with participants who have not used ChatGPT but only heard of it and participants who have heard of ChatGPT and also used it answering the survey.

3.5.4. Interview

For the qualitative data, 21 item list of interview question were prepared. The questionnaire was prepared with the aim of shedding light, more detailed and personalized information and data to the overall topics addressed within the three surveys that are presented together to shape up the quantitative data collection tools for this research. The questions were prepared with the information garnered from the literature around the topic of AI, chatbots, and ChatGPT in education and especially English education to ensure content validity of the 21-item interview questionnaire.

The questionnaire has been taken through a process of revising which resulted in the addition of items, removal of items, rewrites of certain wording, some items were adapted, changed places, and so on, to ensure the data collection through semi-structured interviews which would flow better with the discussion topics to ensure effective data gathering and analysis.

The use of an expert opinion form was also utilized the same way as it was for the other questionnaire items for the revising process of the interview questions. Three experts in ELT, academia, and education-based technology were asked to look at the items of the questionnaire and the interview questions to improve the content and face

validity of the items. This process led to additional items being included to the interview to its final 21 item count.

Items like item 1 “Where and when did you hear about ChatGPT first?” and item 2 “How often do you use ChatGPT for your studies? Can you give a recent example?” were created to understand the background of the students who as a pre-requisite had to be in the category of participants who have used ChatGPT before, their extent of ChatGPT use and overall understanding. While the quantitative data is useful in generalizing the certain aspects around ChatGPT use and understanding, the personalization of ChatGPT experiences and examples was kept in mind when designing the items.

Furthermore, the items were designed to touch on students’ understandings of ChatGPT, and therefore their literacy, in addition to their overall attitude towards it, both as students, ELT students, and prospective teachers. The perceived benefits and limitations of ChatGPT gathered from the literature were addressed in the questionnaire to gather the reactions and responses to the concepts as well.

The interview questions, in addition to the aforementioned, were also designed to correspond with at least one research question. For research question 1: “What is the ChatGPT literacy of ELT students”, interview questions 1, 2, and 3 are solely for the research question. However, while a question like “What can be some ethical considerations for ChatGPT use” directly relates to the participants’ understanding the ethical concepts around ChatGPT use, directly relating to ChatGPT literacy, a question like interview question 21 “Overall, what do you think is the main advantage and disadvantage of ChatGPT?” can be used for both research question 1 and 2, signifying that the answer to this question may give information to the research about the participants’ literacy and also their attitude towards ChatGPT. Interview question from 5 to 21 are all used to answer research question one. Because all interview questions are used to answer research question 1 or 2, or both, they are also used for research question 3 for the reason that participants’ literacy and attitude are correlated in that particular research question. Research question 4 and 5 are indirectly answered through all research questions, especially questions like 6 to 13, 15, 16, and so on, for the reason that the students’ answers may change depending on whether they have taken compulsory pedagogical lessons in the field, which can be observed through their use of ELT-based terminology and information in their answers, and whether such a variable change

depending on the students' university year, which is almost directly related to whether they have taken compulsory ELT-based pedagogical lessons in the field or not.

The implementation of the items was considered in the act of designing and shaping the items and the overall interview questionnaire in regard to the semi-structured nature of the interview. The items were planned around the idea of eliciting responses and unique opinions if they are to occur, the potential for additional clarifying and follow-up questions were planned alongside the main items. Finally, the understanding that while the qualitative data was to be analyzed in a vacuum, the act of combining the quantitative and qualitative data was to occur at a later stage of data analysis in this study in particular, due to the concurrent triangulation design of this mixed-method study, and the questions were designed with these facts in mind.

3.6. Procedure

The data was collected over the course of 2 weeks at the start of the Autumn period for the 2024-2025 educational year in a state university's department of English Language Teaching in the Faculty of Education.

Quantitative data was collected with the consent of instructors after the end of a two-hour lesson for students from all different years and classes. The students were given handouts which included a detailed consent form, a preliminary form for participants' descriptive information like age, gender, their ChatGPT experience, and so on, followed in order by the adapted ChatGPT literacy scale, adapted TAME-ChatGPT, and ChatGPT ELT survey. A short introduction was done by the researcher who was present at every step of the data collection phase, about the nature of the study, and voluntary nature of it. The researcher also asked twice, once at the beginning and once at the ending phase of the quantitative data collection procedure for volunteers who have had any ChatGPT experience for the qualitative data collection procedure, which was also briefly explained. Additional data was gathered from different, randomly selected classes if the number of participants were low enough to make the act of statistical analysis not as valid and/or reliable. The classes overall were chosen randomly added to the fact that the separation of classes for the students studying the same year has been done randomly by the institution as well.

For the quantitative data collection, the administration of the questionnaire usually did not take more than 20 minutes. Overall, students took between 10 to 20 minutes to

consent to the study itself, answer the demographic questions, and answer questionnaires made up of the 3 surveys. Anecdotally, the higher the students' university year, the less time it took for them to answer the questions, presumably due to a better understanding of English and the education-based concepts present in the questionnaire items.

The qualitative data collection was usually done an hour to a number of days after the quantitative data was collected from the class depending on the availability of the participants. The quantitative data was gathered in the physical classrooms after the lessons that have just taken place, and the qualitative data was collected in an office environment in the same building. The semi-structured interviews were voice recorded for the purposes of data collection and analysis. The participants were informed about the nature of the interview and the aspect of voice recording in the written consent form in addition to the researcher asking the participants before the voice recording starts verbally for additional consent.

Usually, the midterms in the university curriculum in Türkiye takes place on the 8th week of the classes out of 14 weeks in total. The data collection process for both the quantitative and the qualitative data was finished more than a month before the start of the midterms. The timing of the data collection was also considered due to the fact that the process would not take time away from students in more time-restrictive and potentially stressful stages of the term closer to the midterms and/or finals, and would rather take place at the start of the term, where the classes are their simplest without any major deadlines for assignments, and so on.

No issue regarding finding volunteers for the qualitative section of the data collection was observed; although, no further restrictions to limit the number of participants were necessary either due to the fact that from the second year students, 3 students volunteered and the interview itself was done with 2 students due to the availability issues of the third student, 3 3rd year students volunteered and were interviewed, and 2 4th year students volunteered. No other participant volunteered for the interviews. No qualitative data was collected from 1st year students due to the certain pedagogical and ELT-based proficiencies not being reached for the 1st year students, like the detailed concepts around lesson planning, skills teaching, and so on, which were present as concepts in the interview questions. Considering the fact that some terms were

needed to be explained to the interviewed 2nd year students as well, the decision to opt out of interviewing 1st year students was deemed to be the appropriate decision.

All questions were asked and answered by the participants without any issues, with answers of the participants leading to further clarification and follow-up questions relating to different concepts. The results of such findings alongside the detailed breakdown of the interviews and the properties surrounding the interviews will be discussed in detail in the following sections of this thesis, as well as the breakdown of how the quantitative data was analyzed, first in isolation, then in conjunction with the qualitative data.

3.7. Analysis of the Data

For the purposes of researching the literacy and attitude of English Language Teaching students towards ChatGPT in Türkiye, both quantitative and qualitative data was collected and analyzed. The quantitative data was collected via a questionnaire made up of 2 adapted scales, ChatGPT Literacy Scale, TAME-ChatGPT, and one constructed survey, ChatGPT ELT Survey. The qualitative data was collected using the constructed semi-structured interview data collection method. The data was analyzed with the tenets of concurrent triangulation design in mind, meaning that the quantitative and qualitative data was analyzed separately before being analyzed together to answer the research questions laid out for the purposes of this study.

The quantitative data was analyzed using IBM SPSS 26 release version. The qualitative data was analyzed without the use of any software. The transcriptions of the interviews were done by the researcher in addition to the thematic analysis.

To answer the research questions, the quantitative data went through different statistical analysis methods depending on the sought-out data. Firstly, to answer the overall ChatGPT literacy, ChatGPT attitude, and ChatGPT ELT attitude, descriptive statistics was used to see the means and standard deviations of the results of the questionnaire. To compare the different sections of the scales in addition to comparing the scales themselves with one another, the Spearman correlation formula was used. The Spearman correlation formula posits that “true correlation may be obtained by dividing the average correlation for the various trials of the two tests by the square root of the product of the correlations for the successive trials for each test” (Wissler, 1905), making it an

appropriate formula to see the correlation between the results of two tests done by the same participant group (Spearman, 1904).

After the act of analyzing the correlation between the results of the scales, the use of independent samples t-test was incorporated to see the correlation between the results of the three separate scales and the independent variables like gender, whether the participants have only heard about ChatGPT or used it, whether the participants have taken branch courses are not. The independent samples t-test is used to compare “the means of two groups” (Ross & Willson, 2017).

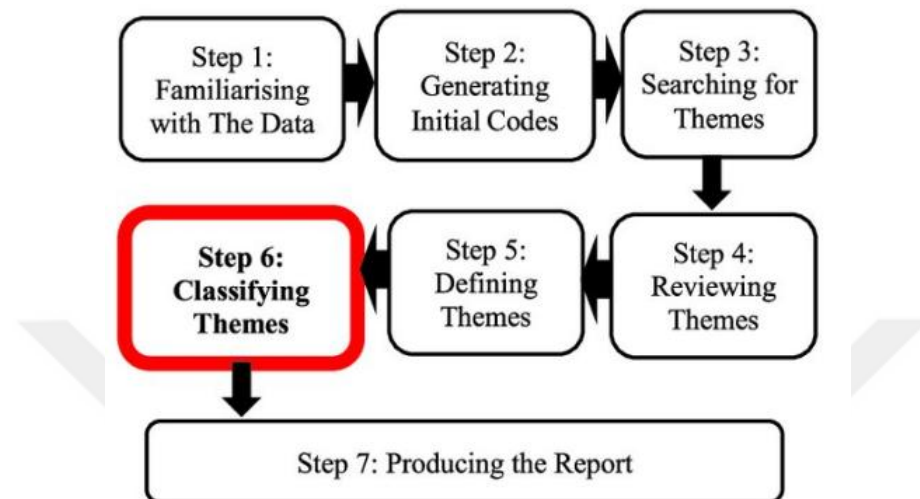
Because of the way the formula for the independent samples t-test is constructed, it can only be used for two groups. For this reason, to analyze the results from the scales and to see their correlation between the year of the student, which is comprised of 4 separate groups from freshman to senior year, One-Way Analysis of Variance (ANOVA) statistical analysis method was used. One-Way ANOVA is explained as the way to determine if there are statistically significant differences between 3 or more groups and is used rather than repeating a number of t-tests (T. Kim, 2017).

For the qualitative data, the method of data analysis was chosen as thematic analysis for the purposes of this study. Thematic analysis is a type of analysis where the themes within the qualitative data is found through the acts of identifying, analyzing, and interpreting (Clarke & Braun, 2017). Within the three different schools of thematic analysis, the one thematic analysis type chosen for this thesis is codebook thematic analysis, instead of codebook reliability thematic analysis or reflexive thematic analysis (Braun et al., 2019). The reason for choosing the particular approach of codebook thematic analysis is for the reason it provides flexibility and researcher-focus during the generation and analysis of themes while also being a middle ground between codebook reliability thematic analysis and reflexive thematic analysis (Ayre & McCaffery, 2022). Additionally, the reason for choosing thematic analysis instead of content analysis which could also be used adequately for the purposes of this thesis is the fact that while they are both similar in their nature of coding, variable and construct manifestation (Marks & Yardley, 2004), content analysis has not been chosen because of its more quantitative nature where the analysis is done through analyzing the qualitative data with a quantitative angle and coding approach. Thematic analysis is a more flexible system in the sense that the data itself can be analyzed alongside the coded data for the purposes of

describing the nature of the data with fewer instances of number and frequency use (Neuendorf, 2018).

Figure 3.4

The seven steps in the Thematic Analysis



Source: Abd Kadir et al. (2021)

The thematic analysis has seven steps that is laid out to be followed in order for a successful thematic analysis to be achieved (Figure 3.4). The first step is to become familiar with the data, where the researcher reads the transcripts multiple times, writes down early notes and impressions. In the second step, the researcher starts to organize the data by generating the initial codes in line with the purposes of the study. The third step is the step where the themes are searched for the first time among the coded data. These broad themes are reviewed and changed accordingly in the fourth step, followed by the definition of themes which happens in step 5, where the research defines the themes in more detail to be ready for the final data analysis of the qualitative data process. In step 6, which can be described as a part of step 5 in different interpretation of thematic data analysis, the research classifies the themes further. For the last step, with the increased in quality of the data over the first six steps where the codes and themes are polished to fit the objectives of the study as adequately as possible, the report is produced by writing down the results (Maguire, 2017). The thematic analysis done in this study follows these steps as described.

As discussed, after the qualitative and quantitative data were analyzed separately, both data were integrated to be analyzed together for the objective of seeing the

similarities and differences between the two types of data, and the implications for the literature, further research, pedagogy, and similar research-related and education-related implications could be made from the results of the study. The data was analyzed for the purpose of answering the research questions. What data collection tool or tools correspond directly to which research question can be seen below in Table 3.2.

Table 3.2

Table for which data collection tool corresponds to which research question.

Research Question	Data Collection Tool
What is the ChatGPT literacy of ELT students?	ChatGPT Literacy Scale Interview Questions
What is the ChatGPT attitude of ELT students?	TAME-ChatGPT ChatGPT ELT Attitude Survey Interview Questions
Is there a significant correlation between ELT students' ChatGPT literacy and ChatGPT attitude?	ChatGPT Literacy Scale TAME-ChatGPT ChatGPT ELT Attitude Survey
Is there a significant correlation between ELT students' ChatGPT literacy and whether if the students have taken compulsory pedagogical lessons in the field?	ChatGPT Literacy Scale Interview Questions
Is there a significant correlation between ELT students' ChatGPT attitude and whether if the students have taken compulsory pedagogical lessons in the field?	TAME-ChatGPT ChatGPT ELT Attitude Survey Interview Questions

3.8. Ethical Declaration

The entirety of the research structure and plan around this thesis was presented to the ethical committee of Anadolu University. The ethical committee named “ANADOLU ÜNİVERSİTESİ SOSYAL VE BEŞERİ BİLİMLER BİLİMSEL ARAŞTIRMA VE YAYIN ETİĞİ KURULU” has deemed this study as a ethically sound study and gave the study a positive response in terms of the ethical structures around it with a protocol number of 744442 on the date of 02.07.2024. The copy of the document approving this study from the ethics department of Anadolu University can be seen in Appendix VII.



4. RESULTS AND DISCUSSION

4.1. Introduction

In this chapter, the detailed results of the study will be given from the data gathered by the three surveys and semi-structured group interviews. The data gathered using the aforementioned three surveys, adapted ChatGPT Literacy Survey, adapted TAME-ChatGPT, and ChatGPT ELT survey, alongside the interview questions constructed through the use of the items in these three surveys, was used to see the results of the objective of this thesis, to find out about the attitude and literacy around the contemporary concept of ChatGPT in university students studying to become English teachers in Türkiye.

The objective of this chapter is to answer the research questions around the topic of ChatGPT regarding ELT students in Türkiye, and to find out the correlation between the various variables this study is set around that includes ChatGPT literacy, ChatGPT attitude, age, gender, university level, and whether the students have taken methodology-related main courses concerning English education. Not only, but students' opinions and thoughts about ChatGPT use as a university student, in academia, in language teaching by teachers and students, and so on will be seen as well as the participants who have joined the group interviews thoughts and opinions about the future of ChatGPT and AI overall regarding education, especially language education and its perceived impacts.

4.2. Results

Five main research questions have been asked for the purposes of this study:

- 1) What is the ChatGPT literacy of ELT students?
- 2) What is the ChatGPT attitude of ELT students?
- 3) Is there a significant correlation between ELT students' ChatGPT literacy and ChatGPT attitude?
- 4) Is there a significant correlation between ELT students' ChatGPT literacy and whether if the students have taken compulsory pedagogical lessons in the field?

- 5) Is there a significant correlation between ELT students' ChatGPT attitude and whether if the students have taken compulsory pedagogical lessons in the field?

To shortly explain the research questions and their purposes, the first research questions are set to understand the level of ChatGPT attitude and literacy of ELT students studying to become English teachers studying in the Turkish context. The aim of these questions is to whether the students have an overall high, low, or average understanding of AI and ChatGPT, and how to use it, how it works, and so on, alongside their thoughts and opinions about various aspects of ChatGPT like its usefulness, positives and shortcomings, ethical issues, and other related general attitude about the technology.

The third question is one of the facets of this study that makes it unique in the literature area it is set in, as no study yet is known to see the correlation between ELT students' ChatGPT literacy and ChatGPT attitude, especially in the Turkish context. Even outside of the field of language teaching, looking at education overall, studies with the same goal are rare to none. The last two questions are asked to see if the students' university levels and whether they have taken compulsory pedagogical courses in the field changes their attitude or literacy towards ChatGPT.

Additionally, while not necessarily being main research questions, the sub-objectives of this study is to also see if there are other correlations between ChatGPT attitude and ChatGPT literacy concerning gender and the university level of the students. Aside from a few cases, most junior and senior students have taken and are taking main pedagogical courses regarding language education, so they will be in the same group when the difference between that variable and ChatGPT attitude and/or ChatGPT literacy are discussed, the same being the case for freshman and sophomore students. For this reason, to see if there are any differences between the years of the students separately as four different groups, the university years of the students will also be used as an additional variable to see if there are any significant findings, alongside gender.

4.2.1. ChatGPT Literacy Scale

The objective of the ChatGPT literacy scale (S. Lee & Park, 2024) is to fit the purposes, population, and context of this study. The survey was given as a part of the questionnaire containing three questionnaires, this survey being the first one to be answered after the participant has read and signed the consent form for the study. The

questionnaire contains 21 different 5-point Likert scale items that are separated into five different categories. The first category, technical proficiency, includes 5 items. The second category, critical evaluation, also contains 5 items, as well as the third category which is communication proficiency. The fourth category, creative application, contains 4 items; and lastly, the fifth category, ethical competence, contains 2 items. The 5 Likert-scale range was designed as follows: “5=Very Much”, “4=Considerably”, “3=Neutral”, “2=Slightly”, “1=Not at all”.

To answer the first research question of this study, “What is the ChatGPT literacy of ELT students?”, the descriptive statistics of the data collected from the participants who have answered the survey, 131 in total, with 124 participants answering every item in the questionnaire.

The descriptive statistics of the results from the adapted ChatGPT literacy scale can be seen in Table 4.1. The results reveal that out of a maximum of 105 and a minimum of a 21 score, the overall mean of the participants’ responses came to be a mean of 76.23. The frequency of the scores can also be seen in Figure 4.1.

Table 4.1

Descriptive statistics of the adapted ChatGPT Literacy Scale results

Variable	<i>n</i>	<i>Min.</i>	<i>Max</i>	<i>M</i>	<i>SE</i>	<i>SD</i>
1. ChatGPT Literacy	124	31	105	76.23	1.383	15.399
2. Technical Proficiency	131	5	25	16.52	.415	4.751
3. Critical Evaluation	129	8	25	18.34	.353	4.005
4. Communication Proficiency	128	6	25	19.62	.374	4.232
5. Creative Application	129	4	20	14.77	.369	4.194
6. Ethical Competence	131	2	10	6.98	.214	2.453

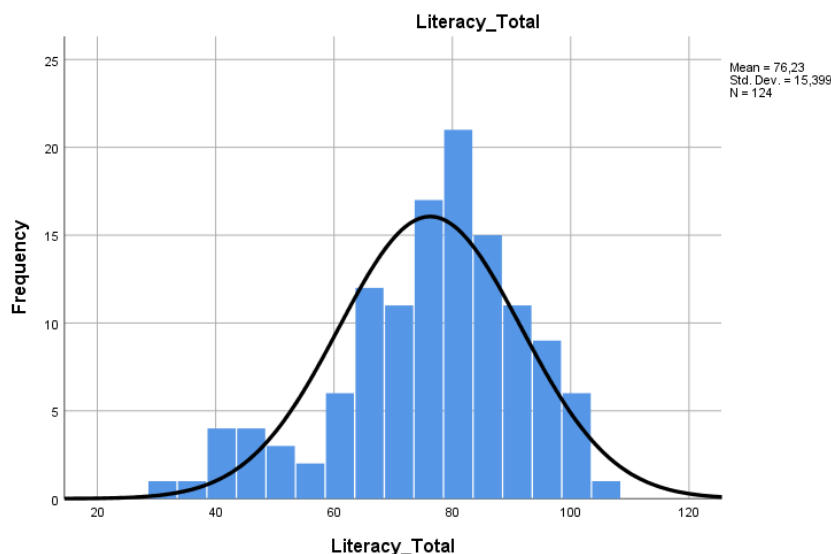
The results indicate that the ChatGPT literacy of the participants are above the possible average mean of the scale. No mean or descriptive statistics of the results are given by the study done by Lee and Park (2024).

Furthermore, to find out if there are any significant correlations between the factors, Spearman's rank-order correlation was used. The results reveal strong, positive correlations between every set of factors, with the strongest correlation being between technical proficiency and communication proficiency ($r(128) = .642, p < .001$), and the least significant correlation being between creative application and ethical competence ($r(129) = .210, p = .012$), still having a p value of $< .05$ and being considered a strong correlation.

The number of studies with the aim of finding out the ChatGPT literacy of the participants are almost zero, with one example being an ahead-of-publishing study done on 543 Chinese EFL learners (Li & Thien, 2025), also using the ChatGPT Literacy Scale. The results of the study display significant correlation between the different factors in the survey. The results of this study present similar results to the significances found in this study, where every combination of correlation between the different factors pertaining to the sections in the questionnaire having a strong correlation with each other. One difference of the study to this thesis is that the participant group of 543 EFL learners are not only made up of students studying for their bachelor's degrees but also students with associate degree and below, undergraduate degrees, master's degrees, and doctoral degrees as well.

Figure 4.1

Frequency graph of the adapted ChatGPT Literacy Scale scores



To find out if there is a statistically significant difference between male and female participants in terms of ChatGPT literacy, independent samples t-tests were employed. In terms of gender, no statistical differences were found between the variable of gender and any singular section in the survey. As a whole, no statistically significant difference was found between male and female participants and the ChatGPT Literacy Scale scores ($t(120) = ,771, p = ,442$).

Additionally, an independent samples t-test was done to see the difference between the students' who have only heard of ChatGPT but did not use it, and students who have both heard of ChatGPT and used it as well concerning their ChatGPT literacy. Levene's test for equality of variances was done to see if the equal variances are to be assumed or not assumed. The results of the test revealed that equal variances were assumed. Furthermore, the results of the independent samples t-test showcase that students who have used ChatGPT have a significantly higher self-reported ChatGPT literacy compared to students who have only heard of ChatGPT ($t(122) = 2,204, p = ,029$).

To find out if there is a statistically significant difference between the students who have taken obligatory ELT-related pedagogical courses and the ones who have not, independent samples t-test formula was utilized. In addition, to see if there is any specific significant distinction between the university year of the students, one-way ANOVA was utilized.

Firstly, to see if there is an apparent significance found between the variable of ELT courses taken or not and ChatGPT literacy, Levene's test for equality of variances was utilized. This test is used for the purposes of finding out where the equality of the variances can be assumed, and therefore have a better possibility of having a statistically significant result (Pratiwi et al., 2025). The results of the Levene's test significance were above .05; therefore, equal variances were assumed. The results revealed that, although very close, no statistically significant difference between ChatGPT literacy scores and whether the students have taken ELT-based pedagogical courses or not ($t(122) = 1.952, p = ,053$). However, looking at the sections of the scale, it can be seen that there is a statistically significant difference between students' technical proficiency and whether they have taken the courses or not ($t(129) = 2.038, p = ,044$). For the overall test, N for the participants who have answered every part of the test was 62 for courses taken, and

62 for courses not taken as well. For the one significant finding, technical proficiency, the N values were 66 for courses taken, and 65 for courses not taken.

In terms of the university year of the students, separated into four categories of freshman, sophomore, junior, and senior, a one-way ANOVA test was utilized to see if there was any statistically significant difference between the students ChatGPT literacy, and its sections, and their university year. The results suggested there was statistically significant difference found for at least two groups concerning the total score of the ChatGPT literacy scale ($F(3, 120) = [4.713]$, $p = .004$), in addition to technical proficiency ($F(3, 127) = [3.956]$, $p = .010$), communication proficiency ($F(3, 124) = [6.029]$, $p = .001$), and creative application ($F(3, 125) = [3.237]$, $p = .025$); in addition, no statistically significant results for at least two groups were observed concerning critical evaluation ($F(3, 125) = [.996]$, $p = .397$) or ethical competence ($F(3, 127) = [.254]$, $p = .858$).

Bonferroni post-hoc test was employed to find the statistically significant correlation individually. The reason for using the Bonferroni correction method, based on Carlo Bonferroni, to compare different groups at baseline and to study the relationship between variables (Armstrong, 2014) in addition to correcting the family-wise error rate, also called the experiment-wise error rate, which is described as the “probability of at least one false positive finding out of a set of significance tests directed at addressing the same research question.” (Yoder et al., 2004). Consequently, while Tukey HSD can also be used to control for false positives, Bonferroni correction is a more conservative approach at the risk of increasing false negatives, but also reducing the probability of Type I errors as a result, which is the family-wise error rate (Nicholson et al., 2022)

The highest and lowest possible scores for the ChatGPT literacy scale and the five factors within it are as follows:

- ChatGPT Literacy survey: 21-105
- Technical Proficiency: 5-25
- Critical Evaluation: 5-25
- Communication Proficiency: 5-25
- Creative Application: 4-20
- Ethical Competence: 2-10

The results of the Bonferroni correction post-hoc test revealed that the mean ChatGPT literacy scale score for freshman students ($M = 68.07$, $SD = 17.629$) were significantly different than sophomore students ($M = 78.39$, $SD = 15.050$). Additionally, the mean score for freshman students ($M = 68.07$, $SD = 17.629$) were also significantly different than senior students ($M = 81.97$, $SD = 15.110$). However, no statistically significant difference between freshman students ($M = 68.07$, $SD = 17.629$) and junior students ($M = 76.03$, $SD = 10.660$); sophomore students ($M = 78.39$, $SD = 15.050$) and junior students ($M = 76.03$, $SD = 10.660$); sophomore students ($M = 78.39$, $SD = 15.050$) and senior students ($M = 81.97$, $SD = 15.110$); or between junior students ($M = 76.03$, $SD = 10.660$) and senior students ($M = 81.97$, $SD = 15.110$) were found (Table 4.2).

Table 4.2

Bonferroni post-hoc results for university levels for ChatGPT literacy scale scores

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-10.325	3.753	.041
Freshman - Junior	-7.962	3.781	.224
Freshman - Senior	-13.898	3.840	.003
Sophomore - Junior	2.363	3.658	1.000
Sophomore - Senior	-3.573	3.720	1.000
Junior - Senior	-5.935	3.747	.695

The results of the Bonferroni correction post-hoc test revealed that the mean score for technical proficiency was only significant for two groups, which were freshman students ($M = 14.23$, $SD = 5.011$) and senior students ($M = 18.09$, $SD = 4.862$) (Table 4.3)

In terms of critical evaluation scores, the results of the of Bonferroni correction post-hoc test displayed that there was no statistically significant correlation between two groups, while the biggest difference between means and standard deviations were found between freshman students ($M = 17.67$, $SD = 4.708$) and senior students ($M = 19.34$, $SD = 3.973$), the results were not statistically significant (Table 4.4).

Table 4.3*Bonferroni post-hoc results for university levels for technical proficiency scores*

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-2.774	1.142	.099
Freshman - Junior	-2.421	1.142	.215
Freshman - Senior	-3.868	1.158	.007
Sophomore - Junior	.353	1.115	1.000
Sophomore - Senior	-1.094	1.132	1.000
Junior - Senior	-1.447	1.132	1.000

The results of the Bonferroni correction post-hoc test revealed that the mean score for the section of communication proficiency of the students exhibited two statistically significant differences between groups, which were between freshman students ($M = 17.16$, $SD = 4.920$) and sophomore students ($M = 20.55$, $SD = 3.519$), and also between freshman students ($M = 17.16$, $SD = 4.920$) and senior students ($M = 21.16$, $SD = 3.407$) (Table 4.5).

Table 4.4*Bonferroni post-hoc results for university levels for critical evaluation scores*

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-.480	1.003	1.000
Freshman - Junior	-.515	1.010	1.000
Freshman - Senior	-1.677	1.018	.612
Sophomore - Junior	-.035	.979	1.000
Sophomore - Senior	-1.197	.986	1.000
Junior - Senior	-1.162	.994	1.000

The results of the Bonferroni correction post-hoc test displayed that the mean score for creative application exhibited a singular statistically significant difference between two groups: between freshman students ($M = 12.90$, $SD = 4.237$) and senior students ($M = 15.97$, $SD = 3.430$) (Table 4.6)

Table 4.5

Bonferroni post-hoc results for university levels for communication proficiency scores

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-3.384	1.001	.006
Freshman - Junior	-2.384	1.001	.112
Freshman - Senior	-4.000	1.016	.001
Sophomore - Junior	1.000	.985	1.000
Sophomore - Senior	-.616	1.001	1.000
Junior - Senior	-1.616	1.001	.653

Table 4.6

Bonferroni post-hoc results for university levels for creative application scores

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-2.482	1.024	.101
Freshman - Junior	-1.806	1.024	.481
Freshman - Senior	-3.068	1.047	.024
Sophomore - Junior	.676	.991	1.000
Sophomore - Senior	-.585	1.015	1.000
Junior - Senior	-1.262	1.015	1.000

In terms of ethical competence, the results of the Bonferroni correction post-hoc test showed no statistically significant difference between the groups (Table 4.7).

Table 4.7*Bonferroni post-hoc results for university levels for ethical competence scores*

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-.352	.614	1.000
Freshman - Junior	-.528	.614	1.000
Freshman - Senior	-.323	.623	1.000
Sophomore - Junior	-.176	.600	1.000
Sophomore - Senior	.029	.609	1.000
Junior - Senior	.206	.609	1.000

Overall, it can be seen that there is a number of statistically significant difference when it comes to students' university years. While not exclusively, a high number of these differences between the groups can be seen belonging to freshman students and senior level students. In fact, the number of significant statistical differences found were 6. Two of these differences belonged to between freshman and sophomore level students, and four of them belonged to between freshman and senior level students. From this, it can be argued that the university year of the student is a predictor of students' ChatGPT literacy to a certain degree, while aside from technical proficiency; overall, whether the students have taken ELT-related pedagogical courses or not is barely not a statistically significant predictor of ChatGPT literacy of the students.

One last variable that was statistically analyzed regarding its correlation with the results of the ChatGPT literacy scale score and scores for its 5 different factors as well, which is whether students have only heard of ChatGPT, or used it as well. For this purpose, the independent samples t-test formula was used. Following the results of the Levene's test for equality of variances, the equal variances for all independent variables were assumed. The results of the independent samples t-test revealed that there was a statistically significant difference between the two groups in terms of total ChatGPT literacy scale scores ($t(122) = 2.204$, $p = .029$). Looking at the different factors individually, only statistically significant difference between the two groups were found was for the communication proficiency score ($t(126) = 3.417$, $p = .001$). Consequentially, it can be argued that using ChatGPT is a predictor for higher ChatGPT literacy; however,

it also needs to be pointed out that the total number of participants who have reported that they have both heard and used ChatGPT is 121, or 85.8% of the total number of participants, while students that reported of only hearing about ChatGPT but not using it is 20, or 14.2% of the total number of participants present in the study.

This section has answered certain hypotheses set before the study in conjunction with the first and fourth research questions; additional discussion of the data of the study in addition to the comparison to other data in the literature will be analyzed further in the fifth section, discussion part of this thesis.

4.2.2. TAME-ChatGPT

The second survey that was present in the entire questionnaire is the adapted form of TAME-ChatGPT, a scale designed to test students' and future teachers' attitude towards ChatGPT (Abdaljaleel et al., 2024).

The test has two different sections. Students who have not used ChatGPT, but at least have heard of it have only answered the first 13 questions of the survey, while students who have both heard of ChatGPT and used it at least once in the past were instructed to answer both sections. Overall, the survey has 22 items, with 13 items being in the 'heard of ChatGPT but did not use it' section and 9 being in the 'heard of ChatGPT and used it' section. Furthermore, the survey is separated into six different sections for six different factors. These factors are perceived risk, which has five items; anxiety, which has three items; technology/social influence, which has five items; perceived usefulness, which has five items; behavior/cognitive factors, which has two items; and finally, perceived ease of use, which has two items.

To answer the research questions; finding out about the ChatGPT attitude score of the participants, and to see if there is a statistically significant difference between students' ChatGPT attitude score and their score in the individual sections of the TAME-ChatGPT scale concerning differing independent variables such as gender, whether the students have taken compulsory ELT-based pedagogical courses or not, their university year, and whether they have only heard of ChatGPT or used it as well. For the purposes of answering these questions, various statistical analysis methods were used.

Firstly, the descriptive statistics for the adapted TAME-ChatGPT scale was calculated. The results of the calculations revealed that students' scores are only slightly

above the very average point of 66 as the minimum possible score of the scale is 22 and maximum is 110, and students have a mean of 69.45. Therefore, it can be expressed that students are neither positive nor negative in their attitude towards ChatGPT (Table 4.8).

Table 4.8

Descriptive statistics of the adapted TAME-ChatGPT results

Variable	<i>n</i>	<i>Min.</i>	<i>Max</i>	<i>M</i>	<i>SE</i>	<i>SD</i>
1. ChatGPT Attitude	65	29	94	69.45	1.743	14.051
2. Attitude (Heard)	84	13	57	38.24	.904	8.284
3. Attitude (Used)	115	13	45	30.94	.700	7.504
4. Perceived Risk	85	5	25	13.04	.517	4.765
5. Anxiety	87	3	15	7.34	.377	3.514
6. Technology/Social Influence	86	5	25	17.73	.500	4.636
7. Perceived Usefulness	116	5	25	17.47	.481	5.178
8. Behavior/Cognitive Factors	115	2	10	5.13	.201	2.158
9. Perceived Ease of Use	116	4	10	8.33	.161	1.733

Firstly, the first section of the scale containing 13 items which is the only section of the scale answered by both students who have only heard of ChatGPT and also by students who have used ChatGPT as well. Independent samples t-tests were done to see if there are any statistically significant differences between these two groups concerning total score of the first section, and the four factors that are present in said section. The results of the Levene's test for equality of variances show the results of the total score of the section's and the perceived risk score's variances can be assumed to be equal, while anxiety score and technology/social influence score cannot. Looking at the results of the t-tests, while the difference is close to being significant between the two groups concerning technology/social influence ($t(58.871) = 1.732$, $p = .089$), no statistically significant difference is found.

With equal variances being assumed for all different sections of scores through the results of the Levene's test for equality of factors, multiple independent samples t-

tests were done to see if there is any statistically significant difference between male and female students. The results of the independent samples t-tests revealed that there is statistically significant difference between male and female students concerning TAME-ChatGPT total score ($t(60) = 2.772, p = .007$), in addition to score for the 'heard but not used' section ($t(79) = 3.011, p = .003$) and perceived risk score ($t(80) = 3.009, p = .004$). It is also important to point that in total, there is 47 male participants making up 33.3% and 91 female participants making up 64.5% of the study, which demonstrates an unequal number of participants in terms of gender.

Lastly, while there are statistically significant differences observed when students who have chosen 'other' as the answer for their gender are compared with participants choosing male and female as their gender; because of the number of the participants who have chosen 'other' as their gender is 3, no valid statistical analysis can be done; however, further studies on this particular subject and participant group can be immensely important for the purposes of filling a significant gap in literature.

The results of the Levene's test for equality of variances concerning the variable of whether students have taken ELT-based compulsory pedagogical courses or not have given results which reveals that for the total score for the scale in addition to the 'used' section, and perceived usefulness, behavior cognition, and ease of use sections, equal variances are not assumed. For the other sections and facets, equal variances are assumed. With this information, several independent samples t-tests were done to see if there are any statistically significant differences between the students who have taken compulsory courses and student who have not.

The results of the independent samples t-tests reveals that there is statistically significant difference between the two groups in terms of the total scores of the adapted TAME-ChatGPT scale ($t(60.172) = 2.718, p = .009$), as well as for the 'heard' section ($t(82) = 2.005, p = .048$), technology/social influences ($t(84) = 4.787, p = .000$), perceived usefulness ($t(90.838) = 2.303, p = .024$), and ease of use ($t(88.705) = 2.678, p = .009$).

The results of the independent samples t-tests display that overall, students have a more significantly more positive attitude towards ChatGPT if they have taken compulsory ELT-based pedagogical courses, as well as having a positive outlook regarding the technology/social influences, perceived usefulness, and ease of use of ChatGPT. Additionally, from the section of the participants who have only heard of

ChatGPT, and not used it, the students who have taken the courses have a more positive outlook towards ChatGPT overall as well.

To see whether there is a significant difference between university levels concerning the results of the adapted TAME-ChatGPT survey, a one-way ANOVA test was utilized. The results of the one-way ANOVA test showcased that aside from two factors in the survey, there were potential statistically significant differences between university levels. To see the specific significances and their degrees, Bonferroni correction post-hoc tests were used.

Table 4.9

Bonferroni post-hoc results for university levels for the adapted TAME-ChatGPT survey score

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-15.467	4.175	.003
Freshman - Junior	-16.133	5.965	.053
Freshman - Senior	-18.162	4.098	.000
Sophomore - Junior	-.667	5.716	1.000
Sophomore - Senior	-2.696	3.727	1.000
Junior - Senior	-2.029	5.661	1.000

Firstly, the post-hoc test was utilized for the reason of finding the specific differences between the university levels of the participants regarding their overall survey scores. The results of the Bonferroni post-hoc test showcases that there is a statistically significant difference between freshman students ($M = 56.53$, $SD = 14.961$) and sophomore students ($M = 72.00$, $SD = 12.988$), as well as freshman students ($M = 56.53$, $SD = 14.961$) and senior students ($M = 74.70$, $SD = 10.585$). It can be seen that there is an upwards trajectory of positive ChatGPT attitude in students as their university years increase, although the difference is only significant between the freshman students and other years, excluding junior year students -although close, but not enough to be significant ($p = .053$)-, and not between the other three years (Table 4.9).

For the ‘heard’ section of the survey, the Bonferroni post-hoc test revealed a sole significant difference between the university years, which is between freshman students ($M = 34.76$, $SD = 8.301$) and senior students ($M = 41.26$, $SD = 7.533$). Depending on these results, it can be stated that the senior-level students are significantly have a more positive outlook compared to freshman-level students even if they have only heard of ChatGPT, and did not use it themselves (Table 4.10).

Table 4.10

Bonferroni post-hoc results for university levels for the ‘heard’ section scores

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-4.626	2.156	.210
Freshman - Junior	-3.741	3.581	1.000
Freshman - Senior	-6.502	2.229	.028
Sophomore - Junior	.885	3.616	1.000
Sophomore - Senior	-1.876	2.285	1.000
Junior - Senior	-2.761	3.660	1.000

To see the specific statistical significances and their degrees between the students’ university level and their ‘used’ section scores, Bonferroni post-hoc test was utilized. The results of the test revealed there was a statistically significant differences between freshman students ($M = 25.45$, $SD = 8.656$) and sophomore students ($M = 32.41$, $SD = 7.971$), freshman students ($M = 25.45$, $SD = 8.656$) and junior students ($M = 31.28$, $SD = 3.579$); and freshman students ($M = 25.45$, $SD = 8.656$) and senior students ($M = 33.03$, $SD = 5.319$). Similarly, a difference can be seen between freshman students and other students, with freshman students having a significantly less positive attitude towards ChatGPT compared to the other years (Table 4.11). It can be seen from the results of the post-hoc test for the ‘heard’ section and this section that this difference between the freshman students and other students are even larger when students have used ChatGPT, compared to only hearing about it.

Table 4.11*Bonferroni post-hoc results for university levels for the 'used' section scores*

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-6.959	1.999	.004
Freshman - Junior	-5.827	1.958	.022
Freshman - Senior	-7.577	1.958	.001
Sophomore - Junior	1.133	1.813	1.000
Sophomore - Senior	-.617	1.813	1.000
Junior - Senior	-1.750	1.768	1.000

When it comes to the particular factors in the study, the Bonferroni post-hoc test revealed that there were no statistically significant differences between students university levels and their perceived risk scores, or anxiety scores (Table 4.12, 4.13).

Table 4.12*Bonferroni post-hoc results for university levels for perceived risk scores*

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-2.383	1.279	.397
Freshman - Junior	-1.488	1.995	1.000
Freshman - Senior	-1.954	1.323	.862
Sophomore - Junior	.896	2.017	1.000
Sophomore - Senior	.430	1.356	1.000
Junior - Senior	-.466	2.045	1.000

Table 4.13*Bonferroni post-hoc results for university levels for anxiety scores*

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-.502	.931	1.000
Freshman - Junior	1.470	1.465	1.000
Freshman - Senior	1.135	.963	1.000
Sophomore - Junior	1.973	1.490	1.000
Sophomore - Senior	1.637	1.002	.636
Junior - Senior	-.335	1.511	1.000

The Bonferroni post-hoc test was utilized to see the particular differences between participants' university levels and their technology/social influences section scores. The results exhibit significant differences between freshman ($M = 15.35$, $SD = 3.929$) and senior students ($M = 21.17$, $SD = 2.640$), as well as sophomore ($M = 17.23$, $SD = 5.070$) and senior students ($M = 21.17$, $SD = 2.640$) (Table 4.14). From these results, it can be stated that senior students have significantly more positive attitude towards ChatGPT concerning technology and social influences. In addition, a static upwards trajectory of increasingly more positive attitude can be seen in university years of freshman ($M = 15.35$), sophomore ($M = 17.23$), junior ($M = 19.00$), and senior ($M = 21.17$).

For the perceived usefulness scores, the Bonferroni post-hoc test revealed that there was a significant difference between freshman students ($M = 14.22$, $SD = 5.954$) and senior students ($M = 19.28$, $SD = 3.787$). From this result, it can be stated that senior students have a more positive outlook towards ChatGPT regarding its perceived usefulness compared to freshman students (Table 4.15).

To find out the specific significant differences between students' behavior/cognitive factors scores and their university year, Bonferroni post-hoc test was used. The results of the test revealed that there was a statistically significant difference between freshman students ($M = 4.27$, $SD = 2.186$) and sophomore students ($M = 6.17$, $SD = 2.300$). From these results, it can be stated that sophomore students have a more positivity attitude towards ChatGPT, compared to freshman students (Table 4.16).

Table 4.14*Bonferroni post-hoc results for university levels for technology/social influences*

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-1.876	1.084	.524
Freshman - Junior	-3.645	1.818	.290
Freshman - Senior	-5.819	1.122	.000
Sophomore - Junior	-1.769	1.847	1.000
Sophomore - Senior	-3.943	1.167	.007
Junior - Senior	-2.174	1.869	1.000

Table 4.15*Bonferroni post-hoc results for university levels for perceived usefulness scores*

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-3.610	1.378	.060
Freshman - Junior	-3.470	1.349	.069
Freshman - Senior	-5.064	1.349	.002
Sophomore - Junior	.140	1.266	1.000
Sophomore - Senior	1.454	1.266	1.000
Junior - Senior	-1.594	1.234	1.000

Lastly, to find out the specific statistically significant differences between students' university levels and their ease of use scores, the Bonferroni post-hoc test was utilized again. The results reveal that there was a statistically significant difference between freshman ($M = 7.13$, $SD = 1.766$) and sophomore ($M = 8.41$, $SD = 1.991$), freshman ($M = 7.13$, $SD = 1.766$) and junior ($M = 8.44$, $SD = 1.523$), and freshman ($M = 7.13$, $SD = 1.766$) and senior level students ($M = 9.00$, $SD = 1.218$). From these results,

it can be stated that junior level students have a significantly more negative attitude towards the ease of use of ChatGPT compared to all other years (Table 4.17).

Table 4.16

Bonferroni post-hoc results for university levels for behavior/cognitive factors scores

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-1.900	.587	.010
Freshman - Junior	-.884	.575	.764
Freshman - Senior	-.477	.575	1.000
Sophomore - Junior	1.016	.532	.353
Sophomore - Senior	1.422	.532	.052
Junior - Senior	.406	.519	1.000

Overall, there have been a total of 14 significant differences between the university years found. From this total number, 6 of the differences were between freshman students and senior students, the years furthest apart from each other. The second number of differences were between freshman students and sophomore students with a total number of 4. The differences between freshman students and junior students, and sophomore students and senior students were both 2, and no difference between sophomore and junior, nor junior and senior was found. From these results, it can be stated that the mean scores most opposed to the rest were the scores of the freshman students.

The results of the freshman students can be due to many factors, like not taking any university courses, and only starting to take their introductory courses; their overall understanding of material and tool use in the language learning classroom as a teacher, inexperience with the field, age, and other similar factors. Because the second number of differences were seen between the freshman and sophomore years, it can be stated that the cutoff point between attitude differences does not seem to be between second and third year students, but between first and second year. However, there also have been significant differences observed when the students who have taken pedagogical compulsory courses concerning ELT and the ones who have not.

Table 4.17*Bonferroni post-hoc results for university levels for ease of use scores*

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-1.283	.455	.034
Freshman - Junior	-1.307	.446	.024
Freshman - Senior	-1.870	.446	.000
Sophomore - Junior	-.024	.418	1.000
Sophomore - Senior	-.586	.418	.981
Junior - Senior	-.562	.407	1.000

This section was utilized for answering the second and fifth research questions in this study. The results illustrate that there are differences for students' attitude towards ChatGPT regarding whether they have taken compulsory pedagogical courses or not, as well as their experience with ChatGPT, university year, and gender.

4.2.3. ChatGPT ELT Attitude Survey

The third and the last survey present in the questionnaire is the ELT Attitude survey, designed for the purposes of this study, and to answer the second and fifth research questions in this study alongside the adapted TAME-ChatGPT scale, but concerning entirely ELT-based attitude towards the large language model.

The test is made up of 4 different sections, which are teaching, evaluation, lesson design/material development, and attitude towards ChatGPT as a teacher, with a total number of items being 37. The same Likert scale as the adapted TAME-ChatGPT survey was utilized for this survey as well. The first section, teaching, includes 8 items. The second section, evaluation, includes 13 items. The third section, lesson design/material development, includes 5 items. Lastly, the fourth section, attitude towards ChatGPT as a teacher, includes 11 items.

To answer the research questions; finding out about the ChatGPT attitude score focusing on the different factors in ELT of the participants, and to see if there is a statistically significant difference between students' ChatGPT attitude score focusing on the aspects of ELT and their score in the different factors within the ChatGPT ELT survey

concerning differing independent variables such as gender, whether the students have taken compulsory ELT-based pedagogical courses or not, their university year, and whether they have only heard of ChatGPT or used it as well, this survey was utilized. For the purposes of answering these questions, various statistical analysis methods were used.

Firstly, to look at the overall findings of the survey and its different sections, the descriptive statistics of the results were calculated. The result reveals that in total, the average mean of the total scores of the ELT Attitude survey is 3.40, between the values of 1 to 5, being slightly higher than average. This highlights a positive attitude towards ChatGPT from the ELT perspective, although not by a large margin. From the averages, the highest mean score from the sections of the ELT attitude survey was from the lesson design/material development section, being 3.86. The lowest average was for the evaluation section, being 3.21. Furthermore, no average was below the average of 3 (Table 4.18).

Table 4.18

Descriptive statistics of the ELT attitude survey results

Variable	<i>n</i>	<i>Min.</i>	<i>Max</i>	<i>M</i>	<i>SE</i>	<i>SD</i>
1. ELT Attitude	114	37	185	126.49	3.009	32.132
2. Teaching	132	8	40	28.56	.642	7.377
3. Evaluation	120	13	65	41.77	1.166	12.771
4. Lesson Design/Material Development	132	5	25	19.30	.420	4.831
5. Attitude Towards ChatGPT as a Teacher	129	11	55	36.02	.986	11.197

To look at the results of the survey from different variables, different types of statistical analysis were done. Firstly, to see if there is a statistically significant difference between students' familiarity with ChatGPT and their scores, an independent samples t-test was incorporated to the study alongside Levene's Test for Equality of Variances. The results were then analyzed to see whether there is a statistically significant differences between students who have reported to only have heard of ChatGPT, and students who have reported to have heard of and also used ChatGPT.

From the results of the Levene's test for equality of variances, the equal variances were assumed for the total scores of the ELT attitude survey. This was also the case for the individual sections of the survey concerning the equality of variances.

With variances assumed to equal, an independent samples t-test was employed to see whether there were any differences between the students' ChatGPT experience -heard or used- and their ELT attitude survey scores. The results of the test revealed that there was no statistically significant difference between whether students have heard of ChatGPT or used it and ELT attitude survey score ($t(112) = -1.005$, $p = .317$). This result presents that students who have only heard of ChatGPT and students who have used ChatGPT have similar attitude towards ChatGPT in terms of ELT.

To see whether there is a significant difference between students' ChatGPT experience and scores of the teaching section of the survey, an independent samples t-test was used. The results reveal that there was no statistically significant difference between the two groups ($t(130) = -.397$, $p = .692$). Similarly, no significant results were found for the sections of lesson design/material development ($t(130) = .150$, $p = .881$), and attitude towards ChatGPT as a teacher ($t(127) = .082$, $p = .935$). However, when an independent samples t-test was used to see if there are any differences for the evaluation section of the survey, a statistically significant difference between students who have heard of ChatGPT and students who have used it was found ($t(118) = -2.153$, $p = .033$).

From these results, it can be stated that ChatGPT experience is not a factor in predicting students' ELT attitude towards ChatGPT, other than evaluation, which students who have only heard of ChatGPT had a significantly more positive outlook towards compared to students who have both heard and used the large language model.

Next, independent samples t-tests were used to see whether there were any differences between the ELT attitude score and scores for its sections regarding gender. The results of the Levene's Test for Equality of Variances all displayed a significance score higher than .05, which means that for every variable, equal variances are assumed.

The preceding results of the independent samples t-tests revealed that there were no differences between the total ELT attitude score and gender ($t(109) = 1.379$, $p = .171$), neither between evaluation score and gender ($t(115) = .909$, $p = .365$), lesson design/material development score and gender ($t(127) = 1.924$, $p = .057$), nor attitude towards ChatGPT as a teacher score and gender ($t(124) = 1.581$, $p = .116$). However, a

statistically significant different was found between ELT attitude survey teaching section score and the participants' gender ($t(127) = 2.346, p = .021$), with male students having a significantly more positive attitude towards teaching concerning ChatGPT. It also needs to be stated that from the valid responses to the teaching section, 45 participants were male and 84 were female, making the participant grouping unequal. Whether the significant difference can be found in a participant group equal or similar in number concerning male and female is a question for a future study.

Lastly, with Levene's Test for Equality of Variances showing that the equal variances can be assumed for all variables aside from the lesson design/material development section, independent samples t-tests were utilized to see whether there are any significant differences between whether students have taken compulsory ELT related pedagogical courses or not. The results of the independent samples t-tests reveal that there was no statistically significant difference between the courses taken and courses not taken group concerning total ELT attitude survey score ($t(112) = 1.165, p = .246$), teaching score ($t(130) = 1.858, p = .065$), evaluation score ($t(118) = .274, p = .785$), and attitude towards ChatGPT as a teacher score ($t(127) = 1.671, p = .097$). However, there was a statistically significant difference observed for the lesson design/material development section scores between students who have taken ELT-related compulsory pedagogical courses or not ($t(118.921) = 3.267, p = .001$).

To see whether there are any significant differences between the ELT attitude survey score and/or the scores for the sections of the survey and students' university years, a one-way ANOVA formula was used to statistically analyze the results of the survey answered by the participants. In addition, to see the individual differences between the university years, post-hoc tests were incorporated to the analysis process.

Firstly, the one-way ANOVA test, and additionally the Bonferroni post-hoc results were used to see whether there was any statistically significant difference between students' university levels and ELT attitude score. No statistically significant different was found (Table 4.19). In addition, no statistically significant difference was found between the students' university levels and evaluation score neither (Table 4.21).

Table 4.19*Bonferroni post-hoc results for university levels for ELT attitude score*

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-10.116	8.654	1.000
Freshman - Junior	-3.067	8.138	1.000
Freshman - Senior	-21.415	8.332	.069
Sophomore - Junior	7.049	8.395	1.000
Sophomore - Senior	-11.299	8.583	1.000
Junior - Senior	-18.348	8.063	.149

Table 4.20*Bonferroni post-hoc results for university levels for teaching score*

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-.296	1.741	1.000
Freshman - Junior	.294	1.714	1.000
Freshman - Senior	-5.483	1.741	.012
Sophomore - Junior	.590	1.741	1.000
Sophomore - Senior	-5.187	1.767	.024
Junior - Senior	-5.778	1.741	.007

The results of the post-hoc test reveal only one statistically significant difference for lesson design/material development score regarding the university level of the students, which is between freshman students ($M = 17.03$, $SD = 4.982$) and senior students ($M = 21.66$, $SD = 3.964$). From these results, it can be seen that senior students are significantly more positive towards ChatGPT in terms of lesson design and material development compared to freshman students (Table 4.22).

Table 4.21*Bonferroni post-hoc results for university levels for evaluation score*

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-2.560	3.317	1.000
Freshman - Junior	.945	3.292	1.000
Freshman - Senior	-4.947	3.286	.810
Sophomore - Junior	3.504	3.292	1.000
Sophomore - Senior	-2.387	3.370	1.000
Junior - Senior	-5.891	3.262	.441

To see the significant differences between university levels and teaching score, the Bonferroni post-hoc test was utilized. The results of the post-hoc test reveal there was statistically significant differences between freshman students ($M = 27.24$, $SD = 6.638$) and senior students ($M = 32.72$, $SD = 6.387$), sophomore students ($M = 27.53$, $SD = 8.195$) and senior students ($M = 32.72$, $SD = 6.387$), as well as junior students ($M = 26.94$, $SD = 6.945$) and senior students ($M = 32.72$, $SD = 6.387$). Overall, it can be seen that senior students have a significantly more positive outlook towards ChatGPT in terms of teaching English compared to any other university year (Table 4.20).

Table 4.22*Bonferroni post-hoc results for university levels for lesson design/material development score*

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-1.939	1.131	.532
Freshman - Junior	-2.618	1.113	.121
Freshman - Senior	-4.627	1.131	.000
Sophomore - Junior	-.678	1.131	1.000
Sophomore - Senior	-2.687	1.148	.124
Junior - Senior	-2.009	1.131	.467

The Bonferroni post-hoc test was utilized to see the specific differences concerning students' university level and their attitude towards ChatGPT as a future teacher. The results reveal that there is a statistically significant difference between freshman students ($M = 31.75$, $SD = 11.416$) and senior students ($M = 40.75$, $SD = 8.254$). From these results, it can be stated that senior students have a more positive attitude towards ChatGPT as a teacher or a future teacher compared to freshman students in a statistically significant degree (Table 4.23).

Table 4.23

Bonferroni post-hoc results for university levels for attitude towards ChatGPT as a teacher score

University Levels	Mean Difference	SE	Sig.
Freshman - Sophomore	-5.282	2.729	.331
Freshman - Junior	-2.926	2.667	1.000
Freshman - Senior	-9.000	2.707	.007
Sophomore - Junior	2.356	2.689	1.000
Sophomore - Senior	-3.718	2.729	1.000
Junior - Senior	6.074	2.667	.147

Overall, there have been 5 individually significant difference between two university years for the scores of the sections and the ELT attitude survey. Three of these results have been between freshman and senior students, one between sophomore and senior, and one between junior and senior. From these results, it can be seen that there are three statistically significant differences between 1st year and 4th year students from the 5 different post-hoc tests for the total ELT attitude survey score and individual scores for 4 of its functions.

This section was utilized for the purposes of answering the second and fifth research questions of this study alongside the second survey, the adapted TAME-ChatGPT survey. From the results, although they will be later discussed in more detail, it can be seen that there are differences in the attitude of ChatGPT concerning the different aspects of ELT in terms of gender, whether students have taken ELT-related pedagogical courses or not, university year, and whether students have used ChatGPT before or only

have heard of it. Overall, students have been found to have a higher than average positive outlook at ChatGPT concerning aspects of ELT and ELT overall.

4.2.4. Interviews

The following section includes the findings from the semi-structured group interviews conducted on ELT students studying to become English teachers in a state university in Western Anatolia, Türkiye. Three separate interviews were conducted, one involving 2 sophomore students; one involving 4 junior students; and one involving 2 senior students. No interview was done on freshman students due to the reasoning that the freshman students are yet to be in the know for specific ELT-related concepts, even introductory ones, since they have not taken any courses and were only starting their first term courses as the data was collected. The collected data was separated into themes and sub-themes. As mentioned previously, the codebook thematic approach steps were taken to code and present the qualitative data. The following data, while it can in certain cases, does not answer the research questions alone, but aimed to present more data towards answering them together with the findings of the quantitative data, alongside presenting background information and the attitude of the participants in a more detailed manner that is highly difficult to gather in a quantitative, survey-like setting.

The following sections, named as the main themes found in the qualitative data, will present the subthemes, numerical data around how many times the concepts were mentioned, from which student group the mentions came from, certain examples and additional anecdotes, alongside the comparison between the three interview sessions.

The students will be mentioned by numbers to make distinction easier. For this reason, the students will be numbered from Student 1 to Student 8, with the order starting from sophomore students and counting upwards to classification of Student 8 up to the senior students who were interviewed for the purposes of this study.

From Table 4.24, the specific information about the interview can be seen. Because of the reason that 4 junior students volunteered to be interviewed while 2 sophomore and senior students volunteered, the length and word count of the interview conducted with the junior student group were longer. Even though the same number of questions were asked to the students, because of the variety of answers and the overall length of the interview, it should be considered that the number of mentions regarding themes and sub-themes could have the potential to be higher in number for junior student

interviews compared to the other two interviews. However, looking at the numbers, this is not always the case, and there are a variety of sub-themes mentioned a similar number of times compared to other times, or lower in certain instances, but this fact should be kept in mind regardless.

Table 4.24

Information about the interviews' length, word count, and participants' number

University Year	Number of Students	Interview Length	Word Count
Sophomore	2	15:50	2290
Junior	4	22:14	3527
Senior	2	14:32	2154

4.2.4.1. Teaching language skills

The interview questions were, even though not solely, focused on the attitude of students regarding ChatGPT and ELT. In terms of learning and teaching English, even if the questions are not directly related, because of the students' background and current studies, they unsurprisingly answered the questions such as 'Overall, what do you think is the main advantage and disadvantage of ChatGPT?' from an angle of ELT, both as a current student, and as a prospective teacher; in some cases, a practicing teacher as well.

The theme of teaching language skills, also directly asked in the first 8 items of the ChatGPT ELT survey, was also coded with 5 different sub-themes in the qualitative data. These themes, a short description of the overall theme, and the number of themes the specific sub-themes were mentioned, and by which group it was mentioned can be seen in Table 4.25.

The most mentioned skill overall regarding ChatGPT use in teaching language skills was writing. Writing was only directly mentioned by the junior ($n = 4$) and senior ($n = 4$) students but not by the sophomore students. For example, student 5 expressed how the students can be directed by the teacher to check their own writing through the use of ChatGPT: "I would say yes to the student checking themselves, because I also input the essays that I write to ChatGPT". Student 7 expressed that "I do it (evaluate through

ChatGPT) for my own writing.” And that if they had ChatGPT for their own process in learning writing in English, it would enable to the student to receive more feedback than it was feasible for their English teacher to give them considering the number of students in classes and other constraints that teachers face.

Table 4.25

Coding table for the theme ‘Teaching Language Skills’ and its sub-themes

Theme	Sub-theme	Description	Number of times mentioned
Teaching Language Skills	Reading	How can ChatGPT be used to teach different language skills, are there more effective skills to be taught with ChatGPT, or can it help with any language skill? These sub-themes regard the students’ mentioning using ChatGPT for any section of lesson preparation and lesson conducting through the use/help of ChatGPT at any point of the process, with the caveat of being centered around teaching a certain language skill.	Sophomore: 0
			Junior: 4
			Senior: 2
	Listening		Sophomore: 0
			Junior: 2
			Senior: 1
	Writing		Sophomore: 0
			Junior: 4
			Senior: 4
	Speaking		Sophomore: 0
			Junior: 2
			Senior: 0
	Other Skills/Sub-Skills		Sophomore: 5
			Junior: 1
			Senior: 2
			Total: 27

The second most mentioned skill was reading. It was mostly mentioned by junior students ($n = 4$) followed by senior students ($n = 2$), while no mention was given by sophomore students. Student 3 expressed that “It is better to be used in that way (to prepare reading material using ChatGPT). I don’t know if I would use it directly in the classroom setting.”. In terms of senior students, the attitude towards it was similar. “...material can be outputted for reading. You can input the words you want to teacher, and it will give you stories that contain those words. It can be very helpful in this way.”

Listening was mentioned again by only junior (n = 2) and senior students (n = 1). Senior students talked about a recent instance where they used ChatGPT to create listening material using the text-to-audio function of ChatGPT. “We used it together last well. We needed an audio for micro teaching; so, we wrote what we wanted it (ChatGPT) to say, and it sent an audio recording.”

For the last main language skill, speaking was the least mentioned, only being mentioned by the junior students (n = 2). The students expressed that they would be open to using it for the evaluation of students’ speaking ability; however, they would not entirely trust the results, and use the output as a tool, rather than as a final say in students’ speaking ability. They expressed that they could also use it to come up with new items for activities around the skill.

For other language skills, there was a high number of mentions by sophomore students (n = 5), followed by mentions from senior students (n = 2) and junior students (n = 1). Sophomore students mainly talked about using ChatGPT to help with grammar-oriented lessons and the teaching process.

“When I teach a grammar topic, I feel that the skill is not explained in detail in the book (teacher’s/coursebook). For example, when teaching tenses, the time expressions used, and so on, there are no such examples (in the book). For that, I use ChatGPT to list them (examples).”

One mention by the junior student, Student 6, that they would “not use it for pronunciation”, explaining that technically, they believed that using ChatGPT to teach such a skill was not feasible.

Overall, the students looked positively, but cautiously at ChatGPT use for teaching language skills, and there were gaps in their knowledge about how the tool could be used for different skills. While they had ideas about using it to teach things such as writing and reading, they had not thought about other potential uses of it in teaching, for example, speaking, listening, pronunciation, intonation, and such.

4.2.4.2. Evaluation

The second main theme found from the interviews was ‘evaluation’. The section was separated into two sub-themes of grading and feedback, considering that while students’ overall ideas about evaluation stayed particularly same for every language skill,

their opinions differed in terms of whether and how much they would use ChatGPT for grading compared to feedback.

Table 4.26

Coding table for the theme 'Evaluation' and its sub-themes

Theme	Sub-theme	Description	Number of times mentioned
Evaluation	Grading	Should evaluation include ChatGPT? Is there are difference between giving feedback and grading the students' work with the use of AI?	Sophomore: 1
			Junior: 2
			Senior: 2
	Feedback		Sophomore: 1
			Junior: 2
			Senior: 1
			Total: 8

It can be seen in Table 4.26 that the amount of mentions in terms of evaluation are lower compared to teaching language skills, the subtheme of grading has been mentioned by all three groups, with junior ($n = 2$) and senior students ($n = 2$) mentioning it the most compared to sophomore students ($n = 1$). Sophomore students were open to evaluation through ChatGPT, although they highlighted the importance of giving ChatGPT the rubric and the context of the material they were inputting. Overall, all groups were more open to the idea of giving feedback, especially for homework and such, rather than grading. "If I make my own point system, I know the level of the students better. A technology coming between (teacher and student) even if it is ChatGPT, I feel like it separates me further away from the student. For this reason, I would like to prepare and evaluate exams myself." was expressed by Student 6.

Furthermore, while senior students were open to the idea of giving feedback to students as well as directing students to receive feedback from ChatGPT for things like their writings, they were not open to the idea of grading students' exam through ChatGPT. "In our education system, grades students receive affect a lot of things, that is why I don't think it would be fair."

Sophomore students were more inclined to trust ChatGPT, although not fully, when evaluating things such as assignments and in-class work that does not necessarily

affect the students' final grades in the lesson. Using it to grade, they would be even more careful to go over the results themselves as well. "Taken it directly would be wrong in my opinion. Even when you use ChatGPT, it says at the bottom that we should not trust the answers given by it a hundred percent." In terms of grading writing assignments, the sophomore students were distant to the idea, but agreed that it could be done. "I hadn't thought about it, but it could be used that way." However, the students also agreed that in terms of grading, the evaluation rubric, or a similar tool, should be given to ChatGPT as context for it to be used effectively, "I don't think it would be right to use artificial intelligence and tell it to give us an evaluation without giving it an evaluation metric/criterion"

While student 2 was open to the idea of using ChatGPT to grade assignments, they were still unsure of the final results that would be given by the AI. "I wouldn't take the results directly. I think taking them directly would be wrong, because when you use ChatGPT it says at the bottom (of the screen) that you shouldn't trust the answers given by ChatGPT a hundred percent" the student also adds "We can use it as a tool, but at the same time, we should use your own mind for evaluation as well. I don't think it would be enough just to trust it (ChatGPT)".

Overall, students were highly cautious towards using ChatGPT for evaluating the students, especially in terms of using ChatGPT for grading the students' work, meaning that they did not trust ChatGPT enough to give the responsibility of rating the students' work without checking the output in detail themselves; and even then, they did not see much use of it and would rather grade the students themselves. For feedback, especially pushing students towards receiving a form of self-feedback with their own ChatGPT use was seen as a positive aspect in the English learning process.

4.2.4.3. Lesson/material design

The third main theme found within the interviews is the 'lesson/material' designed theme, which is separated into four sub-themes as can be seen in Table 4.27. This theme collects the students' attitude and understanding of how ChatGPT can be used to design the lesson itself alongside the materials that will be used in different sections of the lesson, and other materials such as exams and homework.

Firstly, students had various outlooks towards lesson plan design using ChatGPT. The sophomore students' mentions of lesson plan design (n = 8) showcased that the

students were open to the idea of taking inspiration and ideas from how to design a lesson and what to include in it from ChatGPT rather than simply taking a set lesson plan outputted by ChatGPT as a whole. “I would take ideas. Rather than telling ChatGPT to create something, it feels sounder to take ideas from it, like working together with ChatGPT, instead of making it do everything” explaining that there is a high number of variables that go into creating a lesson plan “A lesson has objectives, goals. There are different aspects in goals.”

Junior students’ mention of lesson plan design ($n = 2$) reveal that junior students, especially Student 5, had used ChatGPT to construct a lesson plan, which was their first lesson plan.

“I had used it last year. My teacher had asked me to make a lesson, but we had never done it before. How does one construct a lesson plan? I had to make a lesson plan from scratch, but I didn’t know how. Then, I wrote to ChatGPT, asking about the different sections of a lesson plan and specified the age group, number of students, and so on, and it gave me the information one by one. At least, I understood and took inspiration from it and made my lesson plan that way.”

Student 3, in the same junior level student group expressed that “I also used it for my assignment. I had no idea about how to construct a lesson plan- I had to teach about wild animals to this age group, I asked ChatGPT about a lesson plan for wild animals, what would be the attainment goals for this goals, and so on.”

As a further note, while some of the students in the junior group was explaining how they used ChatGPT for their lesson plans, one of the students exclaimed that “Well, I only learned that you could use it (ChatGPT) this way in this moment. I was like ‘whoa, people are so smart’”. Because of that, I haven’t use ChatGPT this way (lesson planning) before. Would I use it in the future? Maybe, I could use it to get an idea, I wouldn’t just copy-paste everything it says.” The statement points towards the attitude of the student towards ChatGPT; however, it also points out the ChatGPT literacy of the student, especially in regards to what ChatGPT can be used and in which ways, and how the student was not knowledgeable enough towards the capabilities of the large language model to think of using it to construct an English lesson plan before; and when they learned that it was possible, they were intrigued about using it that way, even though they

were still wary of the results. This data has a possibility of pointing towards the idea that as ChatGPT literacy increases, so does positive ChatGPT attitude.

Table 4.27

Coding table for the theme 'Lesson/Material Design' and its sub-themes

Theme	Sub-theme	Description	Number of times mentioned
Lesson/Material Design	Lesson Plan Design	What are the benefits of lesson/material design using ChatGPT? What are the considerations?	Sophomore: 2
			Junior: 2
			Senior: 4
	Material Design		Sophomore: 3
			Junior: 1
			Senior: 4
	Exam Design		Sophomore: 1
			Junior: 1
			Senior: 1
	Homework/Assignment Design		Sophomore: 1
			Junior: 0
			Senior: 3
			Total: 23

In terms of material design, senior students had the most mentions ($n = 4$), closely behind was sophomore students ($n = 3$), with junior students having the least amount of mentions of the topic ($n = 1$). Overall, the students looked at material design using ChatGPT positively. Senior students discussed how they use ChatGPT for activity ideas for certain topics, contexts, lessons, and so on. Student 7 explained “the micro-teaching we mentioned was around task-based teaching, and because some things developed spontaneously in the class, we asked it (ChatGPT) to develop a few back-up activities. We chose the ones that could be useful to us during the lesson.” The sophomore students had less experience with material design, but still had a positive attitude towards using ChatGPT to gather activity ideas and answer questions about material design, but still looked at using activities and materials solely created by ChatGPT to be unethical.

Senior students were looking to find a text with a specific property that they deemed to be difficult to create themselves; so, they asked ChatGPT “I needed to write something rhythmic for the children, for the lesson plan. I couldn’t find it on the internet. So, I asked it (ChatGPT) to write that for this level. It wrote it immediately.” After the anecdote given by student 7, student 8 also added that they also used it to write a short story for their lesson plan material.

The sophomore students expressed that their outlook towards material design was the same as their outlook towards their perceived ethical use of ChatGPT, which was explained by student 2 as “I would take ideas, more like. In my opinion, it feels healthier to take ideas from it rather than ask to it create something. Here, it’s more like working together with ChatGPT, instead of making it do everything.” The student mentions cooperating with ChatGPT, rather than it taking center stage, using it as an idea machine, rather than something that one should use it make any type of material from scratch.

In terms of exam design, all groups mentioned exam design by ChatGPT once. Overall, sophomore students expressed that they could use ChatGPT to prepare items and questions for exams, but not for designing the entire exam with ChatGPT. “I would use it. Not as a “come on, prepare me an exam” way but in what kind of questions could ChatGPT come up with. I would look at those questions, change or add things as necessary and use them.”

All junior students expressed adamantly that they would not use ChatGPT in any way to construct examinations, with Student 5 expressing “I would only use it if I was in a state of “I cannot do it, I can’t find anything”. It (ChatGPT) would not be the first thing I would go to when preparing exams.” Senior students expressed that ChatGPT has a tendency to create wrong question structures, give multiple correct answers in multiple choice questions, and overall be inconsistent.

Lastly, majority of mentions for homework design using ChatGPT came from senior students ($n = 3$), followed by sophomore students ($n = 0$), while no direct mentions were made by junior students. Sophomore students expressed that they would use it for homework/assignment design while expressing more negative attitude towards exam design using ChatGPT. When asked about it, student 7 answered “Because it is more more-ended, I can choose between them (outputs by ChatGPT), but with tests, I can miss certain things (wrong ChatGPT outputs/evaluations).” Sophomore students looked more

positively to the use of ChatGPT for this context. As a note, when they were asked about homework design using ChatGPT, their focus also shifted more towards expressing how students can use ChatGPT to plagiarize or do their homework in a way that goes against the main objectives of the homework.

Overall, the students' attitude towards using ChatGPT for lesson/material design is positive; however, they also stress certain things to look out for. For example, the use of ChatGPT to create entire activities and other similar things was seen as a negative by students, and they expressed that the proper use of ChatGPT to fit in line with the tenet of teaching ethically and with responsibility was that ChatGPT could give individual items, ideas, outlines, and so on, but the teacher always should have a hand in the design process and should not leave the entire process to the large language model.

4.2.4.4. Professional development

The fourth theme taken away from the results of the interviews is the theme of 'professional development'. This section includes students use of ChatGPT as a university student themselves to learn about English education now and in the future, and also using it as a teacher to keep up with current theory in ELT.

The results from table 4.28 display the mention number of the sub-themes. The results reveal that for use as a university student sub-theme, junior students were the most to discuss it ($n = 5$) with sophomore students mentioning it only once, and no mentions from senior students. Junior students have expressed their use of ChatGPT for their courses, homework, to prepare for exams, and so on.

"Online lessons were centered around homework. Sometimes we needed to research articles and I would use it to summarize 3,4 articles, for example. Although I do not trust it a hundred percent, because I caught mistakes before."

Sophomore students also expressed that they have used ChatGPT before for their own homework and assignments. They expressed that rather than using ChatGPT to create material for them to be able to complete their assignments and homework, they rather use it as a tool that would steer them towards the right correction, give them ideas, and help them achieve the objectives of the assignment themselves. They also expressed that they would want the same thing for their future students as well "I wouldn't want it to be like that. It doesn't feel like they are doing their homework by actually thinking that

way. There is a homework, go ask ChatGPT, copy-paste the answer, and let the teacher grade it. I think I would prefer students to use their heads and create something that way, like I said.”

Table 4.28

Coding table for the theme ‘Professional Development’ and its sub-themes

Theme	Sub-theme	Description	Number of times mentioned
Professional Development	Use as a University Student	Can ChatGPT be used as a tool for professional development as an ELT student and/or as a teacher?	Sophomore: 1 Junior: 5 Senior: 0
	Keeping up with Current Literature		Sophomore: 1 Junior: 1 Senior: 1
			Total: 9

For the act of keeping up with current literature, all student groups mentioned the subject once, and agreed on its usefulness overall as a teacher; however, one of the senior students also expressed that “In terms of literature, it (ChatGPT) could be lacking, and could give incorrect information.” Junior students had a more positive attitude towards using ChatGPT to find relevant studies to use in practice as a teacher while still not being a hundred percent trustful of the results “Rather than teaching how to be a teacher, it could give things like different sources and information written around the web.”

Overall, students showed an interest in expressing the usefulness of ChatGPT in their own university process, but did not give much previous thought into how they would keep up with the ever-changing and improving literature of language education after they graduate and do not have active courses teaching them relevant ELT theory. Even so, when asked, they expressed a positive attitude towards ChatGPT’s usefulness in this aspect while also still being concerned with the correctness of data given by the large language model.

4.2.4.5. Ethical considerations

The fifth main theme taken from the group interviews is ‘ethical considerations’ which is separated into 4 relevant sections. This section is around the ethical considerations how ChatGPT used by teacher, students, and overall, the ethical concepts surrounding the development and existence of ChatGPT, large language models, and AI overall. The number of mentions by different interviewed groups can be seen in Table 4.29, alongside the list of sub-themes for the main theme.

In terms of academical use, the most amount of mentions of the sub-theme came from junior students ($n = 5$). The subtheme was mentioned an equal number of times by sophomore ($n = 2$) and senior students ($n = 2$). Student 6 mentions that ChatGPT has been criticized for the potential of it getting information from studies and articles not available to the public. “Sometimes ChatGPT can steal people’s written things not available for public access and showcase it as its own original material. That is the one thing I see as unethical about ChatGPT”. Students also express that people working in academia can create academical work with the use of ChatGPT quickly in an unethical manner. This was expressed with the sophomore students as well. “I heard about some lecturers writing articles through ChatGPT and publishing them to increase their publication numbers.”

Concerning plagiarism, the number of mentions going from highest to lowest is junior ($n = 3$), sophomore ($n = 2$), and senior students ($n = 1$). Student 5 mentioned an instance where the students in a course she was taking was caught plagiarizing their written assignments using AI. “... because plagiarism is too widespread. You give ChatGPT instruction and tell it to write a paragraph about this topic in a B1 (CEFR) level, and it does. For example, some people had done that at the time the courses were done online and the teacher had caught it.” Additionally, junior students were also knowledgeable about potential plagiarism in academic writing, as with the quote also given a paragraph above “I heard about some lecturers writing articles through ChatGPT and publishing them to increase their publication numbers.”. As an additional question, the researcher asked the students that explained that they use ChatGPT to help with their homework, studying, and so on, on whether it feels like they are cheating in some sense. Junior students have explained that it is a tool to be used, and it is not cheating to use it. Senior students, on the other hand, expressed a feeling of “cutting corners” or “taking the easy way out” even if they were not plagiarizing their work, but only using ChatGPT as

a tool to get ideas and similar things from “Yes, it happens. I feel like I am taking the easy way out. We could have me the listening audio ourselves, yes, but we took it from there (ChatGPT).”

Table 4.29

Coding table for the theme ‘Ethical Considerations’ and its sub-themes

Theme	Sub-theme	Description	Number of times mentioned
Ethical Considerations	Academical Use	What are the ethical boundaries that should be considered with the use and development of AI? In what ways can AI be used ethically, or unethically?	Sophomore: 2
			Junior: 5
			Senior: 2
	Plagiarism		Sophomore: 2
			Junior: 3
			Senior: 1
	Cheating		Sophomore: 2
			Junior: 2
			Senior: 1
	Ethical Considerations of Machine Learning		Sophomore: 4
			Junior: 1
			Senior: 0
			Total: 25

For the sub-theme of cheating, sophomore students ($n = 2$) and junior students ($n = 2$) expressed the idea the same number of times, with senior students expressing it only once. Senior students, in their only utterance of cheating as an ethical consideration of ChatGPT when they were asked about what generally could be the ethical considerations of ChatGPT, and Student 8’s first answer was that it could be students using it to cheat at exams. “For example, it can be cheating maybe, in exams and homework, they can copy and paste the answers. That wouldn’t be ethical.” The other groups were also in the know about the capabilities of ChatGPT’s ability to be a tool for cheating at exams, homework, and so on, alongside the considerations of plagiarism for both students and academics alike.

When asked about the ethical considerations of ChatGPT, there was a single utterance by junior students about the ethical boundaries and questions about how ChatGPT was developed. As discussed, ChatGPT is a large language model, developed on data from the internet. This data can be websites, studies, articles, social media sites, and so on. While the general process of the AI is understood, the exact data it is trained on, and the ethical qualms about using such data without explicit permissions from the people who created that data is an ethical consideration well discussed in the literature of artificial intelligence, big data, machine learning, and such. While one utterance was captured from junior students and none from senior students, sophomore students discussed the topic four times during the interview.

“For example, it would be wrong to ask ChatGPT to write you a homework, an article because when ChatGPT writes an article, it takes from articles written by people because there is so many sources on the internet. It takes that information from the internet, combines them.”

In summary, the students were all around knowledgeable around the ethical qualms ChatGPT could bring to the language learning classroom, and more. Overall, the students were even more aware of such ethical considerations regarding being a student, cheating at exams, plagiarism and so on, but have also given thought to other aspects like plagiarism in academia, machine learning algorithms, and so on that they did not directly had experience in, but were knowledgeable in regardless.

4.2.4.6. Outlook as a teacher

The sixth main theme found from the interviews was titled ‘outlook as a teacher’. This theme surrounds the students’ attitudes and beliefs about ChatGPT regarding their future as an English teacher. Additionally, the senior students have already went to their internships as English teachers for a few weeks before the interview took place. While so, because the data was collected intentionally at the start of the fall term, and for that reason, the students did not have extensive teaching experience in their internships but only were given introductory experiences. In addition, one of the sophomore students also taught English at a private school part-time, and therefore had first-hand experience with the language learning classroom, and ChatGPT use in and around such classroom.

Table 4.30*Coding table for the theme 'Outlook as a Teacher' and its sub-themes*

Theme	Sub-theme	Description	Number of times mentioned
Outlook as a Teacher	Time-Saving	What are the benefits of AI in the real life classroom, should students be allowed to use ChatGPT? In what ways?	Sophomore: 1
			Junior: 5
			Senior: 5
	Creativity		Sophomore: 3
			Junior: 5
			Senior: 1
	Permission to use ChatGPT		Sophomore: 1
			Junior: 3
			Senior: 1
			Total: 20

From Table 4.30, it can be seen that for the sub-theme of time saving as a teacher, both junior ($n = 5$) and senior students ($n = 5$) had a high number of mentions around its topic, with a single mention from sophomore students. The students expressed the lack of resources teachers may face in the field, and ChatGPT may be helpful in saving time in certain aspects for the teacher's overall reach can be better. Senior students were especially vocal about this positive aspect of ChatGPT.

"It can work for planning. For example, when you ask "what should I do in the lessons today?" it lists things you can do. For a teacher, I think that can be helpful. In Türkiye, we can teach classes that are highly crowded. It can take so much time to sit down and plan every one of them. The curriculum is also very busy. That's why it can save time in Türkiye."

When asked about one positive and negative aspect of ChatGPT, sophomore students said that the positive was its potential to save time.

For creativity, junior students had the greatest number of mentions ($n = 5$), followed by sophomore ($n = 3$) and senior students ($n = 1$). While there have been mentions of ChatGPT helping creativity by giving new ideas, most mentions were around the consideration that using ChatGPT might lower the creativity of students, teachers, and

so on. Student 2 expressed that “One disadvantage of it is it lowers the creativity of people.” When asked about the disadvantages of ChatGPT, Student 7 expressed the same opinion.

For the permission of the use of ChatGPT for students, the participants had the most mentions in junior students ($n = 3$), followed by one mention both by sophomore and senior students. Overall, students expressed that they would not try to restrict the use of ChatGPT from their students completely, but would rather leave the door open for the benefits of its use while keeping aware of the potential disadvantages. One junior student expressed:

“Like my friends said, even if I ban it, they will still use it. We are not standing over them at their homes. It’s the opposite, they will be curious (if I ban it). prohibitions are always attractive. But I also don’t want to give me a perfect routine when I ask them to write about their day is. I want them to make mistakes, because you learn by making mistakes.”

Overall, students have a positive attitude towards ChatGPT as teacher, both as a prospective teacher and also for their students to use as well; however, they were also aware of the potential pitfalls its use can bring, and while aware of what to do to avoid some of those pitfalls, they were not sure that they could avoid ChatGPT use garnering its disadvantages completely, and therefore were wary of it to a certain degree.

4.2.4.7. Future of AI

The last theme found within the interviews was titled ‘future of AI’. This section looks at the mentions of AI overall, which is an important aspect when discussing AI and large language models, since their development is highly recent and therefore in addition to the discussion around it, the technology behind it is improving at a rapid pace.

From Table 4.31, it can be seen that overall, junior students had the most number of mentions of the topic. Junior students mentioned ‘progression of AI’ 5 times while other groups mentioned it once, and mentioned ‘ChatGPT improvements’ 4 times while other groups once again mentioned the subtheme only once. Overall, a general discussion around how the AI is ever-evolving was had with the students which affected the number of mentions. Overall, junior students mentioned that they were not completely aware of the recent developments and version of ChatGPT, and believed that the technology

overall will improve more as time goes on. Student 3 mentioned that, in regards of ChatGPT and AI tools overall “We are 3rd year, so I don’t know how different it will be when we graduate. Everything is improving much, much faster now.”

Table 4.31

Coding table for the theme ‘Future of AI’ and its sub-themes

Theme	Sub-theme	Description	Number of times mentioned
Future of AI	Progression of AI	What is the future of AI? Will ChatGPT only get better, what should be the outlook towards ChatGPT and other similar AI models’ future?	Sophomore: 1
	ChatGPT Improvements		Junior: 5 Senior: 1
			Sophomore: 1 Junior: 4 Senior: 1
			Total: 13

Junior students were especially interested with the future of AI and how AI should be integrated into the language classroom going forward. Even though they were adamant that AI played a big role in the future, that idea also seemed to be a double-edged sword in their opinion, meaning that the future of AI could prove to be a detriment. Student 6 expressed that:

“I wouldn’t want ChatGPT to be integrated into the education system too much. Whatever happens, ChatGPT is an application that makes writing and homework easier. Even if it is correct or incorrect, it is a fast program, and that speed and ease makes students run away from research techniques and ability, their ability to complete assignment, creating parts, making something. If it is always integrated into education too much, these students, after they graduate, they are not going to have the skills they need in their careers. Because they are attached to it too much, they are going to create problems for themselves, and young students are not aware of that. Therefore, everyone wants to relax, and spend more free time. For that reason, for students’ and people’s future, I wouldn’t want AI to be integrated too much. Aside from that, as I’ve said, there

are useful sides of it as a source that can help you. There are advantages to it because it (ChatGPT) going parallel to the education system does make me anxious.”

In summary, students had a positive attitude towards ChatGPT concerning its future. From the different sides of belief that AI is a ‘fad’, or AI is the future of education, students fall somewhere in the middle. They do not believe that AI will completely go away and it will be integrated into language education, but it will not drive away the human element of it, and the teacher will be an integral part of every process in and outside the classroom, and AI will be used as a tool, like many other tools and resources at the disposal of the teacher.

4.2.5. Comparison

While a number of significant differences and correlations were discovered concerning the participant groups for questionnaire and interview results on their own, there is also the need to compare the results of the different types and categories of data collected for the purposes of this thesis to find out whether there are any significant correlations between students’ ChatGPT literacy, attitude, and ELT attitude.

Firstly, The Spearman rank-order correlation coefficient was utilized to see if there were any statistically significant correlations between the means of the survey results for the adapted ChatGPT literacy scale results and adapted TAME-ChatGPT. The correlation strength was calculated using Figure 4.2. The results reveal that there is a weak but significant relationship between ChatGPT literacy and attitude in total survey scores ($r(61) = .010$, $p = .326$). When separated into the categories of ‘heard’ and ‘used’, while a moderate correlation was found between ChatGPT literacy and scores for the ‘used’ section of TAME-ChatGPT ($r(109) = .000$, $p = .482$), no significant correlation was found between ChatGPT literacy and scores for the ‘heard’ section ($r(69) = .255$, $p = .139$).

Furthermore, the total score of the adapted ChatGPT literacy survey was also utilized with the use of Spearman correlation to see whether there was any significant correlation between different scores of sections from TAME-ChatGPT and ChatGPT literacy survey total score. The results reveal the following correlations: weak correlation with technology/social factors ($r(71) = .012$, $p = .297$), moderate correlation with perceived usefulness ($r(109) = .000$, $p = .481$), weak correlation with behavior/cognitive

factors ($r(109) = .001$, $p = .311$), and weak correlation with ease of use ($r(109) = .001$, $p = .324$).

Next, Spearman correlation was used for the different sections of the adapted ChatGPT literacy scale and adapted TAME-ChatGPT to see whether there was any significant correlation between the factors. First factor used was the technical proficiency section in the ChatGPT literacy survey. The results reveal significant correlation between technical proficiency and the following factors: weak correlation with attitude total scores ($r(65) = .012$, $p = .309$), moderate correlation with attitude used section ($r(115) = .000$, $p = .481$), weak correlation with technology/social influences ($r(76) = .025$, $p = .256$), moderate correlation with perceived usefulness ($r(116) = .000$, $p = .451$), weak correlation with behavior/cognitive factors ($r(115) = .001$, $p = .419$), weak correlation with perceived ease of use ($r(116) = .001$, $p = .311$).

Figure 4.2

Correlation coefficient strength of Spearman coefficient

Perfect	+1	-1
Strong	+0.9	-0.9
	+0.8	-0.8
	+0.7	-0.7
Moderate	+0.6	-0.6
	+0.5	-0.5
	+0.4	-0.4
Weak	+0.3	-0.3
	+0.2	-0.2
	+0.1	-0.1
Zero	0	

Source: Dancey & Reidy (2007)

For the critical evaluation section, the results of the Spearman correlation showcase weak significant negative correlation with perceived risk ($r(74) = .012$, $p = -.292$).

Concerning communication proficiency, the results of the Spearman correlation revealed the following significant correlations: a weak correlation with total attitude survey score ($r(63) = .005, p = .350$), a moderate correlation with attitude ‘used score’ ($r(112) = .000, p = .477$), a weak correlation with technology/social influence ($r(74) = .020, p = .270$), a moderate correlation with perceived usefulness ($r(113) = .000, p = .481$), a weak correlation with behavior/cognitive factors ($r(112) = .002, p = .285$), and a weak correlation with perceived ease of use ($r(113) = .000, p = .333$).

For the creative application section, the results of the Spearman correlation revealed the following significant correlations: moderate correlation with total attitude survey score ($r(63) = .000, p = .530$), weak correlation with the ‘heard’ section ($r(72) = .006, p = .324$), moderate correlation with the ‘used’ section ($r(113) = .000, p = .507$), moderate correlation with technology/social influences ($r(74) = .000, p = .423$), moderate correlation with perceived usefulness ($r(114) = .000, p = .523$), weak correlation with behavior/cognitive factors ($r(113) = .000, p = .362$), and weak correlation with perceived ease of use ($r(114) = .010, p = .242$).

Concerning the ethical competence section, the results of the Spearman correlation showcased the following significant correlations: weak negative correlation with perceived risk ($r(75) = .037, p = -.241$), and weak correlation with behavior/cognitive factors ($r(115) = .030, p = .202$).

Overall, it can be seen that there is overall significant correlation between the results of the adapted ChatGPT literacy survey and adapted TAME-ChatGPT. Furthermore, it can be stated that these correlations are overall weak to moderate. Lastly, the total scores of both surveys have a weak but significant correlation, answering the third research question ‘Is there a significant correlation between ELT students’ ChatGPT literacy and ChatGPT attitude?’

The same steps were followed to further answer the third research questions from an angle of ELT attitude towards ChatGPT and its hypothesized correlation with ChatGPT literacy. Firstly, the total score of the literacy survey was utilized using Spearman correlation to see if there were any significant correlations with the different facets of the ELT attitude survey. The results reveal a moderately strong significant correlation between total literacy score and total ELT attitude score ($r(104) = .000, p = .483$), as well as weak correlation with the teaching section ($r(119) = .000, p = .355$),

weak correlation with evaluation section ($r(110) = .000, p = .385$), moderate correlation with lesson design/material development section ($r(110) = .000, p = .436$), and moderate correlation with attitude towards ChatGPT as a teacher section ($r(117) = .000, p = .496$). The results showcase that category of the ELT attitude survey correlates with the total score for the literacy survey.

Concerning technical proficiency, Spearman correlation was utilized to see if there was any significant correlation concerning the different categories of ELT attitude survey. The results showcase the following significant correlations: moderate correlation with total ELT attitude score ($r(109) = .000, p = .464$), weak correlation with teaching ($r(126) = .001, p = .296$), weak correlation with evaluation ($r(115) = .000, p = .355$), weak correlation with lesson design/material development ($r(126) = .000, p = .393$), and moderate correlation with attitude towards ChatGPT as a teacher ($r(124) = .000, p = .467$).

For critical evaluation, Spearman correlation was utilized to find out the following significant correlations: weak correlation with total ELT attitude score ($r(108) = .005, p = .270$), weak correlation with evaluation ($r(114) = .048, p = .186$), weak correlation with lesson design/material development ($r(124) = .023, p = .203$), weak correlation with attitude towards ChatGPT as a teacher ($r(122) = .002, p = .275$).

For communication proficiency, Spearman correlation was utilized to find out the following significant correlations: weak correlation with total ELT attitude survey score ($r(107) = .000, p = .380$), weak correlation with teaching ($r(123) = .000, p = .318$), weak correlation with evaluation ($r(113) = .001, p = .300$), moderate correlation with lesson design/material development ($r(123) = .000, p = .402$), weak correlation with attitude towards ChatGPT as a teacher ($r(121) = .000, p = .381$).

Concerning creative application, Spearman correlation was utilized to find out the following significant correlations: moderate correlation with total ELT attitude survey score ($r(107) = .000, p = .489$), weak correlation with teaching ($r(124) = .000, p = .343$), weak correlation with evaluation ($r(113) = .000, p = .381$), moderate correlation with lesson design/material development ($r(124) = .000, p = .416$), moderate correlation with attitude towards ChatGPT as a teacher ($r(122) = .000, p = .538$).

For ethical competence, Spearman correlation was utilized to find out the significant correlations between ethical competence and the scores of the sections of the ELT attitude survey; however, no significant correlation was found.

From the results, it can be seen that concerning the total scores of both surveys, the correlation between ChatGPT literacy and ELT attitude is higher for the participants than between ChatGPT literacy and attitude.

Lastly, Spearman correlation was utilized to see the different correlations regarding TAME-ChatGPT attitude scores and ELT attitude scores. Starting from the total score of the ELT attitude survey, a significant moderate correlation between ELT attitude survey score and attitude survey score can be seen ($r(56) = .000$, $p = .629$). Additionally, the following correlations concerning total ELT attitude survey score was found out: weak correlation with 'heard' section ($r(68) = .003$, $p = .356$), moderate correlation with 'used' section ($r(97) = .000$, $p = .619$), weak correlation with perceived risk ($r(69) = .026$, $p = .268$), moderate correlation with technology/social influence ($r(69) = .000$, $p = .414$), moderate correlation with perceived usefulness ($r(97) = .000$, $p = .600$), moderate correlation with behavior/cognitive factors ($r(97) = .000$, $p = .499$), and weak correlation with ease of use ($r(97) = .004$, $p = .290$).

Furthermore, Spearman correlation was utilized to see whether there were any significant correlations between the scores of the different sections of the ELT attitude survey and different scores of TAME-ChatGPT. Firstly, ELT attitude teaching score was used for the correlation which revealed the following significant correlations: moderate correlation with total attitude score ($r(65) = .000$, $p = .425$), weak correlation with 'heard' section ($r(80) = .003$, $p = .324$), moderate correlation with 'used' section ($r(112) = .000$, $p = .425$), moderate correlation with technology/social influences ($r(81) = .000$, $p = .451$), moderate correlation with perceived usefulness ($r(113) = .000$, $p = .447$), weak correlation with behavior/cognitive factors ($r(112) = .006$, $p = .261$), weak correlation with ease of use ($r(113) = .009$, $p = .243$).

Concerning ELT evaluation, Spearman correlation was utilized to find out the following significant correlations: moderate correlation with total attitude score ($r(58) = .001$, $p = .439$), weak correlation with 'heard' section ($r(70) = .038$, $p = .248$), moderate correlation with 'used' section ($r(103) = .000$, $p = .497$), weak correlation with technology/social influences ($r(72) = .029$, $p = .257$), moderate correlation with perceived usefulness ($r(103) = .000$, $p = .446$), moderate correlation with behavior/cognitive factors ($r(103) = .000$, $p = .457$), weak correlation with perceived ease of use ($r(103) = .026$, $p = .219$).

For lesson design/material development, Spearman correlation was utilized to find out the following significant correlations: moderate correlation with total attitude score ($r(64) = .000, p = .586$), weak correlation with ‘heard’ section ($r(79) = .002, p = .337$), moderate correlation with ‘used’ section ($r(112) = .000, p = .582$), moderate correlation with technology/social influence ($r(81) = .000, p = .508$), moderate correlation with perceived usefulness ($r(113) = .000, p = .548$), moderate correlation with behavior/cognitive factors ($r(112) = .000, p = .402$), weak correlation with perceived ease of use ($r(113) = .000, p = .375$).

Lastly, for attitude towards ChatGPT as a teacher, Spearman correlation was utilized to find out the following significant correlations: a strong correlation with total attitude survey score ($r(63) = .000, p = .724$), moderate correlation with ‘heard’ section ($r(77) = .000, p = .467$), moderate correlation with ‘used’ section ($r(110) = .000, p = .681$), weak correlation with perceived risk ($r(78) = .007, p = .303$), moderate correlation with technology/social influence ($r(79) = .000, p = .569$), moderate correlation with perceived usefulness ($r(111) = .000, p = .664$), moderate correlation with behavior/cognitive factors ($r(110) = .000, p = .500$), weak correlation with perceived ease of use ($r(111) = .000, p = .328$).

Overall, the strongest correlation for ChatGPT literacy was attitude towards ChatGPT as a teacher. The strongest correlation for ChatGPT attitude was attitude towards ChatGPT as a teacher. Lastly, the strongest correlation for ChatGPT ELT attitude was ChatGPT attitude, in the form of the total score for the adapted TAME-ChatGPT. This also coincides with the results of the interviews, where the students’ attitude was observed to be stronger in its answers with less room for uncertainty and ‘middle-of-the-road’ approaches when the subject was around attitude concerning ELT exclusively compared to general attitude towards ChatGPT. However, this difference was not observed to be too high, since students saw many similarities in their attitude towards ChatGPT in general and concerning ELT.

Furthermore, having more experience with ChatGPT, rather than only hearing about it has been shown to have a significant effect on ChatGPT attitude, where the general attitude and ELT attitude of the participants were found out to be higher if they have interacted with ChatGPT before. While no students who have not interacted with ChatGPT was interviewed, the interviewed students seemed to also have a nuanced

approach to their attitude towards ChatGPT if they were more familiar with using ChatGPT, and its potential overall, with students who have given anecdotes of using ChatGPT also expressing concerns over less considered considerations concerning ChatGPT. For example, a senior student has expressed their use of ChatGPT for academic purposes, and also went on to discuss the ethical considerations around how ChatGPT was improved using machine learning techniques by acquiring input from the digital web without permission of the creators of various material which could be and was seen as unethical in AI development, even recognizing that these data fed to the AI could have been material with monetary gates and restricted access in general and public use.

In terms of anxiety, while students have expressed concern over ChatGPT and its current and future use in the interviews, discussing the idea that ChatGPT has entered into the education as a construct, and in the future, teachers and students will have to learn to live with it, since all the interviewed students saw ChatGPT as an inevitability and not something that could be ignored. However, even though students were aware of this, they were not anxious about the fact's prospects, which could also be seen in the low scores of the anxiety section of the adapted TAME-ChatGPT survey.

In overall ChatGPT use, students have expressed ease and practicality in ChatGPT, which was also reflected in the scores of sections such as perceived usefulness and ease of use both having a high mean score, especially with ease of use having a mean score of 8,33 between the possible minimum of 2 and a maximum of 10.

In terms of ELT attitude, students have scores higher overall than TAME-ChatGPT survey results, with many students seeing the benefits that ChatGPT could potentially bring to the education process of foreign language education, although they were hesitant in its use in aspects of education such as evaluation, which is not necessarily echoed with the score of the evaluation section of the ELT attitude survey being higher than average. Higher scores in different sections of the ELT attitude survey can also be seen, with teaching being higher on average than evaluation, which was also seen in interviews where students were more willing to use ChatGPT to create material for the classroom and to use it to teach language skills compared to using it for creating exams, for evaluation and rating of said exams, whether they be writing assignments, or other types of examinations or homework. Overall, a positive ELT attitude was observed for both the interviews and the results of the questionnaire.

4.3. A Brief Summary of the Results

To briefly summarize the results, the ChatGPT literacy of the participants were found to be above average in their mean scores. The total score of the adapted ChatGPT literacy scale was found to not be affected by gender or whether students have taken ELT-based pedagogical courses or not. In terms of university years, statistically significant difference between various years were found, with the most significant being between freshman and senior students. Afterwards, the same statistical methods were utilized to see the various statistically significant differences between the scores of the sub-sections of the survey and other various variables, with significances found in many facets.

For TAME-ChatGPT, the results reveal a near-average mean score of ChatGPT attitude concerning the results of the adapted TAME-ChatGPT survey. The total score was found to be statistically significant in its difference concerning gender. This difference was also existent in the ‘heard’ section of the survey, but not on the ‘used’ section of the survey. Furthermore, statistically significant differences were found between participants who have taken ELT-based pedagogical compulsory courses or not, as well as university levels, with the most significant difference being between freshman and senior students. Additional significant differences were found concerning the sub-sections of the survey as well.

For ChatGPT ELT attitude survey, the results display a higher-than-average mean score for ELT attitude, overall higher than TAME-ChatGPT scores. No statistically significant difference was found between total ELT attitude score and gender in addition to between students who have heard of ChatGPT and ones who have both heard of and used it. Although statistically significant differences were found between the sub-sections of the ELT attitude survey and university year, as well as other variables, no statistically significant difference between university levels and total ELT attitude survey score was found, as well as a difference between students who have taken ELT-based pedagogical courses and ones who have not.

The results of the interview exhibit similar findings to the results of the surveys. Students are overall reserved against the use of ChatGPT in various areas of life and work, especially when it comes to using it as a prospective English teacher. It was also observed that students were surer of their attitude towards ChatGPT as their university years were higher, which can be shown in the various statistically significant differences between

lower university years and higher university years found for general attitude, ELT attitude, and also literacy concerning ChatGPT.

From the table below, how the results of the study answer reach research question can be seen.

Table 4.32

Table for research questions and the related results.

Research Question	Result
What is the ChatGPT literacy of ELT students?	The ChatGPT literacy of the students are above average, 65.8 out of a 100 considering the minimum and maximum possible scores for the ChatGPT Literacy Scale.
What is the ChatGPT attitude of ELT students?	From the results of TAME-ChatGPT, the ChatGPT attitude of the students are overall 53,9 out of 100, only slightly above average. Their ELT attitude is 60.4 out of 100, still only above average.
Is there a significant correlation between ELT students' ChatGPT literacy and ChatGPT attitude?	Overall, students' ChatGPT literacy and ChatGPT attitude weak a weak but significant correlation. Between ChatGPT literacy and ChatGPT ELT attitude, the significant correlation is even higher.
Is there a significant correlation between ELT students' ChatGPT literacy and whether if the students have taken compulsory pedagogical lessons in the field?	From the data, it can be seen that whether students' have taken compulsory pedagogical lesson in the field does play a significant part in their ChatGPT literacy.
Is there a significant correlation between ELT students' ChatGPT attitude and whether if the students have taken compulsory pedagogical lessons in the field?	From the data, it can be seen that whether students' have taken compulsory pedagogical lesson in the field significantly affects students' ChatGPT attitude.

4.4. Discussion of the Findings

For the purposes of answering the research questions set by this thesis and therefore see whether the null hypotheses stand or are thrown away when the results of the study are given, this section will look at the research questions set for the purposes of this study through the findings and results shown through the data collected.

Firstly, the first research questions “What is the ChatGPT literacy of ELT students?” were strived to be answering with participants from a state university in Western Türkiye participating in a questionnaire with its first survey being an adapted version of the ChatGPT literacy scale (S. Lee & Park, 2024). The results revealed that the students self-reported to be literate in ChatGPT, although only moderately and not to a high degree. The accompanying interviews also illustrated that while students were aware of ChatGPT and how to use it overall, they were mostly not aware of the intricacies of its use in edge-cases, uncommon situations, and its potential to create material, ideas, and so on for other aspects of education for both teachers and students other than the most commonly recognized and researched potentials of large language models such as creating text, evaluating writing, giving activity ideas, helping with spell checking, and so on.

For the ChatGPT attitude of students, to answer the second research question, “What is the ChatGPT attitude of ELT students?”, two different surveys were utilized. The first survey is the adapted TAME-ChatGPT (Abdaljaleel et al., 2024) and ELT attitude survey constructed for the purposes of this study. The results showcased a higher than average ChatGPT attitude for the students, with ELT attitude being even higher than the results of the TAME-ChatGPT scale. The preceding interviews also revealed that the students were more inclined to use ChatGPT, and believed in their ability more when using it, when it comes to the area of ELT, this can be hypothesized to be because of their knowledge in the area being more specialized and overall higher than when looking at ChatGPT from every general angle to possess a general stance on different aspects of it. When asked about ChatGPT in general terms, they were more unsure of their answers and stance, while in the form of ELT, they were surer. For example, the junior students were adamant in their understanding that they would not use ChatGPT to rate their future students’ exams, and would also not use ChatGPT in the language learning and teaching

process as the sole tool and designer of any material or activity, but rather a tool that could help them create their own material and teaching.

The third research question “Is there a significant correlation between ELT students’ ChatGPT literacy and ChatGPT attitude?” was strived to be also answered by comparing the means of the results for the ChatGPT literacy survey and both TAME-ChatGPT and ChatGPT ELT attitude survey. The results do showcase significant positive correlation between literacy and attitude. For the fourth and fifth research questions, “Is there a significant correlation between ELT students’ ChatGPT literacy and whether if the students have taken compulsory pedagogical lessons in the field?” and “Is there a significant correlation between ELT students’ ChatGPT attitude and whether if the students have taken compulsory pedagogical lessons in the field?” were also found to have positive answers overall, and the results were given in detail about different subsections and factors of ChatGPT literacy and attitude concerning the variables of whether the students have taken ELT-based compulsory pedagogical lessons and in general, their university year/level. Similar results were also observed in the interviews, where if a student understood ChatGPT more as a large language model and its capabilities, the more they had positive attitude and overall, more nuanced outlook towards it. Additionally, as the students’ university level increased, they seemed to be more capable of understanding how to use ChatGPT practically in the field of ELT, but also as university students themselves to help them produce better material and learn more efficiently and effectively.

While not necessarily linear for all the aspects and factors, it seems that in general, if the university year of the student is higher, than they significantly more literate concerning ChatGPT than the lower years, especially if the year gap is 2 or 3. The highest number of significant differences were observed to be between freshman and senior students. Overall, freshman students were shown to have a lower ChatGPT literacy compared to all other 3 years, even though the differences are not always significant. For example, the differences were not observed to be significant when it was concerning the section for critical evaluation, even less so for ethical competence. This has the potential to point to the idea that while the ability to see the downsides and potential pitfalls of ChatGPT can be seen by all years similarly, seeing how the tool can be utilized and used

for the benefit of the one using it and their objectives differ for students from different years, and between students who have taken important, ELT related pedagogical courses.

While no statistically significant difference was found for gender, gender was also used as a variable to see if it would significantly differ in the results of the second survey, TAME-ChatGPT. Even if gender as a variable does not concern the research questions written for this thesis, it was still checked as additional data and as a minor objective of the study. The results reveal significant differences for the score of TAME-ChatGPT. 33.3% of the participants were male and 64.5% of them were female, making the overall sample not equal in its categorization. Even so, the results were demonstrated.

In terms of pedagogical courses, there were significant differences found for the total score of the survey as well as some of its sections alone. Additionally, more differences were found when university years were compared. Similarly, to the last survey, the differences were seen to be the highest overall when it was between freshman and senior students, in addition to freshman and other years. Similar to the first survey, the section for anxiety did not differ in its score depending on the year of the student, which was the only factor that did not have at least one statistically significant difference between the university years of the students.

Similar differences and results to the TAME-ChatGPT survey was seen for the ChatGPT ELT attitude survey. For a different result, no significant difference was seen between the courses taken and courses not taken group concerning total ELT attitude survey score. For gender, only the section of teaching was found to have a statistically significant difference, with male students have a more positive outlook towards using ChatGPT to teach English. In terms of years, no statistically significant difference was found for the ELT attitude score, while some differences were seen in the subsections. Even though not significant, the most difference was observed between freshman and senior students in terms of their ELT attitude score. While the students were overall more interested and principled in their ideas of using ChatGPT for ELT compared to ChatGPT attitude in general, the difference could not be observed statistically. The reason for this can be the idea that while making the attitude area about ELT does make students more focused on their attitude and outlook towards ChatGPT, this can take the results in to a more extreme positive, or a more extreme negative position.

In the literature, AI and chatbots have been instrumental and highly practiced and researched area when it comes to language education. The potential of the use of artificial intelligence in language education has been a topic of consideration before the technology was thought have reached a point where its potential could be realized (Wilson & Bates, 1981). By some researchers, the ultimate goal of AI in education was seen as its potential to one day be advanced enough to replicate the teacher and be a successful tool in helping the learners reach their language goals (Bailin, 1988). While it is argued whether this goal is reached with the current state of AI and large language models, one thing that can be said for certain is that if the goal has the possibility to be reached, the current state of technology and advancements makes humanity closer than ever to it.

The opinions of language learners as well as language teachers have been a topic of study in almost all areas of language education, ChatGPT included. The results from a case study interviewing 4 language learners studying in a school of foreign languages of a state university in Türkiye displayed the benefits and drawbacks of ChatGPT concerning speaking (Üstünbaş, 2024). Some of the discussed topics by the students were similar to the answers given by the participants participating in the interviews for the present study. For drawbacks, the potential for incorrect responses were given by both participant groups, as well as the benefits for providing opportunity to practice, and constructive feedback. Furthermore, in the study done by Üstünbaş (2024), ‘the importance of prompts’ was given as both a benefit and drawback. This was similar in the current study, since the students were adamant in the importance of giving context to the AI through the prompts to receive more adequate output from ChatGPT. They discussed the importance of the aspect, but overall, they were leaning more towards it being a drawback than a benefit. Even though they saw the benefit in receiving correct and adequate answers from ChatGPT when it is giving appropriate prompts, they were still not trustful towards the answers, and would always double-check the output they were given, since they believed ChatGPT was prone to giving incorrect or misleading answers. For example, a junior student, Student 3, complained about ChatGPT making mistakes even with simple prompts and requests, ‘I caught its mistakes a lot when I tried it. For example, I would ask for it summarize something about I topic I knew about, and I caught it making mistakes a lot.’”

In terms of ChatGPT literacy, even if research with a focus on researching ELT students' ChatGPT are low in number, the number of researchers pointing to the importance of having students and teachers with a high digital literacy and literacy around AI are not (Baidoo-anu & Owusu Ansah, 2023; Bender, 2024; McCallum, 2024). AI literacy can be described as the ability “to know and understand, use and apply, as well as evaluate and create AI” (Ng et al., 2021), fitting in line with the competencies given in Bloom's Taxonomy. The idea of having specialized courses for students to help them learn the benefits and risks of ChatGPT are ideas put forward by researchers (Meniado, 2023), pointing out once again the importance of AI and teaching literacy around AI and ChatGPT to help teachers and students use the tool to their benefit and improvement while avoiding the pitfalls of the potential negatives of it.

One study aiming to find out the AI literacy of different teacher and student groups was designed around finding out ChatGPT literacy of university teachers, student teachers and others, and school teachers (Q. Ma et al., 2024). The overall mean scores of the sub-groups defined for the survey created for the purposes of the study, which were 'benefits', 'limitations', 'prompts', 'evaluation', 'assessment', and 'ethics', were overall much higher in their scores compared to the results found in this study. For example, the self-reported literacy of the participants for the current thesis for thesis came to be a 3.49 mean score between 1 and 5, while the mean score for other study was 4.58 for student teachers & others. For this reason, it can be stated that the students' ChatGPT literacy in this thesis was lower in average scores compared to the limited amount of research done with students, using other scales.

The study done to develop the scale ChatGPT literacy scale was developed with 822 college student participants. While the main aim of the study was to develop the aforementioned scale, and therefore did not give the results of the application of the questionnaire, but presented the finding that there was positive association with ChatGPT and user satisfaction, and the scale was developed to a basis for “future studies exploring the effect of ChatGPT literacy on various outcome variables such as user satisfaction, academic achievements, creativity, and productivity.” (S. Lee & Park, 2024).

One study done using an adapted version of the same ChatGPT literacy scale used for this study was done to see if ChatGPT literacy was a predictor of perceived ease of use, perceived usefulness, and a moderator for the relationship between behavioral

intention and actual use (Li & Thien, 2025). The study was done on 543 Chinese EFL learners, the largest group within the sample was undergraduate students, which was 199, making up 36.65% of the total sample. The sample also included 169 master's degree students, 98 doctoral students, and 77 students studying to achieve associate degrees or below. There was a total of nine hypotheses laid out for the purposes of the study, which were:

“H1. Perceived ease of use positively predicts perceived usefulness.

H2. Perceived usefulness positively predicts attitude.

H3. Perceived ease of use positively predicts attitude.

H4 Attitude positively predicts behavioral intention.

H5. Behavioral intention positively predicts actual use.

H6. Perceived usefulness mediates the perceived ease of use-attitude association.

H7. ChatGPT literacy positively predicts perceived ease of use.

H8. ChatGPT literacy positively predicts perceived usefulness

H9. ChatGPT literacy moderates the existing relationship between behavioral intention and actual use” (Li & Thien, 2025)

The study adapted the ChatGPT literacy scale developed by Lee and Park (2024) in addition to adapting attitude items from a study done with the technology acceptance model (Abdul Ghani et al., 2022) to find out the participants' ChatGPT literacy to see if the hypotheses 7, 8, and 9 were correct. To collect data for the other variables, the subscales were adapted from other informal digital learning of English studies (G. Liu & and Ma, 2024; G. L. Liu et al., 2024).

The results of the study showcased that all nine hypotheses given were statistically proven to be correct for the aims of the study. The results of the study done for the purposes of this thesis also show similar results to the study done by Li and Thien (2025). After linear regression analysis was done on the variables for this study, similarly, perceived ease of use was found to be a significant positive predictor of perceived usefulness ($p = ,000$). Additionally, perceived usefulness was found to be a significant positive predictor of attitude ($p = ,000$). Furthermore, perceived ease of use was found to be a significant predictor of attitude ($p = ,000$). These results reveal a high value of

similarity between the results found for the first three hypotheses set for the study done by Li and Thien (2025) and the results of this thesis.

Linear regression analyses done also showcased ChatGPT literacy as a significant positive predictor of perceived ease of use ($p = ,000$), and perceived usefulness ($p = ,000$). This also exhibits a similarity for the results of hypothesis 7 and 8.

As a study done which includes the aims of seeing a relationship between ChatGPT literacy and attitude over its sub factors and facets, the study done by Lie and Thien (2025) stands alone in its particular place in the literature currently. Although it can be assumed that more studies will be done on the subject, the sole published study on the topic is also focused on EFL learners striving to learn English, which is different than the sample and population for the thesis being ELT students studying to become English teachers. Even so, as expressed, the results of both studies show more similarities in their results and findings than not.

While not focused on ChatGPT in particular, a study done on 409 university students on AI-related digital literacy and attitude found that there was a significant correlation between digital literacy and the students' use of AI tools, meaning that digital literacy is increased with more experience with AI tools (Joseph et al., 2024). In this thesis, it can be seen that separating the students who have only heard of ChatGPT and ones who have used it as well, the difference between the two groups present that those students with an experience of ChatGPT use have significantly more self-reported AI literacy, which is indicated from the independent samples t-test results ($t(122) = 2,204$, $p = ,029$).

Another study done on 212 pre-service early childhood teachers from the United States revealed a correlation between the variables of digital literacy, self-efficacy, and perception of AI education for young children (Lim, 2023). This is also similar to the findings of this research that showcases a significant correlation between ChatGPT literacy and attitude. Such similarities were observed comparing the two studies, but the difference in participant characteristics and selection should also be a factor that needs to be addressed when making such comparison. A different study also revealed that after a course involving AI, some students “developed a more negative view after witnessing how AI-generated answers and images could be potentially harmful or lacked broad representation” (Hu, 2025) which differs from the general findings of this thesis

showcasing students generally have more positive attitude towards AI, ChatGPT specifically, when they are more experienced with it.

For ChatGPT literacy, there may be certain variables not accounted for within the confines of this study, or not directly related to the study, which may be the affecters of the results seen in this study. These variables can in the control of the participants, or they may be outside of the participants' control. For example, in terms of ChatGPT literacy, there may be several reasons outside the students' control that may play a part in their literacy being only average from what can be seen from the resulting statistics gained from the results of the ChatGPT literacy scale. Significant relations regarding some variables used for this study has been statistically calculated to have a significant effect on ChatGPT literacy. Similar variables can be hypothesized to have a possibility of playing a similar role for ChatGPT literacy; for example, students' overall digital literacy may play a significant role in their ChatGPT literacy as well. In addition to overall digital literacy, students' screentime, time they spend on smartphones, computers, their hobbies, and so on may present significant relationship with ChatGPT literacy as well.

While a certain degree of overall digital literacy is expected from language educators, in student profiles, there can be students who are more inclined to have more digital literacy that may be in relation to their hobbies and overall interest in technology. There is a possibility of such aspects translating to ChatGPT literacy as well, more so if a connection between digital literacy and ChatGPT literacy can be established, with a possibility of such a concept being a future study and a way to improve the literature of the subject further and fill the research gaps present in the area this particular topic is included.

There is also the possibility of a connection between ChatGPT literacy and overall GPA. In addition to GPA, students' overall willingness to communicate, participate, and be present in different lessons and overall have motivation towards the subject of ELT is also a possibility and there is a possibility of a bridge between motivation and ChatGPT literacy as well. While the students' academic success was not put forward as a concrete variable in this study, researcher's anecdotal observation during the interview process of this study brings out the idea that there may be a positive correlation between academic motivation and ChatGPT literacy, and this topic may also be a future study as well, which is also discussed in the last chapter of this thesis.

Many of the same possibilities regarded above may also apply for ChatGPT attitude as well. However, while the aforementioned variables are looked at through their possibility of affecting ChatGPT literacy in a positive way where; for example, if a student has more academic motivation, they have more ChatGPT literacy as well. When it comes to ChatGPT attitude, the spectrum has more of a possibility of going both ways; meaning that the researcher should focus on whether different variables affect ChatGPT attitude positively or negatively. There is also the possibility that some ChatGPT attitude factors may be affected positively while others may be affected negatively through the lens of the same variable and its connection with attitude. For example, the interviews showcased that students' literacy with ChatGPT also affected how much they knew and recognized the negative aspects and potential pitfalls of ChatGPT as well. Students who had experience and overall idea around the topic of ChatGPT were also aware of different argument points for ChatGPT that were less known by the overall public. For example, while an aspect of ChatGPT like its tendency to be used by students potentially immorally to complete assignments or cheat at exams, a topic like the machine learning algorithm having access to copyrighted material and using such material to improve its model and capabilities, which was pointed out by one of the junior year students in the interview process.

There may be variables like age, motivation, academic success, academic motivation, personality, socioeconomic background, digital literacy, and so on that may negatively or positively affect attitude towards ChatGPT and the multiple factors that encapsulate different aspects of ChatGPT attitude.

In addition to the different factors, there are also possible reasons for why the variables present in this study affect other variables in different ways. For example, the apparent difference between students' university year and their ChatGPT literacy and attitude were all highly relevant in understanding for what reasons university year may affect these two variables in a significant manner. There can be significant differences seen when the years are separated into two categories, first and second year students, and third and fourth year students, which the variable of whether the students have taken pedagogy-related ELT-based compulsory lessons accomplishes in doing so, which may be the result of various differences between these two groups. While aging itself may not be the sole reason, the cognitive improvement of the participants' can be a factor in their

overall understanding and therefore their attitude towards ChatGPT improve and be more detailed and informed. Aside from cognitive factors, their academic experience throughout university also has a possibility of being one of the most significant reasons for this difference that can be seen between university years. Even if the university years are not separated in half, and looked at one by one, the lowest number of significant differences are seen between years that are next to each other, and further apart the years are, more differences are seen, with the highest number of differences in both ChatGPT literacy and attitude seen between freshman and senior, first and fourth years students.

The differences concerning ChatGPT literacy and attitude observed regarding the university years of the participants may stem from their ELT knowledge and the number of lessons they have taken through their time studying to become English teachers. Methodology, pedagogy, and overall education-based courses starting from the 101 courses over the main topics of education to small but important aspects of language education may have shaped the students overall outlook towards the use of such technology as a student and/or as a teacher, in addition to improving their ability to seek out new tools for technology use, and be more fluent in those new tools. Considering the fact that there are a few technology-based courses regarding ELT curriculum in Türkiye in many universities, the university where the data of this study was collected included, considering when they have taken those courses, there is also the possibility that ChatGPT was explicitly mentioned if they have taken the courses after ChatGPT come to public and literature-based interest, which also has a possibility of shaping the participants' ChatGPT literacy and attitude.

While no direct relation can be drawn to recognize why the students' overall ELT attitude is higher concerning ChatGPT compared to general ChatGPT attitude, and because a significant moderate correlation between ELT attitude survey score and attitude survey score was found from the collected data, the overall experience of students can be seen to affect their ChatGPT ELT attitude and overall ChatGPT attitude, which may theorize the idea that overall experience with ChatGPT, whether ELT-oriented or not, does showcase a more positive approach to both attitude types from the participants. While the notion that people who use ChatGPT more has a more positive, or at least more educated attitude towards it is a statement that cannot be stated without doubt from the

resulting data and findings from this study, it may be a steppingstone towards improving the likelihood of the hypothesis being correct.

From what can be seen from the results of the surveys and the interview questions, even though a general statement cannot be made, it may be a possibility that ChatGPT literacy increases with positive ChatGPT attitude; and literacy over any technological tool in education is highly important for the student and the teacher, there can be a number of additions to the curriculum, or the overall tools the language student and teachers use together or separately that can improve ChatGPT literacy, and attitude, and break the barrier between the user and what this new technology is capable of, and what it can be used to mitigate the risks and potential shortcomings of the tool, and to improve and benefit from its potential advantages. For example, workshops and/or additional courses to learn and use ChatGPT can be constructed in language education faculties for teachers and students alike. In the interviews, students have expressed that hiding or forbidding the use of AI and large language models would not be beneficial actions, since even if the teacher forbids the students from using ChatGPT, they could still use it, and hide that information from the teacher, creating a barrier in the teaching and learning process; therefore, rather than hiding it, interviewed students agreed that ChatGPT should be used; however, teachers and students should learn how to use it effectively and ethically. While the literature around the idea of ChatGPT workshops and AI education is low to none, especially in the language education section of the literature, it can be a topic that could be studied in the future to find out whether AI education in language learning classrooms can be beneficial to the process of AI use, or the possibility of such additional education affecting other aspects in the process like academic success, motivation, participation, and so on.

Overall, the findings of this thesis point out a high number of correlations between factors surrounding ChatGPT literacy and attitude involving ELT as both ELT university students and future language educators. The highly limited literature around the topic point out similar results to the findings of this thesis, although there are still a high number of gaps present in the literature around ChatGPT literacy and attitude, especially regarding the population of prospective English teachers, and the main aim of this thesis is to fill a small gap in the literature regarding the ChatGPT literacy and attitude of the aforementioned population, in addition to other correlations regarding variables such as

age, student experience with ChatGPT, whether the students' have taken ELT-based pedagogical courses, university year, and gender.



5. CONCLUSIONS AND IMPLICATIONS

5.1. Introduction

In this chapter, the study will overall be looked at from the beginning of preparation, data collection, data analysis, presentation of the findings, the results, and the conclusions and implications that can be gathered from the other steps of this process of research and study in this particular context. The main aim of this chapter is, through the findings of the study, to present what implications can be garnered by researchers, practitioners, students, and the readers overall.

The implications are research-based, whether this study was successful in filling the small research gap it has aimed to fill by presenting the data it has presented, and what other research gaps and questions exist that can be researched by other studies and researchers to fill the gaps existent even further. Furthermore, implications for practitioners of ELT will also be discussed in detail, and whether the findings of this study can be used to help the process of teaching English and what other points of consideration can be given to the practitioners. Furthermore, implications around not only teaching, but also learning English in the era of ChatGPT will be discussed. Lastly, the overall implications around AI and ChatGPT regarding research and education will be given that could be garnered from the results of this study.

5.2. An Overview of the Present Study

The current research is a concurrent triangulation design mixed-method study conducted to investigate the literacy and attitude of English language teaching students towards ChatGPT in Türkiye. Specifically, the study was conducted on 141 ELT students studying to become English teachers in a state university in western Türkiye. This mixed-method study had four different tools to collect data. The first three tools were survey designed to collected quantitative data regarding students' attitude and literacy concerning ChatGPT. First of these tools was the adapted ChatGPT literacy scale, adapted from the scale designed by Lee and Park (2024) for the purposes of finding out about the participants' ChatGPT literacy. The second tool was the adapted TAME-ChatGPT, a scale designed for the purposes of collecting data about the participants' ChatGPT attitude (Abdaljaleel et al., 2024). The third tool was the ChatGPT ELT attitude survey, designed with the items from TAME-ChatGPT with the purpose of collecting data concerning the participants ChatGPT attitude around the specific area of ELT. The fourth tool was the

interview questions developed for the purposes of being asked to the participants in semi-structured group interviews. In total, 8 participants were interviewed that volunteered from the main sample group of 141 participants: 2 of the participants were 2nd year student, 4 of the participants were 3rd year students, and 2 of the participants were 4th year students.

The data was collected through the aforementioned tools. In addition, students' age, whether they have heard and/or used ChatGPT before, their university year, their gender, and whether have taken ELT-based compulsory pedagogical courses before were also collected as data.

The quantitative data was collected through the three surveys and the data was analyzed using IBM SPSS 26 release version. In addition, the interviews were analyzed using the analysis method of codebook thematic analysis. The quantitative statistical analysis displayed that the students' ChatGPT literacy was above average, similarly to their overall ChatGPT attitude. Furthermore, a significant correlation between ChatGPT literacy and ChatGPT attitude was observed, as well as between other variables such as the university year of students. Overall, with the help of the results from the interviews, students' ChatGPT literacy and attitude, especially attitude regarding ELT was exhibited. The results revealed that ELT students were overall more positive than negative towards ChatGPT; however, they were also, in an above-average degree, aware of the potential shortcomings and downfalls of ChatGPT, and they were wary of it. While they were not as wary about using it altogether, they were especially wary about using it as the sole tool for text, idea, material, and so on, creation and usage. They express that ChatGPT is an advantageous tool, but its output should be double-checked by the user, and its correctness and appropriateness should be observed by whoever is using it as well, whether they be English teachers or learners.

5.3. Implications

This study focused on ELT students from a university in Türkiye regarding their ChatGPT literacy and attitude. The results from the collected and analyzed data for this purpose has exhibited a number of implications for various facets of language education, which will be listed and explained in detail in this section.

Firstly, the results highlighting that there was significant correlation between overall ChatGPT literacy and attitude reveal a relationship between the two variables

which can potentially benefit from being explored in practice for English language teaching and learning. Bender (2024) expresses AI, and AI literacy, as an integral part of digital literacy, which in turn has been showcased to have a significant positive effect on different facets of language learning (Ali et al., 2024; J. Wang, 2024; Zheng et al., 2025). Therefore, language teachers have the ability to facilitate AI literacy through various methods in their students, and have the possibility of increasing the students' likelihood of success in different areas and facets of language learning. It can be seen from other studies that students' potential for improving their digital literacy is highly mediated with teacher support (Zheng et al., 2025), therefore the role of the teacher in helping the learner increase their digital literacy is a highly important aspect that should not be ignored by the practitioners.

The students interviewed for this study agree that students and teachers alike should know how to use AI properly if they are to facilitate its use in teaching and/or learning. From this, it can be pointed out that facilitating language teachers' AI literacy and attitude is also highly important for the improvement of digital literacy and therefore language teaching and learning. The results of this study also showcase a stronger correlation between AI literacy and ChatGPT ELT attitude compared to AI literacy and general ChatGPT attitude. This demonstrates that language teachers have potentially more positive outlook towards ChatGPT as language teachers and from the lens of overall language learning and teaching when they are more literate in ChatGPT.

Overall, from the two findings that present the importance of ChatGPT literacy for students and teachers, it can be stated that additional developments can be made to improve these two parties' ChatGPT literacy. Additional courses, more active use of ChatGPT for both students and teachers, workshops dedicated to digital literacy and ChatGPT literacy, short presentations, courses, or lessons showcasing how to use ChatGPT and its intricacies at the beginning of education years or singular courses have the potential to help with increasing ChatGPT literacy for students. For teachers, additional courses can be developed to reveal how and in what different ways ChatGPT can be utilized to help make teaching more efficient and effective at the same time. Although, from the results of the interviews showcasing that many pitfalls exist in ChatGPT use, it can also be argued that these developments in ChatGPT literacy can highly benefit from also showcasing where ChatGPT is strong and weak, how it can be

used to avoid the pitfalls it possesses. Aspects such as creativity loss, incorrect prompting, AI hallucinations, incorrect responses, misleading responses, ethical boundaries and argument areas around AI, and other potential negative aspects of ChatGPT and AI in general should be taught to both teachers and students to make sure that the potential negative sides are understood as much as potential positive sides of it.

From the results of the study, it can be seen that while the participants have thought of ChatGPT as a potential tool for lesson planning, material creation, and other similar steps of language teaching, the potential of what ChatGPT can accomplish, especially in its more current versions are not thought as detailed by the students. For example, while it is understood that ChatGPT is proficient in creating reading material for students, helping create material, for example, pronunciation activities through its text-to-speech function, or even facilitating the students to use it whenever they need help in pronunciation is a potential use of ChatGPT not well thought of by the students; on the contrary, when asked about its use in different language skills and other sub-skills, like pronunciation, one of the students expressed that they could not imagine how the large language model could be used for teaching pronunciation even if one wanted to. This also showcases the rapid developments in AI, considering that ChatGPT was not able to listen to input, or produce text-to-speech patterns when it was first released to the public, but subsequent versions saw improvements in these aspects.

Different variety of versions and with them new features to ChatGPT are always available, and not only ChatGPT, but other large language models are also available as well, and being developed by large companies all over the world. Therefore, it can be stated that using a single piece of software to teach AI literacy may be simpler in its education than showcasing various different models and interfaces to the learners of AI literacy, but it can also be beneficial in helping the learners learn variety in AI and large language models, and various different features available in different models.

For AI education, different methods and tools are being developed to help in AI literacy education in teaching (Zha et al., 2025). From the results of this study, it can be seen that choosing and using the effective tools created for the purposes of teaching AI literacy to teachers and/or students can also be highly beneficial in helping facilitate learners' positive attitude towards ChatGPT strengthen. However, if the opposite is assumed, that ChatGPT attitude affects ChatGPT literacy, in that case, the same

implications apply, but only for teaching ChatGPT attitude to teachers and students, which can potentially develop, negative or positive, from more exposure to AI and using ChatGPT for personal and/or educational purposes.

Another implication for ChatGPT use can be found in the results of the ChatGPT ELT Attitude survey and also from the answers given in the interviews. Overall, the participants can be regarded as increasingly wary when the importance of the material created by ChatGPT increases. For example, students are increasingly more reserved against using ChatGPT when the material created is something like an exam, although they are less reserved when they hypothetically use the large language model to create something like homework material that is ungraded. From this, it can be stated that one of the most important implications for the language teacher and language learner that can be garnered from the results of this study is for the person to be aware of the shortcomings of AI and large language models, and use them accordingly while keeping those potential negatives in mind. One of the most prominent arguments against ChatGPT is that its potential to give incorrect, misleading, incomplete, or made-up answers. While the developments in AI have shown decreased number and percentages of these aforementioned shortcomings appearing in the output given by ChatGPT, it is still a prevalent issue that needs to be addressed. This is not only relevant in language education, but in almost every field and sub-field. Failure to recognize these potential dangers of AI and using it without regarding its potential for bad information can lead to “dangerous consequences” for various fields, some of which involve human lives in a daily fashion (Scheinkman et al., 2025).

In language learning, the implications for avoiding the shortcomings of AI can come from educating the user, whether they are teachers or students, on how to use AI, the intricacies and different facets of AI. In addition to education in various forms concerning AI and ChatGPT, the results of this study also showcase that the learner and the educator should be open to new tools and technology for their field, and should strive to learn and educate themselves on how to use these new tools for their benefit. This applies to ChatGPT, but is not limited to it. It also stands that the traditional method of academy and publishing leads to a gap between current technology and its study, which can also be highly apparent when said technology improves on a faster basis. Therefore, the contemporary teacher should have the tools available to them to discern and analyze

new technology, tools, and developments through a critical lens to help them understand how they can use the new tool for their benefit while also recognizing the potential downsides of it. For this reason, while AI literacy and ChatGPT literacy is highly important in this particular context, the person in language education, whether teacher or student, should be well-equipped with digital literacy, and most importantly, critical literacy to help them self-govern their way through contemporary technology.

One of the main benefits mentioned at various steps on the way is the potential to save time via ChatGPT use for teachers, which points to a widely-known issue in the practical field for teachers that they believe they have too many classes, too many students, too many working hours, but not enough compensation and employee rights, among many other issues and problems (Yıldırım Taştı & Engin, 2024). While it is beyond the scope of this study to suggest that such problems should be fixed, AI use seems to be self-reported as a small remedy to some of the problems mentioned. However, this can also cause AI to be used inappropriately, not checking its output before presenting the material created by it to the learners, which can cause disruptions in the language learning process for both the learners and the teachers.

ChatGPT and AI overall are tools that need to be used appropriately and with consideration, not unlike many other tools, digital, technological, or otherwise. Much like how an insufficient and inconsiderate use of a coursebook can cause the language learning process to falter and be negatively affected, the same goes for the most current of technological tools like large language models.

In education, integrating ChatGPT and AI into different aspects with respect to different considerations about AI and its various facets can prove to be beneficial, although more studies need to be conducted on various different aspects of language education to see how they are affected with AI use, integrating AI into school curricula in addition to promoting professional development and lifelong learning with the help of AI, “support reskilling and upskilling programs”, and promoting “diversity and inclusion on AI education and careers” have the potential to be beneficial towards education altogether (George et al., 2023), in addition to teaching and learning English. With that in mind, the integration of such aspects should be researched to see whether they would provide benefit or disadvantage towards different aspects of education, and if benefit is

observed, it still needs to be integrated with intent, focus, and well-prepared steps of integration and education of AI.

Large language models are developing in a way that, currently, literature cannot keep up with. Even with preprints and conference proceedings, the act of research and presenting that research in a research paper takes a considerable amount of time. There have been versions and models of ChatGPT, let alone other competitors developed by different countries all around the world. ChatGPT itself has seen different public version from GPT 3.5, GPT-4, GPT-4o, o3-mini, o1, and GPT-4.5, with more models in development. Other large language models like Google Gemini, Microsoft Copilot, Perplexity AI, Meta AI, and DeepSeek have also been released in the wake of the first public release of ChatGPT. These large language models are overall similar to ChatGPT in their use, with some of them specializing in different areas and topics. For example, Microsoft Copilot is specialized for task automation and overall office automation, and something like Socratic is used in software engineering for monitoring code quality.

Additionally, there are also many factors not studied throughout this study that have potentially important implications for ChatGPT use in language education by both learners and educators. It was revealed that ChatGPT literacy and attitude have significant correlation with each other, meaning that if the goal is to educate the person so they are more aware of the new technological tool, how to use, how to avoid its potential shortcomings, and benefit from its advantages to the best of their ability, their attitude will also improve to be more positive as a result, at least from what was seen from the participants of the study. Through these potential results, certain implications can be made to improve the ChatGPT of students and teachers. Firstly, as explained, additional courses, workshops, demonstrations, and so on concerning ChatGPT can work to improve students' ChatGPT literacy and help them be aware of all the technological developments of ChatGPT and the potential uses of it. Because of the fact that ChatGPT is ever-improving similarly to how new technologies do for the first years of their inception, since there are many kinks that can be ironed out with haste, and because of the potential of public interest and the monetary potential that carries, combined with the idea that educational curriculum, because of its planning and implementation, can lag behind years in many cases, there is the need for keeping the education of ChatGPT current, contemporary, and alive throughout its developmental stages. From the answers given to

some interview question, it was seen that while students were aware of ChatGPT, and have used them before, because of the rapid new developments of AI, some of them were behind on information about how it could achieve. For example, one of the students in the same group interview was aware of the idea that text-to-speech was possible through ChatGPT use, while the other student was not sure that was a feature of ChatGPT. This can derive from the fact that when ChatGPT first entered the public consciousness, text-to-speech was not available as a full-fledged feature given to users; however, after some time, it was released to the public as a new feature of ChatGPT. For similar reasons, if large language model educational courses are planned to be given to students, or teachers, the material present should be kept up to date at all times.

In addition, there is the possibility of overall GPA being connected with ChatGPT literacy and/or attitude, in addition to other aspects of academic success like academic motivation, participation, student profile, learning styles, curriculum and syllabus design, education-adjacent activities, and so on. While no study exists yet to provide results that would prove the connection between ChatGPT literacy and/or attitude and the education-based variables given above, that would signify that improving ChatGPT literacy, and therefore potentially improving ChatGPT attitude in addition to all the other variables, in a positive direction is highly important for language education. There remains need to practically trial such hypotheses to see what the existence and use of ChatGPT directly or indirectly has an effect on in the language learning process.

Overall, the field of AI is competitive, and full of potential developments; however, it is also filled with ethical quandaries, contemporary political debate, potential harms in its use, or through its widespread appeal and ability for automation and easy output. These aspects of large language models and AI overall can pose a real threat to job security in various fields (J. Cao & Song, 2025). Currently less so the case in education due to active involvement of teachers being highly difficult to be replicated in a real-life setting by AI, the future of AI is yet to come, and while the rapid development of AI can be proposed to slow down similar to other rapid developing new advancements and tools in technology slowing down in development as they become more optimized and well-rounded, the potential of AI replicating the teacher altogether is currently a highly distant idea; however, the teacher using AI to improve their ability to teach English, and students using AI to improve their ability to learn English is not a distant

idea, but a contemporary one seen in practice all around the world. Therefore, while disagreements are widespread concerning AI, large language models, machine learning, big data, and public tools like ChatGPT, there is agreement that ChatGPT possesses a high potential to change things, or at least be a more powerful tool at the hand of the ones using it, even though no agreement is made in the argument whether that potential is for good, or for bad.

From the points mentioned, and the overall possible implications from the results of the study, it can be stated that whether by the name of ChatGPT or something else, AI has become a significant part of language education, and while large language models becoming a part of language education seems an inevitability, using it as good and effective as possible is not a given. Therefore, adequate education to teachers and students should be given concerning AI literacy and how to look at, understand, and analyze AI, giving input to it, and receiving output from it. Careful integration is an area that needs to be kept in mind by students and teachers alike by knowing the potential positives and negatives of AI. While ChatGPT can be integrated into the classroom, it should be done with the upmost care, and AI's current potential shortcomings should be well considered, and completed whole by the teacher if it is to be used.

5.4. Limitations of the Study

This study focused on ELT students from a university in western Türkiye concerning the students' ChatGPT literacy and ChatGPT attitude. Through the planning, implementation, and reporting stages of the study, and outside of these 3 stages, there have been various limitations for the study.

The most prevalent limitation of the study is the self-reported nature of the ChatGPT attitude, but more prominently ChatGPT literacy of the participants. The Likert scale surveys adapted and designed for this study have self-reporting as its main aspect of data collection. The students rate themselves on various items concerning their ChatGPT literacy and attitude. For this reason, the data collected is prone to being affected by self-report bias that can be seen in various fields, including social sciences. "There are many reasons individuals might offer biased estimates of self-assessed behavior, ranging from a misunderstanding of what a proper measurement is to social-desirability bias, where the respondent wants to 'look good' in the survey, even if the survey is anonymous" (Rosenman et al., 2011). There are questions in the scales that can

be regarded as direct self-rating items, and others that can be seen as indirect self-report items. Overall, even though face reliability and other validity and reliability types were checked extensively both by the designers of the scales adapted for the purposes of this study, and by checked by this study as well, it is still a possibility that rather a self-report, a test designed with the principles and framework designed for the ChatGPT literacy scale and TAME-ChatGPT can be used to test the literacy and attitude of students regarding ChatGPT. It is unknown whether the results would be different than the self-reporting of the students, but it still is a limitation of this study.

In terms of data analysis, one limitation of the study is that the qualitative data was not checked by a co-rater. Due to the time-related factors in addition to factors such as the feasibility of finding another rater to code the qualitative data, this thesis did not include a second rater. Although the results were checked in detail by the researcher and the thesis supervisor, and most diligence and attention to detail was used when qualitative data was coded and overall analyzed, due to the absence of a co-rater, this study has a lack of inter-rater reliability.

Another limitation of the study is the source of participants, number, and selection of participants. To explain, all the participants were taken from a single university, in a single department where the students are studying to become English teachers. To summarize the population to the whole of Türkiye, or a smaller but still big section of Türkiye, the participants could be taken from multiple institutions, and the results can be compared as a whole, and as different sample groups as well. In addition, because of the way the participants were selected for the interviews, via volunteering, the participant group chosen for the interview may not directly correlate to the rest of the sample, meaning that because of the fact that they volunteered, presumptions such as the volunteering group being more outgoing students, more social, more knowledgeable of AI, have self-perceptions about AI that they like to express, and so on, should be kept in mind. If the participants were selected randomly from the sample group, the generalization could have been better; however, because volunteering and consent is above all in research, this was developed to be the most ethical way of choosing the participants for the interviews.

One group of limitation for the study is the fact that other variables could have been researched with these findings as well. As it stands, this study only briefly mentions

academic success and language learning/teaching success in relation to ChatGPT literacy and attitude, and cannot be used as a study that displays whether or not ChatGPT literacy and/or attitude affects these successes; however, this can also be said as a suggestion for further studies, rather than a limitation of this particular study. Still, it can be stated that the findings from this study, when correlated with academic and language success in both learning and teaching can shine light on a vast number of AI-related questions in language education.

5.5. Suggestions for Further Studies

This study focused on the ChatGPT literacy and ChatGPT attitude of students from a state university in western Türkiye. In addition to variables and sub-categories of the main variables of ChatGPT literacy, ChatGPT attitude, and ChatGPT ELT attitude, different variables were collected as data and correlated with the three main data collected from the surveys and interviews: age, gender, university year, whether they have taken ELT-based pedagogical courses or not, ChatGPT awareness. However, due to the scope of this study, not every research gap in the field of AI research in language education could be filled; rather, this study was aimed at filling a small section of the research gap present, and various other studies can be conducted on the field to build on the emerging field of AI in language education.

One of the suggestions for further studies on the topic, as briefly discussed in the implications section, is that there can be various studies conducted on the topics of AI literacy and attitude, together or separate, and their correlation with language learning success. While the current study reveals, at least for the sample and the comparatively small population, that AI literacy and AI attitude are related in various ways and factors, no study exists where students learning language are used as a population where their AI literacy and attitude are collected as data and compared with their language learning success. To do so, their current language learning success can be learned through assessment; or, experimental research can be conducted where, for a singular example, the experimental group goes through an AI literacy course, and the control group does not. After, the groups can both have courses that utilize ChatGPT to see whether the experimental groups fairs better in their learning, worse, or not significantly different. Another example can be through an experimental design that is similar, but two

experimental groups are present, where one group is made up by the students with higher ChatGPT attitude, or the other group is for students with lower ChatGPT attitude.

While such examples can go on, the fact is that such studies do not exist as of yet, and ChatGPT's different facets, not only attitude and/or literacy, but other, broader or narrower factors and aspects can affect different areas and properties of language learning can all become different studies with high merit behind their role of filling the research gaps existent in the field of AI in language learning. Not only so, but the same gaps exist for different participants, different student groups from young learners to adult learners, teachers teaching in different levels, academicians, ELT students studying to become English teachers like in this study, and other participant groups from different backgrounds, different cultures, and so on. The same research steps that were taken during this study can be taken with different samples. For example, this study was focused on students in a state university. The same study can focus on private university students, and there is even the possibility of comparing the results gathered from state university students and private university students. The data can be collected from different universities in different parts of Türkiye, in different countries, and so on.

The idea that ChatGPT attitude and literacy are connected through their significant correlation that was found in this study can be studied further. A study structure where the participants are given the same questionnaire format as this study, the three surveys, ChatGPT literacy scale, TAME-ChatGPT, and ChatGPT ELT attitude survey, and after they data is collected from their concerning these three surveys in addition to the interview format seen in this study; after, the participants given a ChatGPT courses that aims to develop their ChatGPT and overall AI literacy. After the course is done, the participants are given the questionnaires, and the interviews are conducted again. The results of the study, and the differences between the pre and post-tests of this proposed study can aim to more concretely answer the question of whether ChatGPT literacy directly affects ChatGPT attitude as well, in addition to ELT-based ChatGPT attitude.

There is also the need for research in terms of ChatGPT attitude and literacy that collected the data around the two variables not from self-reporting questionnaires, but from assessment tools designed specifically to collect data for ChatGPT attitude and ChatGPT literacy. Such a study can work to minimize self-reporting bias and provide more objective data towards students' real ChatGPT attitude and literacy. There is also

space for space where students are both given self-reporting questionnaire, and tests after, to see whether their self-reporting matches their real ChatGPT literacy and/or ChatGPT attitude.

Overall, the research gaps in the area of AI research in language educations are vast and many, due to the fact that the technology is highly recent, and its integration with language education even more so. Even though AI as a concept in its modern form was first mentioned in 1955 as a proposal (J. McCarthy et al., 2006), between that point and 70 years later, 2025, AI is more relevant than ever, and research on the topic in any field of research is important. As so in ELT.



REFERENCES

- Abd Kadir, S., Lokman, A., & Tsuchiya, T. (2021). Emotional responses towards unity YouTube videos: Experts vs. viewers perspectives. *International Journal of Affective Engineering*, 20. <https://doi.org/10.5057/ijae.IJAE-D-20-00033>
- Abdaljaleel, M., Barakat, M., Alsanafi, M., Salim, N. A., Abazid, H., Malaeb, D., Mohammed, A. H., Hassan, B. A. R., Wayyes, A. M., Farhan, S. S., Khatib, S. El, Rahal, M., Sahban, A., Abdelaziz, D. H., Mansour, N. O., AlZayer, R., Khalil, R., Fekih-Romdhane, F., Hallit, R., ... Sallam, M. (2024). A multinational study on the factors influencing university students' attitudes and usage of ChatGPT. *Scientific Reports*, 14(1), 1983. <https://doi.org/10.1038/s41598-024-52549-8>
- Abdul Ghani, M. T., Hamzah, M., Wan Daud, W. A. A., & Muhamad Romli, T. R. (2022). The impact of mobile digital game in learning Arabic language at tertiary level. *Contemporary Educational Technology*, 14(1).
- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., & Anadkat, S. (2023). Gpt-4 technical report. *ArXiv Preprint ArXiv:2303.08774*.
- Acosta-Enriquez, B. G., Arbulú Ballesteros, M. A., Arbulu Perez Vargas, C. G., Orellana Ulloa, M. N., Gutiérrez Ulloa, C. R., Pizarro Romero, J. M., Gutiérrez Jaramillo, N. D., Cuenca Orellana, H. U., Ayala Anzoátegui, D. X., & López Roca, C. (2024). Knowledge, attitudes, and perceived Ethics regarding the use of ChatGPT among generation Z university students. *International Journal for Educational Integrity*, 20(1), 10. <https://doi.org/10.1007/s40979-024-00157-4>
- Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *Artificial Intelligence Applications and Innovations* (pp. 373–383). Springer International Publishing.
- Ahmad, Z., Kaiser, W., & Rahim, S. (2023). Hallucinations in ChatGPT: An unreliable tool for learning. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 15. <https://doi.org/10.21659/rupkatha.v15n4.17>
- Al-Amin, M., Ali, M., Salam, A., Khan, A., Ali, A., Ullah, A., Alam, M. N., & Chowdhury, S. (2024). History of generative Artificial Intelligence (AI) chatbots: past, present, and future development. <https://doi.org/10.48550/arXiv.2402.05122>
- Alamri, B. (2021). Challenges of implementing technology in ESL writing classrooms: A case study. *English Language Teaching*, 14, 36. <https://doi.org/10.5539/elt.v14n12p36>
- Alexander, K., Savvidou, C., & Alexander, C. (2023). Who wrote this essay? Detecting AI-generated writing in second language education in higher

- education. *The Journal of Teaching English with Technology*, 23(2), 25–43.
<https://doi.org/10.56297/BUKA4060/XHLD5365>
- Ali, Salah M, Yunus, Kamariah, Alshaikhi, Turkey, & Aliia, Abdussalam Marie. (2024). The use of digital learning in the EFL classroom and its effect on self-perceived knowledge, learning, and English use at the University of Tabuk. *E-Learning and Digital Media*, 22(1), 67–85.
<https://doi.org/10.1177/20427530241229125>
- Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*, 15(2), e35179.
<https://doi.org/10.7759/cureus.35179>
- Allehyani, S., & Alghamdi, M. (2023). Digital competences: Early childhood teachers' beliefs and perceptions of ChatGPT application in teaching English as a Second Language (ESL). *International Journal of Learning Teaching and Educational Research*, 22(11), 343–363.
<https://doi.org/10.26803/ijlter.22.11.18>
- Alm, A., & Watanabe, Y. (2023). Integrating ChatGPT in language education: A Freirean perspective. *Iranian Journal of Language Teaching Research*, 11(3), 19–30. <https://doi.org/10.30466/ijltr.2023.121404>
- AlTwijri, L., & Alghizzi, T. M. (2024). Investigating the integration of artificial intelligence in English as foreign language classes for enhancing learners' affective factors: A systematic review. *Heliyon*, 10(10), e31053.
<https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e31053>
- Arefin, S. E., Heya, T., Al-Qudah, H., Ineza, Y., & Serwadda, A. (2024). Unmasking the giant: A comprehensive evaluation of ChatGPT's proficiency in coding algorithms and data structures.
<https://doi.org/10.48550/arXiv.2307.05360>
- Armstrong, R. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, 34. <https://doi.org/10.1111/opo.12131>
- Atwell, E. (1999). *The Language Machine*.
- Ayre, J., & McCaffery, K. J. (2022). Research Note: Thematic analysis in qualitative research. *Journal of Physiotherapy*, 68(1), 76–79.
<https://doi.org/https://doi.org/10.1016/j.jphys.2021.11.002>
- Bai, Y., Li, J., Shen, J., & Zhao, L. (2024). Investigating the efficacy of ChatGPT-3.5 for tutoring in Chinese elementary education settings. *IEEE Transactions on Learning Technologies*, PP, 1–16.
<https://doi.org/10.1109/TLT.2024.3464560>
- Baidoo-anu, D., & Owusu Ansah, L. (2023). Education in the era of generative Artificial Intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52–62.
<https://doi.org/10.61969/jai.1337500>

- Bailey, D., Southam, A., & Costley, J. (2021). Digital storytelling with chatbots: Mapping L2 participation and perception patterns. *Interactive Technology and Smart Education*. <https://doi.org/10.1108/ITSE-08-2020-0170>
- Bailin, A. (1988). Artificial intelligence and computer-assisted language instruction: A perspective. *CALICO Journal*, 5(3), 25–45. <http://www.jstor.org/stable/24147434>
- Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, 33(1), 63. <https://doi.org/10.1007/s12525-023-00680-1>
- Barchan, J., Woodmansee, B., & Yazdani, M. (1986). A prolog-based tool for French grammar analysis. *Instructional Science*, 15(1), 21–48. <https://doi.org/10.1007/BF00139602>
- Barnard, Y. F., & Sandberg, J. A. C. (1988). Applying artificial intelligence insights in a CAI program for “open sentence” mathematical problems in primary schools. *Instructional Science*, 17(3), 263–276. <https://doi.org/10.1007/BF00048345>
- Baron, P. (2024). Are AI detection and plagiarism similarity scores worthwhile in the age of ChatGPT and other Generative AI? *Scholarship of Teaching and Learning in the South*, 8, 151–179. <https://doi.org/10.36615/sotls.v8i2.411>
- Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/https://doi.org/10.1016/j.inffus.2019.12.012>
- Bartle, R. (2015). Multi-User Dungeons (MUDs). In *The International Encyclopedia of Digital Communication and Society*. <https://doi.org/10.1002/9781118767771.wbiedcs009>
- Bender, S. M. (2023). Coexistence and creativity: screen media education in the age of artificial intelligence content generators. *Media Practice and Education*, 24(4), 351–366. <https://doi.org/10.1080/25741136.2023.2204203>
- Bender, S. M. (2024). Awareness of artificial intelligence as an essential digital literacy: ChatGPT and gen-AI in the classroom. *Changing English*, 31(2), 161–174. <https://doi.org/10.1080/1358684X.2024.2309995>
- Benotti, L., Martinez, M., & Schapachnik, F. (2017). A tool for introducing computer science with automatic formative assessment. *IEEE Transactions on Learning Technologies*, PP, 1. <https://doi.org/10.1109/TLT.2017.2682084>
- Biever, C. (2023). ChatGPT broke the Turing test — the race is on for new ways to assess AI. *Nature*, 619(7971), 686–689. <https://doi.org/10.1038/d41586-023-02361-7>

- Birk, J. P. (1992). The computer as student: An application of artificial intelligence. *Journal of Chemical Education*, 69, 294–295.
<https://api.semanticscholar.org/CorpusID:93117792>
- Block, N. (2022). The mind as the software of the brain. In *An Invitation to Cognitive Science: Visual cognition*. (pp. 377–425). MIT Press.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Handbook I: cognitive domain. *New York: David McKay*, 483–498.
- Bohlmann, M., & Berger, A. M. (2024). ChatGPT and the writing of philosophical essays: An experimental study with prospective teachers on how the Turing Test inverted. *Teaching Philosophy*, 47(2), 233–253.
- Bordt, S., & von Luxburg, U. (2023). Chatgpt participates in a computer science exam. *ArXiv Preprint ArXiv:2303.09461*.
- Boudouaia, A., Mouas, S., & Kouider, B. (2024). A study on ChatGPT-4 as an innovative approach to enhancing English as a foreign language writing learning. *Journal of Educational Computing Research*, 62(6), 1509–1537.
- Boyd, danah, & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662–679.
<https://doi.org/10.1080/1369118X.2012.678878>
- Bozkurt, A., Xiao, F., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., Farrow, R., Bond, M., Nerantzi, C., Honeychurch, S., Dron, J., Mir, K., Stewart, B., Stewart, B., Costello, E., Mason, J., Stracke, C., Romero-Hall, E., & Jandric, P. (2023). Speculative futures on ChatGPT and generative Artificial Intelligence (AI): A collective reflection from the educational landscape. *Asian Journal of Distance Education*, 18(1), 53–130.
<https://doi.org/10.5281/zenodo.7636568>
- Bradeško, L., & Mladenović, D. (2012). A survey of chatbot systems through a Loebner Prize competition. *Proceedings of Slovenian Language Technologies Society Eighth Conference of Language Technologies, C*, 34.
- Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic Analysis. In P. Liamputtong (Ed.), *Handbook of Research Methods in Health Social Sciences* (pp. 843–860). Springer Singapore. https://doi.org/10.1007/978-981-10-5251-4_103
- Breithaupt, F., Otenen, E., Wright, D. R., Kruschke, J. K., Li, Y., & Tan, Y. (2024). Humans create more novelty than ChatGPT when asked to retell a story. *Scientific Reports*, 14(1), 875. <https://doi.org/10.1038/s41598-023-50229-7>
- Bringula, R. (2024). ChatGPT in a programming course: benefits and limitations. *Frontiers in Education*, 9.
<https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2024.1248705>

- Brown, J. S., & Burton, R. R. (1978). A paradigmatic example of an artificially intelligent instructional system. *International Journal of Man-Machine Studies*, 10(3), 323–339. [https://doi.org/https://doi.org/10.1016/S0020-7373\(78\)80050-9](https://doi.org/https://doi.org/10.1016/S0020-7373(78)80050-9)
- Bryman, Alan. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative Research*, 6(1), 97–113. <https://doi.org/10.1177/1468794106058877>
- Çakmak, F. (2022). Chatbot-human interaction and its effects on EFL students' L2 speaking performance and anxiety. *Novitas-ROYAL (Research on Youth and Language)*, 16(2), 113–131.
- Campbell, M., Hoane, A. J., & Hsu, F. (2002). Deep Blue. *Artificial Intelligence*, 134(1), 57–83. [https://doi.org/https://doi.org/10.1016/S0004-3702\(01\)00129-1](https://doi.org/https://doi.org/10.1016/S0004-3702(01)00129-1)
- Cao, J., & Song, Z. (2025). An incoming threat: the influence of automation potential on job insecurity. *Asia-Pacific Journal of Business Administration*, 17(1), 116–135. <https://doi.org/10.1108/APJBA-07-2022-0328>
- Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P., & Sun, L. (2023). A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. <https://doi.org/10.48550/arXiv.2303.04226>
- Carbonell, J. R. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems*, 11(4), 190–202. <https://doi.org/10.1109/TMMS.1970.299942>
- Carstens, K. J., Mallon, J. M., Bataineh, M., & Al-Bataineh, A. (2021). Effects of technology on student learning. *Turkish Online Journal of Educational Technology-TOJET*, 20(1), 105–113.
- Casey, D. (2024). ChatGPT in public policy teaching and assessment: An examination of opportunities and challenges. *Australian Journal of Public Administration*. <https://doi.org/10.1111/1467-8500.12647>
- Çelik, F., Ersanlı, C., & Arslanbay, G. (2024). Does AI simplification of authentic blog texts improve reading comprehension, inferencing, and anxiety? A one-shot intervention in Turkish EFL context. *The International Review of Research in Open and Distributed Learning*, 25, 287–303. <https://doi.org/10.19173/irrodl.v25i3.7779>
- Černý, M. (2022). The history of chatbots: the journey from psychological experiment to educational object. *Journal of Applied Technical and Educational Sciences*, 12(3), 322.
- Chang, C.-Y., & Tang, K.-Y. (2024). Roles and research trends of ChatGPT-based learning: A bibliometric analysis and systematic review. *Educational Technology & Society*.

- Chang, W.-L., & Sun, J. C.-Y. (2024). Evaluating AI's impact on self-regulated language learning: A systematic review. *System*, 126, 103484. <https://doi.org/https://doi.org/10.1016/j.system.2024.103484>
- Chen, D., Amos, J., Shiyu, S., & and Kaptur, B. (2023). Use of technology-based assessments: a systematic review covering over 30 countries. *Assessment in Education: Principles, Policy & Practice*, 30(5–6), 396–428. <https://doi.org/10.1080/0969594X.2023.2270181>
- Chen, L., & Kurtz, B. L. (1989). XTRA-TE: Using natural language processing software to develop an ITS for language learning. *Fourth International Conference on Artificial Intelligence and Education*, 54–70.
- Chen, R., & Zhao, H. (2024). ChatGPT in creative writing courses in Chinese universities: Application and research. *2024 12th International Conference on Information and Education Technology (ICIET)*, 243–247. <https://doi.org/10.1109/ICIET60671.2024.10542824>
- Chen, X., Xie, H., & Hwang, G.-J. (2020). A multi-perspective study on Artificial Intelligence in Education: grants, conferences, journals, software tools, institutions, and researchers. *Computers and Education: Artificial Intelligence*, 1, 100005. <https://doi.org/10.1016/j.caeai.2020.100005>
- Choi, J. H., Hickman, K. E., Monahan, A. B., & Schwarcz, D. (2021). ChatGPT goes to law school. *J. Legal Educ.*, 71, 387.
- Clancey, W. J. (1979). Tutoring rules for guiding a case method dialogue. *International Journal of Man-Machine Studies*, 11(1), 25–49. [https://doi.org/https://doi.org/10.1016/S0020-7373\(79\)80004-8](https://doi.org/https://doi.org/10.1016/S0020-7373(79)80004-8)
- Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology*, 12(3), 297–298. <https://doi.org/10.1080/17439760.2016.1262613>
- Cohen-Azaria, Yael, & Zamir, Sara. (2021). “Exemplary” teachers: The perception of prospective teachers of role model teachers and their contribution to their professional identity. *Journal of Education*, 203(3), 630–638. <https://doi.org/10.1177/00220574211046555>
- Coleman, D. W. (1987). Comments on John Higgins’s “Artificial unintelligence: Computer uses in language learning”. A reader reacts: Let’s put the artificial back into artificial intelligence. *TESOL Quarterly*, 21(4), 780–784. <https://doi.org/10.2307/3586997>
- Commentaries on the ten most highly cited psychometrika articles from 1936 to the present. (2016). *Psychometrika*, 81(4), 1177–1211. <https://doi.org/10.1007/s11336-016-9540-y>
- Coniam, D. (2008). Evaluating the language resources of chatbots for their potential in English as a second language. *ReCALL*, 20, 98–116. <https://doi.org/10.1017/S0958344008000815>
- Cotton, D. R. E., Peter A., C., & and Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education*

- and *Teaching International*, 61(2), 228–239.
<https://doi.org/10.1080/14703297.2023.2190148>
- Creswell, J., Clark, V., Gutmann, M., & Hanson, W. (2003). Advanced mixed methods research designs. In *Handbook of mixed methods in social and behavioral research* (pp. 209–240).
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Creswell John, W., & Creswell, J. D. (2009). Research design: Qualitative, quantitative, and mixed methods approaches. *Los Angeles London New Dehli Singapore: Sage Publications*.
- Croes, Emmelyn A J, & Antheunis, Marjolijn L. (2020). Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships*, 38(1), 279–300. <https://doi.org/10.1177/0265407520959463>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Cruz, K., Noa, S., Turpo Gebera, O., Montesinos-Valencia, C., Bazán-Velasquez, S., & Pérez-Postigo, G. (2023). Use of gamification in English learning in Higher Education: A systematic review. *Journal of Technology and Science Education*, 13, 480. <https://doi.org/10.3926/jotse.1740>
- Cumming, G. (2003). Artificial intelligence in education: An exploration. *Journal of Computer Assisted Learning*, 14, 251–259. <https://doi.org/10.1046/j.1365-2729.1998.1440251.x>
- Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, 22, 811–817. <https://doi.org/10.1017/S1351324916000243>
- Dancey, C. P., & Reidy, J. (2007). Correlational analysis: Pearson's r. *Statistics without Maths for Psychology*, 176–185.
- Dascalu, M., Jacovina, M. E., Soto, C. M., Allen, L. K., Dai, J., Guerrero, T. A., & McNamara, D. S. (2017). Teaching iSTART to understand Spanish. *Artificial Intelligence in Education: 18th International Conference, AIED 2017, Wuhan, China, June 28–July 1, 2017, Proceedings 18*, 485–489.
- Davis, F. D. (1989). Perceived usefulness, Perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- de Winter, J. C. F. (2024). Can ChatGPT pass high school exams on English language comprehension? *International Journal of Artificial Intelligence in Education*, 34(3), 915–930. <https://doi.org/10.1007/s40593-023-00372-z>
- Derakhshan, A., & Ghiasvand, F. (2024). Is ChatGPT an evil or an angel for second language education and research? A phenomenographic study of

- research-active EFL teachers' perceptions. *International Journal of Applied Linguistics*, 34, 1246–1264. <https://doi.org/10.1111/ijal.12561>
- Deryugina, O. (2010). Chatterbots. *Scientific and Technical Information Processing*, 37, 143–147. <https://doi.org/10.3103/S0147688210020097>
- Dillon, A. (2024). The use of large language model tools such as ChatGPT in academic writing in English medium education postgraduate programs: A grounded theory approach. *Journal of Educators Online*. <https://api.semanticscholar.org/CorpusID:268755438>
- Dodds, B. L., & Lefler, R. W. (1946). Influence of science and technology on education. *Review of Educational Research*, 16(1), 30–38. <https://doi.org/10.2307/1168816>
- Drigas, A., & Dourou, A. (2013). A review on ICTs, e-learning and artificial intelligence for dyslexic's assistance. *International Journal of Emerging Technologies in Learning (IJET)*, 8, 63–67. <https://doi.org/10.3991/ijet.v8i4.2980>
- Drigas, A., & Ioannidou, R.-E. (2012). Artificial intelligence in special education: A decade review. *International Journal of Engineering Education*.
- 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/https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Dwyer, C. A. (1999). Using emerging technologies to construct effective learning environments. *Educational Media International*, 36(4), 300–309. <https://doi.org/10.1080/0952398990360409>
- Eisenstein, M., Shuller, S., & Bodman, J. (1987). Learning english with an invisible teacher: An experimental video approach. *System*, 15(2), 209–216. [https://doi.org/https://doi.org/10.1016/0346-251X\(87\)90069-8](https://doi.org/https://doi.org/10.1016/0346-251X(87)90069-8)
- El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, 38. <https://doi.org/10.1111/exsy.12667>
- Epstein, R. (1992). The quest for the thinking computer. *AI Magazine*, 13, 81–95. https://doi.org/10.1007/978-1-4020-6710-5_1
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1), 57. <https://doi.org/10.1186/s41239-023-00425-2>

- Farghaly, A. (1989). A model for intelligent computer assisted language instruction (MICALI). *Computers and the Humanities*, 23(3), 235–250.
<https://doi.org/10.1007/BF00056146>
- Farrelly, T., & Baker, N. (2023). Generative artificial intelligence: Implications and considerations for higher education practice. *Education Sciences*, 13, 1109. <https://doi.org/10.3390/educsci13111109>
- Feurzeig, W., & Lukas, G. (1972). LOGO—a programming language for teaching mathematics. *Educational Technology*, 12(3), 39–46.
<http://www.jstor.org/stable/44417811>
- Field, A. (2009). *Discovering Statistics Using SPSS: Introducing Statistical Method (3rd ed.)*. Sage.
- Fiialka, S., Kornieva, Z., & Honcharuk, T. (2024). The use of ChatGPT in creative writing assistance. *XLinguae*, 17, 3–19.
<https://doi.org/10.18355/XL.2024.17.01.01>
- Finn, J. D. (1960). Automation and education: III. Technology and the instructional process. *Audiovisual Communication Review*, 8(1), 5–26.
<https://doi.org/10.1007/BF02713371>
- Fryer, L. (2006). Bots for language learning. *The Language Teacher*, 30(8), 33–34.
- Fum, D., Giangrandi, P., & Tasso, C. (1989). Tense generation in an intelligent tutor for foreign language teaching: Some issues in the design of the verb expert. *Fourth Conference of the European Chapter of the Association for Computational Linguistics*.
- Gallagher, J. P. (1981). The effectiveness of man-machine tutorial dialogues for teaching attribute blocks problem-solving skills with an Artificial Intelligence CAI System. *Instructional Science*, 10(4), 297–332.
<https://doi.org/10.1007/BF00162731>
- Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *Npj Digital Medicine*, 6(1), 75. <https://doi.org/10.1038/s41746-023-00819-6>
- Gasiba, T. E., Oguzhan, K., Kessba, I., Lechner, U., & Pinto-Albuquerque, M. (2023). I’m sorry Dave, I’m afraid I can’t fix your code: On ChatGPT, cybersecurity, and secure coding. *International Computer Programming Education Conference*. <https://api.semanticscholar.org/CorpusID:260777564>
- Ge, X., Yin, Y., & Feng, S. (2018). Application research of computer artificial intelligence in college student sports autonomous learning. *Educational Sciences: Theory & Practice*, 18. <https://doi.org/10.12738/estp.2018.5.114>
- Geckin, V., Kızıldaş, E., & Çınar, Ç. (2023). Assessing second-language academic writing: AI vs. Human raters. *Journal of Educational Technology and Online Learning*, 6. <https://doi.org/10.31681/jetol.1336599>

- George, A. S., George, A. S. H., & Martin, A. S. G. (2023). ChatGPT and the future of work: a comprehensive analysis of AI'S impact on jobs and employment. *Partners Universal International Innovation Journal*, 1(3), 154–186.
- Gilson, A., Safraneck, C., Huang, T., Socrates, V., Chi, L., Taylor, R., & Chartash, D. (2023). How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9, e45312. <https://doi.org/10.2196/45312>
- Gozali, I., Wijaya, A. R. T., Lie, A., Cahyono, B. Y., & Suryati, N. (2024). ChatGPT as an Automated Writing Evaluation (AWE) tool: Feedback literacy development and AWE tools' integration framework. *JALT CALL Journal*, 20(1), n1.
- Greene, Jennifer, & McClintock, Charles. (1985). Triangulation in evaluation: design and analysis issues. *Evaluation Review*, 9(5), 523–545. <https://doi.org/10.1177/0193841X8500900501>
- Gu, H., Yao, J., Bai, P., Zhou, L., Cheung, A. C. K., & Abrami, P. C. (2021). Does Abracadabra help improve the English reading ability of Chinese elementary school students? A quasi-natural experimental study. *Science Insights Education Frontiers*, 9(2), 1221–1240.
- Gunderson, K. (1964). The imitation game. *Mind*, 73(290), 234–245.
- Güner, H., Er, E., Akçapinar, G., & Khalil, M. (2024). Students' attitudes and perspectives on use of ChatGPT in educational settings. *Educational Technology & Society*, 27(2), 386–404. <https://www.jstor.org/stable/48766180>
- Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435–8463.
- Guo, Y., & Lee, D. (2023). Leveraging ChatGPT for enhancing critical thinking skills. *Journal of Chemical Education*, 100(12), 4876–4883. <https://doi.org/10.1021/acs.jchemed.3c00505>
- Haenlein, Michael, & Kaplan, Andreas. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, 61(4), 5–14. <https://doi.org/10.1177/0008125619864925>
- Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 2(4), 100089. <https://doi.org/https://doi.org/10.1016/j.tbench.2023.100089>
- Hapsari, I. P., & Wu, T.-T. (2022). AI chatbots learning model in English speaking skill: Alleviating speaking anxiety, boosting enjoyment, and fostering critical

- thinking. In Y.-M. Huang, S.-C. Cheng, J. Barroso, & F. E. Sandnes (Eds.), *Innovative Technologies and Learning* (pp. 444–453). Springer International Publishing.
- Haque, M. (2023). A brief analysis of “ChatGPT” – A revolutionary tool designed by OpenAI. *EAI Endorsed Transactions on AI and Robotics*, e15. <https://doi.org/10.4108/airo.v1i1.2983>
- Hays, L., Jurkowski, O., & Sims, S. K. (2024). ChatGPT in K-12 education. *TechTrends*, 68(2), 281–294.
- Helm, M., Swiergosz, A., Haeberle, H., Karnuta, J., Schaffer, J., Krebs, V., Spitzer, A., & Ramkumar, P. (2020). Machine learning and artificial intelligence: definitions, applications, and future directions. *Current Reviews in Musculoskeletal Medicine*, 13. <https://doi.org/10.1007/s12178-020-09600-8>
- Higgins, J. (1987a). Artificial unintelligence: Computer uses in language learning. *TESOL Quarterly*, 21(1), 159–165. <https://doi.org/10.2307/3586364>
- Higgins, J. (1987b). Comments on John Higgins’s “Artificial unintelligence: Computer uses in language learning”. The author responds. *TESOL Quarterly*, 21(4), 784–785. <https://doi.org/10.2307/3586998>
- Howe, J., & O’Shea, T. (1978). Learning mathematics through LOGO. *ACM SIGCUE Outlook*, 12(1), 2–11.
- Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly*, 37(1), 81–88. <https://doi.org/10.1080/02763869.2018.1404391>
- Hsu, Mei-Hua. (2023). Mastering medical terminology with ChatGPT and Termbot. *Health Education Journal*, 83(4), 352–358. <https://doi.org/10.1177/00178969231197371>
- Hu, Y. (2025). Generative AI, communication, and stereotypes: Learning critical AI literacy through experience, analysis, and reflection. *Communication Teacher*, 39(1), 6–12. <https://doi.org/10.1080/17404622.2024.2397065>
- Huang, X. (2021). Chatbot: design, architecture, and applications. In *Univesity of Pennsylvania: School of Engineering and Applied Science, Pennsylvania* (Vol. 1). University of Pennsylvania.
- Ilic, D., Nordin, R., Glasziou, P., Tilson, J., & Villanueva, E. (2013). Implementation of a blended learning approach to teaching evidence based practice: A protocol for a mixed methods study. *BMC Medical Education*, 13, 170. <https://doi.org/10.1186/1472-6920-13-170>
- Imlah, W. G., & du Boulay, J. B. H. (1985). Robust natural language parsing in computer-assisted language instruction. *System*, 13(2), 137–147. [https://doi.org/https://doi.org/10.1016/0346-251X\(85\)90017-X](https://doi.org/https://doi.org/10.1016/0346-251X(85)90017-X)
- Jacquet, B., & Baratgin, J. (2021). Mind-reading chatbots: We are not there yet. In T. Ahram, R. Taiar, K. Langlois, & A. Choplin (Eds.), *Human Interaction*,

Emerging Technologies and Future Applications III (pp. 266–271). Springer International Publishing.

- Jamshed, M., Alam, I., Sultan, S., Abdul, S., & Corresponding. (2024). Using artificial intelligence for English language learning: Saudi EFL learners' opinions, attitudes and challenges. *Journal of Education and E-Learning Research*, 11, 135–141. <https://doi.org/10.20448/jeelr.v11i1.5397>
- Jia, J., & Chen, W. (2008). Motivate the learners to practice English through Playing with Chatbot CSIEC. In Z. Pan, X. Zhang, A. El Rhalibi, W. Woo, & Y. Li (Eds.), *Technologies for E-Learning and Digital Entertainment* (pp. 180–191). Springer Berlin Heidelberg.
- Jia, J., & Ruan, M. (2008). Use chatbot CSIEC to facilitate the individual learning in English instruction: A case study. In B. P. Woolf, E. Aïmeur, R. Nkambou, & S. Lajoie (Eds.), *Intelligent Tutoring Systems* (pp. 706–708). Springer Berlin Heidelberg.
- Jiang, Y., Wang, Q., & Weng, Z. (2022). The influence of technology in educating English language learners at-risk or with disabilities: A systematic review. *CEPS Journal*, 12(4), 53–74.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112–133.
- Johnson, R. Burke, & Onwuegbuzie, Anthony J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26. <https://doi.org/10.3102/0013189X033007014>
- Joseph, G. V., Athira, P., Thomas, M. A., Jose, D., Roy, T. V., & Prasad, M. (2024). Impact of digital literacy, use of AI tools and peer collaboration on AI assisted learning: Perceptions of the university students. *Digital Education Review*, 45, 43–49.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36. <https://doi.org/10.1007/BF02291575>
- Kaiser, Henry F, & Rice, John. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, 34(1), 111–117. <https://doi.org/10.1177/001316447403400115>
- Kaliterna, M., Žuljević, M. F., Ursić, L., Krka, J., & Duplančić, D. (2024). Testing the capacity of Bard and ChatGPT for writing essays on ethical dilemmas: A cross-sectional study. *Scientific Reports*, 14(1), 26046.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual*

- Differences*, 103, 102274.
<https://doi.org/https://doi.org/10.1016/j.lindif.2023.102274>
- Kelley, T. L. (1942). The reliability coefficient. *Psychometrika*, 7(2), 75–83.
<https://doi.org/10.1007/BF02288068>
- Kerly, A., & Bull, S. (2006). The potential for chatbots in negotiated learner modelling: A Wizard-of-Oz Study. In M. Ikeda, K. D. Ashley, & T.-W. Chan (Eds.), *Intelligent Tutoring Systems* (pp. 443–452). Springer Berlin Heidelberg.
- Kerly, A., Hall, P., & Bull, S. (2007). Bringing chatbots into education: Towards natural language negotiation of open learner models. *Knowledge-Based Systems*, 20(2), 177–185.
<https://doi.org/https://doi.org/10.1016/j.knosys.2006.11.014>
- Kim, H.-S. (2021). Is it beneficial to use AI chatbots to improve learners' speaking performance? *Journal of Asia TEFL*, 18(1), 161–178.
- Kim, S. S., & Son, J.-Y. (2009). Out of dedication or constraint? A dual model of post-adoption phenomena and its empirical test in the context of online services. *MIS Quarterly*, 49–70.
- Kim, T. (2017). Understanding one-way ANOVA using conceptual figures. *Korean Journal of Anesthesiology*, 70, 22.
<https://doi.org/10.4097/kjae.2017.70.1.22>
- Kipp, M. (2024). From GPT-3.5 to GPT-4.o: A leap in AI's medical exam performance. *Information*, 15, 543. <https://doi.org/10.3390/info15090543>
- Klimova, B., Pikhart, M., & Al-Obaydi, L. (2024). Exploring the potential of ChatGPT for foreign language education at the university level. *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2024.1269319>
- Knight, W. (2012). Social intelligence: Siri may not be the smartest AI in the world, but it's the most socially adept. *MIT Technology Review*, 115, 80–82.
- Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J., Bielaniec, J., Gruza, M., Janz, A., Kanclerz, K., Kocoń, A., Koptyra, B., Mieszczenko-Kowszewicz, W., Miłkowski, P., Oleksy, M., Piasecki, M., Radliński, Ł., Wojtasik, K., Woźniak, S., & Kazienko, P. (2023). ChatGPT: Jack of all trades, master of none. *Information Fusion*, 99, 101861.
<https://doi.org/https://doi.org/10.1016/j.inffus.2023.101861>
- Koubaa, A. (2023). GPT-4 vs. GPT-3.5: A concise showdown. *Preprints. Org*, 2023030422.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73(4), 440–464.

- Kurt, G., & Kurt, Y. (2024). Enhancing L2 writing skills: ChatGPT as an automated feedback tool. *Journal of Information Technology Education: Research*, 23, 024.
- Kurt, U., & Çayır, A. (2023). A modern Turkish poet: Fine-tuned GPT-2. *2023 8th International Conference on Computer Science and Engineering (UBMK)*, 1–5.
- Kuşcu, O., Pamuk, A. E., Sütay Süslü, N., & Hosal, S. (2023). Is ChatGPT accurate and reliable in answering questions regarding head and neck cancer? *Frontiers in Oncology*, 13, 1256459.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lee, J., & Friedman, A. M. (2009). *Research on technology in social studies education*. IAP.
- Lee, S., & Park, G. (2023). Exploring the impact of ChatGPT literacy on user satisfaction: The mediating role of user motivations. *Cyberpsychology, Behavior and Social Networking*, 26, 913–918. <https://doi.org/10.1089/cyber.2023.0312>
- Lee, S., & Park, G. (2024). Development and validation of ChatGPT literacy scale. *Current Psychology*, 43(21), 18992–19004. <https://doi.org/10.1007/s12144-024-05723-0>
- Lee, Y.-J. (2024). Can my writing be polished further? When ChatGPT meets human touch. *ELT Journal*, 78(4), 401–413. <https://doi.org/10.1093/elt/ccae039>
- Leinonen, J., Denny, P., MacNeil, S., Sarsa, S., Bernstein, S., Kim, J., Tran, A., & Hellas, A. (2023). Comparing code explanations created by students and large language models. *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, 124–130. <https://doi.org/10.1145/3587102.3588785>
- Leon, A. J., & Vidhani, D. (2023). ChatGPT needs a chemistry tutor too. *Journal of Chemical Education*, 100(10), 3859–3865. <https://doi.org/10.1021/acs.jchemed.3c00288>
- Li, X., & Thien, L. M. (2025). Exploring ChatGPT literacy as predictor and moderator in enhancing informal digital learning of English (IDLE): a partial least square analysis. *Quality Assurance in Education*, 33, 336–356. <https://doi.org/10.1108/QAE-08-2024-0152>
- Liang, J.-C., Gwo-Jen, H., Mei-Rong Alice, C., & Darmawansah, D. (2023). Roles and research foci of artificial intelligence in language education: an integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments*, 31(7), 4270–4296. <https://doi.org/10.1080/10494820.2021.1958348>

- Lim, E. M. (2023). The effects of pre-service early childhood teachers' digital literacy and self-efficacy on their perception of AI education for young children. *Education and Information Technologies*, 28(10), 12969–12995. <https://doi.org/10.1007/s10639-023-11724-6>
- 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>
- Lintner, T. (2024). A systematic review of AI literacy scales. *Npj Science of Learning*, 9(1), 50. <https://doi.org/10.1038/s41539-024-00264-4>
- Liu, G., & and Ma, C. (2024). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*, 18(2), 125–138. <https://doi.org/10.1080/17501229.2023.2240316>
- Liu, G. L., Ron, D., & and Ma, C. (2024). Exploring AI-mediated informal digital learning of English (AI-IDLE): a mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*, 1–29. <https://doi.org/10.1080/09588221.2024.2310288>
- Liu, Sylvia, Reynolds, Barry Lee, Thomas, Nathan, & Soyooof, Ali. (2024). The use of digital technologies to develop young children's language and literacy skills: A systematic review. *SAGE Open*, 14(1), 21582440241230850. <https://doi.org/10.1177/21582440241230850>
- Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: a systematic review of empirical studies. *Smart Learning Environments*, 11(1), 50. <https://doi.org/10.1186/s40561-024-00342-5>
- Lund, B., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News*, 40. <https://doi.org/10.1108/LHTN-01-2023-0009>
- Łupkowski, P. (2006). Some historical remarks on Block's "Aunt Bubbles" argument. *Minds and Machines*, 16(4), 437–441. <https://doi.org/10.1007/s11023-006-9043-1>
- Ma, C., Wang, F., & Lv, Y. (2024). Teaching effects of using bullet-screen technology during classes on students' learning: The mediating effect of perceived interactivity. *Education and Information Technologies*, 29(5), 5283–5307. <https://doi.org/10.1007/s10639-023-12054-3>
- Ma, Q., Crosthwaite, P., Sun, D., & Zou, D. (2024). Exploring ChatGPT literacy in language education: A global perspective and comprehensive approach. *Computers and Education: Artificial Intelligence*, 7, 100278. <https://doi.org/https://doi.org/10.1016/j.caeai.2024.100278>

- Mabbott, A., & Bull, S. (2006). Student preferences for editing, persuading, and negotiating the open learner model. In M. Ikeda, K. D. Ashley, & T.-W. Chan (Eds.), *Intelligent Tutoring Systems* (pp. 481–490). Springer Berlin Heidelberg.
- Mabuan, R. (2024). ChatGPT and ELT: Exploring teachers' voices. *International Journal of Technology in Education*, 7, 128–153.
<https://doi.org/10.46328/ijte.523>
- Maguire, M. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Higher Education*, 9(3), 3351.
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. *Smart Learning Environments*, 11(1), 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Marangunić, N., & Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal Access in the Information Society*, 14(1), 81–95. <https://doi.org/10.1007/s10209-014-0348-1>
- Marks, D. F., & Yardley, L. (2004). *Research Methods for Clinical and Health Psychology*. SAGE.
- Mauldin, M. L. (1994). Chatterbots, tinymuds, and the turing test entering the loebner prize competition. *Proceedings of the Twelfth AAAI National Conference on Artificial Intelligence*, 16–21.
- Mayor, A. (2018). *Myths, Machines, and Ancient Dreams of Technology*. Princeton University Press. <https://doi.org/10.2307/j.ctvc779xn>
- McCallum, L. (2024). New takes on developing intercultural communicative competence: using AI tools in telecollaboration task design and task completion. *Journal for Multicultural Education*, 18(12), 153–172.
<https://doi.org/https://doi.org/10.1108/JME-06-2023-0043>
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI Magazine*, 27(4), 12.
- McCarthy, K. S., Soto, C. M., de Blume, A. P. G., Palma, D., González, J. I., & McNamara, D. S. (2020). Improving reading comprehension in Spanish using iSTART-E: a pilot study. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 10(4), 66–82.
- McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers*, 36(2), 222–233.
<https://doi.org/10.3758/BF03195567>
- Mengual-Andrés, S., Roig-Vila, R., & Mira, J. B. (2016). Delphi study for the design and validation of a questionnaire about digital competences in higher

- education. *International Journal of Educational Technology in Higher Education*, 13(1), 12. <https://doi.org/10.1186/s41239-016-0009-y>
- Meniado, J. (2023). The impact of ChatGPT on English language teaching, learning, and assessment: A rapid review of literature. *Arab World English Journal*, 14, 3–18. <https://doi.org/10.24093/awej/vol14no4.1>
- Merç, A. (2015). Using technology in the classroom: A study with Turkish pre-service EFL teachers. *Turkish Online Journal of Educational Technology-TOJET*, 14(2), 229–240.
- Metze, K., Morandin-Reis, R., Lorand-Metze, I., & Florindo, J. (2023). Bibliographic research with ChatGPT may be misleading: The problem of hallucination. *Journal of Pediatric Surgery*, 59. <https://doi.org/10.1016/j.jpedsurg.2023.08.018>
- Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Education Sciences*, 13(9), 856.
- Minsky, M. (1961). Steps toward artificial intelligence. *Proceedings of the IRE*, 49(1), 8–30.
- 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. *Education and Information Technologies*, 29(3), 3195–3217. <https://doi.org/10.1007/s10639-023-11917-z>
- Mompean, J. A. (2024). ChatGPT for L2 pronunciation teaching and learning. *ELT Journal*, 78(4), 423–434. <https://doi.org/10.1093/elt/ccae050>
- Mone, G. (2016). The edge of the uncanny. *Commun. ACM*, 59(9), 17–19. <https://doi.org/10.1145/2967977>
- Mulcahy, C. A., Kristine A., C., & Fenty, N. S. (2024). A systematic review of research on technology use among students with high incidence disabilities. *Exceptionality*, 32(3), 149–167. <https://doi.org/10.1080/09362835.2023.2285272>
- Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, 2(2), 109–138. <https://doi.org/DOI: 10.1017/XPS.2015.19>
- Muñoz-Basols, J., Neville, C., Lafford, B., & Godev, C. (2023). Potentialities of applied translation for language learning in the era of artificial intelligence. *Hispania*, 106, 171–194. <https://doi.org/10.1353/hpn.2023.a899427>
- Mutammimah, H., Rejeki, S., Kustini, S., & Amelia, R. (2024). Understanding teachers' perspective toward ChatGPT acceptance in English Language Teaching. *International Journal of Technology in Education*, 7, 290–307. <https://doi.org/10.46328/ijte.656>

- Nagata, N. (1996). Computer vs. workbook instruction in second language acquisition. *CALICO Journal*, 14(1), 53–75.
<http://www.jstor.org/stable/24157094>
- Namaziandost, E., & Nasri, M. (2019). The impact of social media on EFL learners' speaking skill: A survey study involving EFL teachers and students. *Journal of Applied Linguistics and Language Research*, 6, 199–215.
<https://api.semanticscholar.org/CorpusID:213706812>
- Neuendorf, K. A. (2018). Content analysis and thematic analysis. In *Advanced research methods for applied psychology* (pp. 211–223). Routledge.
- Newel, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3), 113–126.
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041.
- Nghi, T., Huu Phuc, T., & Nguyen Tat, T. (2019). Applying AI chatbot for teaching a foreign language: An empirical research. *International Journal of Scientific & Technology Research*, 8, 897–902.
- Nicholson, K. J., Sherman, M., Divi, S. N., Bowles, D. R., & Vaccaro, A. R. (2022). The role of family-wise error rate in determining statistical significance. *Clinical Spine Surgery*, 35(5).
https://journals.lww.com/jspinaldisorders/fulltext/2022/06000/the_role_of_family_wise_error_rate_in_determining.7.aspx
- Noble, H., & Heale, R. (2019). Triangulation in research, with examples. *Evidence Based Nursing*, 22(3), 67. <https://doi.org/10.1136/ebnurs-2019-103145>
- Nov, O., Singh, N., & Mann, D. (2023). Putting ChatGPT's medical advice to the (Turing) Test: Survey study. *JMIR Med Educ.*, 10(9).
<https://doi.org/doi:10.2196/46939>
- Nuruzzaman, M., & Hussain, O. K. (2020). IntelliBot: A dialogue-based chatbot for the insurance industry. *Knowledge-Based Systems*, 196, 105810.
<https://doi.org/https://doi.org/10.1016/j.knosys.2020.105810>
- Nyns, R. R. (1989). Is intelligent computer-assisted language learning possible? *System*, 17(1), 35–47. [https://doi.org/https://doi.org/10.1016/0346-251X\(89\)90058-4](https://doi.org/https://doi.org/10.1016/0346-251X(89)90058-4)
- Obermeier, K. K. (1989). *Natural language processing technologies in artificial intelligence: The science and industry perspective*. Prentice-Hall, Inc.
- Oflaz, A. (2019). The effects of anxiety, shyness and language learning strategies on speaking skills and academic achievement. *European Journal of Educational Research*, 8(4), 999–1011. <https://doi.org/10.12973/eu-jer.8.4.999>

- OpenAI. (2025). *ChatGPT (gpt-4-turbo-2024-04-09 version)*.
<https://chat.openai.com/chat>.
- Ouh, E. L., Gan, B. K. S., Jin Shim, K., & Wlodkowski, S. (2023). ChatGPT, can you generate solutions for my coding exercises? An evaluation on its effectiveness in an undergraduate Java programming course. *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, 54–60. <https://doi.org/10.1145/3587102.3588794>
- Ozfidan, B., El-Dakhs, D. A. S., & Alsalim, L. A. (2024). The use of AI tools in English academic writing by Saudi undergraduates. *Contemporary Educational Technology*, 16(4).
- Ozfidan, B., Fayez, O., & Ismail, H. (2021). Student perspectives of online teaching and learning during the COVID-19 pandemic. *Online Learning*, 25(4), 461–485.
- Pardeshi, S., Ovhal, S., Shinde, P., Bansode, M., & Birajdar, A. (2020). A survey on Different Algorithms used in Chatbot. *International Research Journal of Engineering and Technology (IRJET)*, 7(5), 6092–6098.
- Paul, G., & Blau, I. (2009). Online teaching effectiveness: A tale of two instructors. *International Review of Research in Open and Distance Learning*, 10. <https://doi.org/10.19173/irrodl.v10i3.712>
- Phosa, S. (2024). The study of EFL students' attitudes and behaviors toward using ChatGPT in English language learning:(in progress). *Journal of English Language and Linguistics*, 5(3), 332–344.
- Pillare, T., Chaoudhari, M., Hiwrale, V., Ade, S., Dudhe, P., & Chinte, S. (2023). A survey paper on chatbot. *IRJMETs*, 5, 10.
- Plengkham, B., & Wasanasomsithi, P. (2023). Effects of integrated performance assessment modules on English speaking ability of Thai EFL undergraduate students. *LEARN Journal: Language Education and Acquisition Research Network*, 16(1), 448–472.
- Pokrivcakova, S. (2024). Pre-service teachers' attitudes towards artificial intelligence and its integration into EFL teaching and learning. *Journal of Language and Cultural Education*, 11, 100–114.
<https://doi.org/10.2478/jolace-2023-0031>
- Pratiwi, A., Azizah, N., Baharudin, B., Saputra, M., & Mustofa, M. (2025). Student learning outcomes: Innovative learning with an experimentation blended learning model. *Jurnal Pendidikan Dan Pembelajaran Indonesia (JPPI)*, 5, 63–74. <https://doi.org/10.53299/jppi.v5i1.938>
- Priest, A. G. (1981). Artificial intelligence and learning: Conference reports. *Instructional Science*, 10(3), 277–285. <http://www.jstor.org/stable/23368366>
- Qi, W., Pan, J., Lyu, H., & Luo, J. (2024). Excitements and concerns in the post-ChatGPT era: Deciphering public perception of AI through social media

- analysis. *Telematics and Informatics*, 92, 102158.
<https://doi.org/https://doi.org/10.1016/j.tele.2024.102158>
- Qualter, T. H. (1959). THE Venn Diagrams. *Higher Education Quarterly*, 13(2), 196–199.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.
- Raiola, R. (2023). ChatGPT, can you tell me a story? An exercise in challenging the true creativity of generative AI. *Communications of the ACM*, 66(5), 104–ff.
- Rajaraman, V. (2023). From ELIZA to ChatGPT. *Resonance*, 28(6), 889–905.
<https://doi.org/10.1007/s12045-023-1620-6>
- Ramadhani, R., Aulawi, H., & Ulfa, R. L. (2023). Readability of reading texts as authentic materials issued by ChatGPT: A systemic functional perspective. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 8(2), 149–168.
- Redican, K., Matteo, G., & Zizzamia, B. (2025). Assessing ChatGPT for GIS education and assignment creation. *Journal of Geography in Higher Education*, 49(1), 113–129. <https://doi.org/10.1080/03098265.2024.2397332>
- Rodosthenous, C., Lyding, V., Sangati, F., König, A., ul Hassan, U., Nicolas, L., Horbacauskiene, J., Katinskaia, A., & Aparaschivei, L. (2020). Using crowdsourced exercises for vocabulary training to expand ConceptNet.
- Rodríguez, A. C. (2024). Evaluation of Artificial Intelligence tools (ChatGPT and Google Bard) in higher education for the history of America: answering questions and crafting assignments with a gender perspective. *Building Knowledge: Visions from Education and the Humanities*, 49–68.
- Roh, Y., Heo, G., & Whang, S. E. (2021). A survey on data collection for Machine learning: A big data - AI integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, 33(4), 1328–1347.
<https://doi.org/10.1109/TKDE.2019.2946162>
- Roscoe, R. D., Allen, L. K., Weston, J. L., Crossley, S. A., & McNamara, D. S. (2014). The writing pal intelligent tutoring system: Usability testing and development. *Computers and Composition*, 34, 39–59.
<https://doi.org/https://doi.org/10.1016/j.compcom.2014.09.002>
- Rosenman, R., Tennekoon, V., & Hill, L. (2011). Measuring bias in self-reported data. *Int. J. of Behavioural and Healthcare Research*, 2, 320–332.
<https://doi.org/10.1504/IJBHR.2011.043414>
- Ross, A., & Willson, V. L. (2017). Independent Samples T-Test. In A. Ross & V. L. Willson (Eds.), *Basic and Advanced Statistical Tests: Writing Results Sections and Creating Tables and Figures* (pp. 13–16). SensePublishers.
https://doi.org/10.1007/978-94-6351-086-8_3

- Rossi, C., & Russo, F. (2017). Automata (Towards Automation and Robots). In C. Rossi & F. Russo (Eds.), *Ancient Engineers' Inventions: Precursors of the Present* (pp. 353–380). Springer International Publishing.
https://doi.org/10.1007/978-3-319-44476-5_21
- Ruan, S., Willis, A., Xu, Q., Davis, G. M., Jiang, L., Brunskill, E., & Landay, J. A. (2019). Bookbuddy: Turning digital materials into interactive foreign language lessons through a voice chatbot. *Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale*, 1–4.
- Saarna, C. (2024). Identifying whether a short essay was written by a university student or ChatGPT. *International Journal of Technology in Education*, 7(3), 611–633.
- Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E. L., Ghasemipour, K., Gontijo Lopes, R., Karagol Ayan, B., & Salimans, T. (2022). Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems*, 35, 36479–36494.
- Sahin, M. C., Sozer, A., Kuzucu, P., Turkmen, T., Sahin, M. B., Sozer, E., Tufek, O. Y., Nernekli, K., Emmez, H., & Celtikci, E. (2024). Beyond human in neurosurgical exams: ChatGPT's success in the Turkish neurosurgical society proficiency board exams. *Computers in Biology and Medicine*, 169, 107807.
<https://doi.org/https://doi.org/10.1016/j.compbiomed.2023.107807>
- Salaberry, M. (1996). A theoretical foundation for the development of pedagogical tasks in computer mediated communication. *CALICO Journal*, 14.
<https://doi.org/10.1558/cj.v14i1.5-34>
- Samarakou, M., Papadakis, A., Fylladitakis, E., Hatziapostolou, A., Tsaganou, G., & Früh, W.-G. (2014). An open learning environment for the diagnosis, assistance and evaluation of students based on artificial intelligence. *International Journal of Emerging Technologies in Learning (IJET)*, 9, 36–44. <https://doi.org/10.3991/ijet.v9i3.3367>
- Samir Abdelhafiz, A., Ali, A., Maaly, A., Ziady, H., Sultan, E., & Mahgoub, M. (2024). Knowledge, perceptions and attitude of researchers towards using ChatGPT in research. *Journal of Medical Systems*, 48.
<https://doi.org/10.1007/s10916-024-02044-4>
- Sayed, B. T., Bani Younes, Z. B., Alkhayyat, A., Adhamova, I., & Teferi, H. (2024). To be with artificial intelligence in oral test or not to be: a probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy. *Language Testing in Asia*, 14(1), 49.
<https://doi.org/10.1186/s40468-024-00321-0>
- Saygin, A. P. (1999). *Turing Test and Conversation*. Bilkent University.
- Saygin, A. P., Cicekli, I., & Akman, V. (2000). Turing Test: 50 years later. *Minds and Machines*, 10(4), 463–518. <https://doi.org/10.1023/A:1011288000451>

- Schaeffer, J., Björnsson, Y., Kishimoto, A., Müller, M., Lake, R., Lu, P., & Sutphen, S. (2007). Checkers Is Solved. *Science*, 317, 1518–1522. <https://doi.org/10.1126/science.1144079>
- Scheinkman, R., Tordjman, L., Sharifi, S., & Nouri, K. (2025). The ethical considerations of artificial intelligence hallucination and misinformation in dermatological and medical laser documentation. *Lasers in Medical Science*, 40(1), 110. <https://doi.org/10.1007/s10103-025-04364-4>
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424. <https://doi.org/DOI: 10.1017/S0140525X00005756>
- Self, J. A. (1974). Student models in computer-aided instruction. *International Journal of Man-Machine Studies*, 6(2), 261–276. [https://doi.org/https://doi.org/10.1016/S0020-7373\(74\)80005-2](https://doi.org/https://doi.org/10.1016/S0020-7373(74)80005-2)
- Shabara, R., ElEbyary, K., & Boraie, D. (2024). Teachers or ChatGPT: The issue of accuracy and consistency in L2 assessment. *Teaching English with Technology*, 24(2), 71–92.
- Shahriar, S., Lund, B. D., Mannuru, N. R., Arshad, M. A., Hayawi, K., Bevara, R. V. K., Mannuru, A., & Batool, L. (2024). Putting gpt-4o to the sword: A comprehensive evaluation of language, vision, speech, and multimodal proficiency. *Applied Sciences*, 14(17), 7782.
- Shaikh, S., Yildirim Yayilgan, S., Klimova, B., & Pikhart, M. (2023). Assessing the usability of ChatGPT for formal English language learning. *European Journal of Investigation in Health, Psychology and Education*, 13, 1937–1960. <https://doi.org/10.3390/ejihpe13090140>
- Sharif, K., & Tenbergen, B. (2020). Smart home voice assistants: A literature survey of user privacy and security vulnerabilities. *Complex Systems Informatics and Modeling Quarterly*, 0, 15–30. <https://doi.org/10.7250/csimq.2020-24.02>
- Shen, Y., Heacock, L., Elias, J., Hentel, K., Reig, B., Shih, G., & Moy, L. (2023). ChatGPT and other large language models are double-edged swords. *Radiology*, 307. <https://doi.org/10.1148/radiol.230163>
- Shimray, Y. P. (2023). The effectiveness of Online Speaking Tasks (OST) on English communication ability of Thai EFL undergraduates. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 8(2), 25–40.
- Shin, D., Kim, H., Lee, J. H., & Yang, H. (2021). Exploring the use of an artificial intelligence chatbot as second language conversation partners. *Korean Journal of English Language and Linguistics*, 21(21), 375–391.
- Shortliffe, E. (1976). Computer-based medical consultations: MYCIN. *Artificial Intelligence - AI*, 388. <https://doi.org/10.1097/00004669-197610000-00011>
- Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. *IEEE Access*, 11, 38805–38818. <https://doi.org/10.1109/ACCESS.2023.3268224>

- Singh, V., & Singh, A. (2023). Revolutionizing the hospitality industry: how ChatGPT empowers future hoteliers (pp. 192–203). <https://doi.org/10.4018/979-8-3693-0815-8.ch011>
- Sleeman, D. H., & Brown, J. S. (1979). Intelligent tutoring systems. In *International Journal of Man-Machine Studies* (Vol. 11, Issue 1, pp. 1–3). Academic Press.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72–101. <https://doi.org/10.2307/1412159>
- Sumbal, A., Sumbal, R., & Amir, A. (2024). Can ChatGPT-3.5 pass a medical exam? A systematic review of ChatGPT’s performance in academic testing. *Journal of Medical Education and Curricular Development*, 11. <https://doi.org/10.1177/23821205241238641>
- Taecharungroj, V. (2023). “What can ChatGPT do?” Analyzing early reactions to the innovative AI chatbot on Twitter. *Big Data and Cognitive Computing (BDCC)*, 7(1). <https://doi.org/10.3390/bdcc7010035>
- Tang, K.-Y., Ching-Yi, C., & and Hwang, G.-J. (2023). Trends in artificial intelligence-supported e-learning: a systematic review and co-citation network analysis (1998–2019). *Interactive Learning Environments*, 31(4), 2134–2152. <https://doi.org/10.1080/10494820.2021.1875001>
- Tatar, Ç. (2015). *Structural equation modeling (SEM) and its applications*. Dokuz Eylül University.
- Tercanlioglu, L. (2004). Exploring gender effect on adult foreign language learning strategies. *Issues in Educational Research*, 14, 181–193.
- Tseng, Y.-C., & Lin, Y.-H. (2024). Enhancing English as a Foreign Language (EFL) learners’ writing with ChatGPT: A university-level course design. *Electronic Journal of E-Learning*, 22, 0. <https://doi.org/10.34190/ejel.21.5.3329>
- Tüfekçi, A., & Köse, U. (2013). Development of an artificial intelligence based software system on teaching computer programming and evaluation of the system. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 28(2), 469–481.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59, 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- Tyagi, A. K. (2024). Transformative effects of ChatGPT on the modern era of education and society: From society’s and industry’s perspectives. In *Machine Learning Algorithms Using Scikit and TensorFlow Environments* (pp. 374–387). IGI Global Scientific Publishing.
- Ulla, M. B., William F., P., & and Busbus, S. O. (2023). ‘To generate or stop generating response’: Exploring EFL teachers’ perspectives on ChatGPT in English language teaching in Thailand. *Learning: Research and Practice*, 9(2), 168–182. <https://doi.org/10.1080/23735082.2023.2257252>

- Ulum, Ö. (2024). Unveiling the layers: Analyzing ChatGPT implementations in Turkish state universities. *Base for Electronic Educational Sciences*, 5, 114–134. <https://doi.org/10.29329/bedu.2024.651.7>
- Üstünbaş, Ü. (2024). Hey, GPT, can we have a chat?: A case study on EFL learners' AI speaking practice. *International Journal of Modern Education Studies*, 8. <https://doi.org/10.51383/ijonmes.2024.318>
- Vo, T., & Nguyễn, H. (2024). Generative artificial intelligence and ChatGPT in language learning: EFL students' perceptions of technology acceptance. *Journal of University Teaching and Learning Practice*, 21. <https://doi.org/10.53761/fr1rkj58>
- Vygotsky, L. S., Cole, M., Jolm-Steiner, V., Scribner, S., & Souberman, E. (1978). *Mind in Society: Development of Higher Psychological Processes* (M. Cole, V. Jolm-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Waisberg, E., Ong, J., Masalkhi, M., Zaman, N., Sarker, P., Lee, A. G., & Tavakkoli, A. (2024). Google's AI chatbot "Bard": a side-by-side comparison with ChatGPT and its utilization in ophthalmology. *Eye*, 38(4), 642–645. <https://doi.org/10.1038/s41433-023-02760-0>
- Wallace, R. S. (2009). The Anatomy of A.L.I.C.E. In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the Turing Test: Philosophical and methodological issues in the quest for the thinking computer* (pp. 181–210). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6710-5_13
- Waltzer, T., Cox, R., & Heyman, G. (2023). Testing the ability of teachers and students to differentiate between essays generated by ChatGPT and high school students. *Human Behavior and Emerging Technologies*, 2023, 1–9. <https://doi.org/10.1155/2023/1923981>
- Wang, J. (2024). The effect of Chinese EFL students' digital literacy on their technostress and academic productivity. *The Asia-Pacific Education Researcher*, 33(4), 987–996. <https://doi.org/10.1007/s40299-023-00794-2>
- Wang, X., Zhong, Y., Huang, C., & Huang, X. (2024). ChatPRCS: A personalized support system for English reading comprehension based on ChatGPT. *IEEE Trans. Learn. Technol.*, 17, 1762–1776. <https://doi.org/10.1109/TLT.2024.3405747>
- Wang, Y. F., Petrina, S., & Feng, F. (2017). VILLAGE—Virtual Immersive Language Learning and Gaming Environment: Immersion and presence. *British Journal of Educational Technology*, 48(2), 431–450. <https://doi.org/https://doi.org/10.1111/bjet.12388>
- Ware, P. D., & Kramsch, C. (2005). Toward an intercultural stance: Teaching German and English through telecollaboration. *The Modern Language Journal*, 89(2), 190–205. <http://www.jstor.org/stable/3588680>

- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45.
- Wells, F. L. (1907). A statistical study of literary merit.(Columbia Univ. Cont. to Phil. & Psych., 16, 3.). *Archives of Psychology*.
- Wenger, E. (1987). Chapter 3 - Tutorial dialogues: from semantic nets to mental models. In E. Wenger (Ed.), *Artificial Intelligence and Tutoring Systems* (pp. 29–50). Morgan Kaufmann. <https://doi.org/https://doi.org/10.1016/B978-0-934613-26-2.50010-4>
- Wilson, K., & Bates, M. (1981). Artificial intelligence in computer-based language instruction. *The Volta Review*, 83(5), 336–349.
- Wissler, C. (1905). The Spearman correlation formula. *Science*, 22(558), 309–311. <https://doi.org/10.1126/science.22.558.309>
- Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, 10(5), 1122–1136.
- Xing, R. (2023). Advancements in English listening education: ChatGPT and convolutional neural network integration. *Journal of Pedagogical Research*. <https://doi.org/10.33902/JPR.202323980>
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28(11), 13943–13967. <https://doi.org/10.1007/s10639-023-11742-4>
- 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/DOI: 10.1017/S0958344022000039>
- Yang, Y. (2024). The reliability of using ChatGPT in rating EFL writings. *Shanlax International Journal of Education*, 12, 49–59. <https://doi.org/10.34293/education.v12i4.7855>
- Yazdani, M., & Uren, J. (1988). Generalising language-tutoring systems: A French/Spanish case study, using LINGER. *Instructional Science*, 17(2), 179–188. <https://doi.org/10.1007/BF00052705>
- Yıldırım Taştı, Ö., & Engin, C. (2024). Teaching in neoliberal times: Unraveling teacher voices at basic high schools in Türkiye. *International Journal of Curriculum and Instructional Studies*, 14(1), 83–108. <https://doi.org/10.31704/ijocis.1466226>
- Yıldız, C. (2024). ChatGPT Integration in EFL education: A path to enhanced speaking self-efficacy. *Novitas-Royal*, 18(2), 167–182. <https://doi.org/10.5281/zenodo.13861137>
- Yılmaz, C. (2010). The relationship between language learning strategies, gender, proficiency and self-efficacy beliefs: a study of ELT learners in Turkey.

- Procedia - Social and Behavioral Sciences*, 2(2), 682–687.
<https://doi.org/https://doi.org/10.1016/j.sbspro.2010.03.084>
- Yoder, P. J., Jennifer Urbano, B., Niels G., W., & and Kim, G. (2004). Enhancing power while controlling family-wise error: An illustration of the issues using electrocortical studies. *Journal of Clinical and Experimental Neuropsychology*, 26(3), 320–331.
<https://doi.org/10.1080/13803390490510040>
- Yu, C., Yan, J., & Cai, N. (2024). ChatGPT in higher education: factors influencing ChatGPT user satisfaction and continued use intention. *Frontiers in Education*, 9, 1354929.
- Yu, Z., Liheng, Y., Qiaoyun, X., Wei, X., & and Wu, P. (2022). Effects of mobile learning technologies and social media tools on student engagement and learning outcomes of English learning. *Technology, Pedagogy and Education*, 31(3), 381–398. <https://doi.org/10.1080/1475939X.2022.2045215>
- Yüksel, D., & Tanrıverdi, B. (2009). Effects of watching captioned movie clip on vocabulary development of efl learners. *Turkish Online Journal of Educational Technology*, 8, 48–54.
- Zdravkova, K. (2000). Conceptual framework for an intelligent chatterbot. *ITI 2000. Proceedings of the 22nd International Conference on Information Technology Interfaces (Cat. No. 00EX411)*, 189–194.
- Zeevy-Solovey, O. (2024). Comparing peer, ChatGPT, and teacher corrective feedback in EFL writing: Students’ perceptions and preferences. *Technology in Language Teaching & Learning*, 6, 1–23.
<https://doi.org/10.29140/tltl.v6n3.1482>
- Zemčík, M. (2019). A brief history of chatbots. *DEStech Transactions on Computer Science and Engineering*, 10, 14–18.
- Zha, S., Chen, L. K., Hung, W., Gong, N., Moore, P., & Klemetsrud, B. (2025). Predicting students’ interest from small group conversational characteristics: Insights from an AI literacy education with high school students. *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 2*, 1677–1678.
- Zhang, D., Zhao, J. L., Zhou, L., & Nunamaker Jr, J. F. (2004). Can e-learning replace classroom learning? *Communications of the ACM*, 47(5), 75–79.
- Zhang, K., Li, X., Ye, H., Lin, K. J., Chen, S., & Law, R. (2024). How can ChatGPT assist hospitality and tourism education? *Journal of Teaching in Travel & Tourism*, 24(4), 438–448.
- Zheng, Q., Yuan, Z., & Pan, X. (2025). Examining the influencing effect of EFL students’ digital literacy on their online learning power: the mediating role of perceived teacher support. *Asia Pacific Journal of Education*, 45(1), 20–34.

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.

Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., Zhang, K., Ji, C., Yan, Q., He, L., Peng, H., Li, J., Wu, J., Liu, Z., Xie, P., Xiong, C., Pei, J., Yu, P. S., & Sun, L. (2024). A comprehensive survey on pretrained foundation models: a history from BERT to ChatGPT. *International Journal of Machine Learning and Cybernetics*. <https://doi.org/10.1007/s13042-024-02443-6>



APPENDICES

Appendix I. Survey 1 – Adapted ChatGPT Literacy Scale

ChatGPT Literacy Survey	Very Much	Considerably	Neutral	Slightly	Not at all
Factor 1: Technical Proficiency (TP)					
1. I can train and fine-tune (adjust prompts according to your needs) ChatGPT for specific purposes or applications.	5	4	3	2	1
2. I have the ability to identify and solve technical problems of ChatGPT.	5	4	3	2	1
3. I have the ability to use ChatGPT in conjunction with other tools or technologies.	5	4	3	2	1
4. I can compare and evaluate the functions of ChatGPT and other language models.	5	4	3	2	1
5. I can understand how ChatGPT works.	5	4	3	2	1
Factor 2: Critical Evaluation (CE)					
6. I can evaluate the accuracy of ChatGPT responses.	5	4	3	2	1
7. I can determine whether ChatGPT's response is true.	5	4	3	2	1
8. I can identify errors in ChatGPT's responses.	5	4	3	2	1
9. I can evaluate the completeness of ChatGPT's responses.	5	4	3	2	1
10. I can recognize bias in ChatGPT's responses.	5	4	3	2	1
Factor 3: Communication Proficiency (CP)					
11. I can communicate effectively with ChatGPT.	5	4	3	2	1
12. I can ask appropriate and effective questions to ChatGPT.	5	4	3	2	1
13. I can use technical terms related to giving prompts to ChatGPT in conversations with ChatGPT.	5	4	3	2	1
14. I can ask accurate questions to ChatGPT.	5	4	3	2	1
15. I can elicit a response from ChatGPT to suit a particular situation.	5	4	3	2	1
Factor 4: Creative Application (CA)					
16. I can use ChatGPT to generate new ideas or solutions.	5	4	3	2	1
17. I can do creative writing and/or storytelling using ChatGPT.	5	4	3	2	1

18. I can use ChatGPT to analyze large sets of data.	5	4	3	2	1
19. I can improve my creativity and/or innovation skills using ChatGPT.	5	4	3	2	1
Factor 5: Ethical Competence (EC)					
20. I can identify potential ethical issues (e.g., plagiarism, copyright) associated with the use of ChatGPT.	5	4	3	2	1
21. I can recognize potential privacy issues related to the use of ChatGPT.	5	4	3	2	1



Appendix II. Survey 2 – Adapted TAME-ChatGPT

ChatGPT Attitude Survey	Strongly Agree	Agree	Neutral / No Opinion	Disagree	Strongly Disagree
Heard of ChatGPT but did not use it					
1. I am concerned about the reliability/accuracy/completeness of the information provided by ChatGPT.	5	4	3	2	1
2. I am concerned that using ChatGPT would get me accused of plagiarism.	5	4	3	2	1
3. I am concerned about the potential security risks of using ChatGPT.	5	4	3	2	1
4. I am concerned that the use of ChatGPT would be a violation of academic and university policies.	5	4	3	2	1
5. I am concerned about the potential privacy risks that might be associated with using ChatGPT.	5	4	3	2	1
6. I am afraid of relying too much on ChatGPT and not developing my critical thinking skills.	5	4	3	2	1
7. I am afraid of becoming too dependent on technology like ChatGPT.	5	4	3	2	1
8. I am afraid that using ChatGPT would result in a lack of originality in my university assignments.	5	4	3	2	1
9. I am enthusiastic about using technology such as ChatGPT for learning and research.	5	4	3	2	1
10. I believe technology such as ChatGPT is an important tool for academic success.	5	4	3	2	1
11. I think that technology such as ChatGPT is attractive and fun to use.	5	4	3	2	1
12. I am always keen to learn about new technologies like ChatGPT.	5	4	3	2	1
13. I trust the opinions of my friends or colleagues about using ChatGPT.	5	4	3	2	1
Heard of ChatGPT and used it					
14. ChatGPT helps me to save time when searching for information.	5	4	3	2	1
15. For me, ChatGPT is a reliable source of accurate information.	5	4	3	2	1
16. I recommend ChatGPT to my friends to facilitate their academic duties.	5	4	3	2	1

17. ChatGPT is more useful than other sources of information (Google, social media, etc.) that I have used previously.	5	4	3	2	1
18. I believe that using ChatGPT can save time and effort in my university assignments.	5	4	3	2	1
19. I have used tools or techniques similar to ChatGPT in the past.	5	4	3	2	1
20. ChatGPT is the first source I use when I need information for my university assignments.	5	4	3	2	1
21. It does not take a long time to learn how to use ChatGPT.	5	4	3	2	1
22. ChatGPT does not require extensive technical knowledge.	5	4	3	2	1

Appendix III. Survey 3 – ChatGPT ELT Attitude Survey

ChatGPT ELT Survey	Strongly Agree	Agree	Neutral / No Opinion	Disagree	Strongly Disagree
a. Teaching					
1. I believe ChatGPT can be used to help teachers while teaching reading .	5	4	3	2	1
2. I believe ChatGPT can be used to help teachers while teaching listening .	5	4	3	2	1
3. I believe ChatGPT can be used to help teachers while teaching writing .	5	4	3	2	1
4. I believe ChatGPT can be used to help teachers while teaching speaking .	5	4	3	2	1
5. I believe ChatGPT can be used to help teachers while teaching vocabulary .	5	4	3	2	1
6. I believe ChatGPT can be used to help teachers while teaching grammar .	5	4	3	2	1
7. I believe ChatGPT can be used to help teachers while teaching pronunciation .	5	4	3	2	1
8. I believe ChatGPT can be used to help teachers while teaching spelling .	5	4	3	2	1
b. Evaluation					
9. I believe ChatGPT can help teachers test/evaluate/grade students' speaking skills.	5	4	3	2	1
10. I believe ChatGPT can help teachers test/evaluate/grade students' writing skills.	5	4	3	2	1
11. I believe ChatGPT can help teachers test/evaluate/grade students' reading skills.	5	4	3	2	1
12. I believe ChatGPT can help teachers test/evaluate/grade students' listening skills.	5	4	3	2	1
13. I believe ChatGPT can help teachers test/evaluate/grade students' vocabulary skills.	5	4	3	2	1
14. I believe ChatGPT can help teachers test/evaluate/grade students' grammar skills.	5	4	3	2	1

15. I believe ChatGPT can help teachers test/evaluate/grade students with integrated tests.	5	4	3	2	1
16. I believe ChatGPT can help teachers give feedback on students' speaking skills.	5	4	3	2	1
17. I believe ChatGPT can help teachers give feedback on students' writing skills.	5	4	3	2	1
18. I believe ChatGPT can help teachers give feedback on students' reading skills.	5	4	3	2	1
19. I believe ChatGPT can help teachers give feedback on students' listening skills.	5	4	3	2	1
20. I believe ChatGPT can help teachers give feedback on students' vocabulary skills.	5	4	3	2	1
21. I believe ChatGPT can help teachers give feedback on students' grammar skills.	5	4	3	2	1
c. Lesson Design / Material Development					
22. I believe ChatGPT can help teachers develop assignments.	5	4	3	2	1
23. I believe ChatGPT can help teachers in lesson plan development.	5	4	3	2	1
24. I believe ChatGPT can help develop materials according to students' needs.	5	4	3	2	1
25. I believe ChatGPT can help develop materials according to students' levels.	5	4	3	2	1
26. I believe ChatGPT can help develop materials according to students' learning styles.	5	4	3	2	1
d. Attitude Towards ChatGPT as a Teacher					
27. I plan to allow my students to use ChatGPT while in the classroom.	5	4	3	2	1
28. I plan to allow my students to use ChatGPT while doing their homework.	5	4	3	2	1
29. I plan to allow my students to use ChatGPT when studying.	5	4	3	2	1
30. I believe ChatGPT can be used for the professional development of the teacher.	5	4	3	2	1

31. ChatGPT is a good source for the teacher to keep up with current literature in ELT.	5	4	3	2	1
32. ChatGPT is a good source for the teacher to learn and/or develop new ways for classroom management.	5	4	3	2	1
33. ChatGPT is a good source for the teacher to learn and/or develop new ways to communicate with the students.	5	4	3	2	1
34. ChatGPT is a good source for the teacher to learn about new techniques and strategies to teach language.	5	4	3	2	1
35. I plan on using ChatGPT often in the future as a teacher.	5	4	3	2	1
36. I plan on recommending ChatGPT to my students.	5	4	3	2	1
37. I plan on recommending ChatGPT to my colleagues	5	4	3	2	1

Appendix IV – Interview Questions

Interview Questions
1. Where and when did you hear about ChatGPT first?
2. How often do you use ChatGPT for your studies? Can you give a recent example?
3. What can be some ethical considerations for ChatGPT use?
4. Do you think ChatGPT helps you academically? If so, in what aspects?
5. Did you ever recommend ChatGPT to your friends? if not, would you in the future?
6. Do you plan to use ChatGPT as an English teacher to teach and/or evaluate reading? How do you plan on using it?
7. Do you plan to use ChatGPT as an English teacher to teach and/or evaluate listening? How do you plan on using it?
8. Do you plan to use ChatGPT as an English teacher to teach and/or evaluate writing? How do you plan on using it?
9. Do you plan to use ChatGPT as an English teacher to teach and/or evaluate speaking? How do you plan on using it?
10. Do you plan to use ChatGPT as an English teacher to teach and/or evaluate vocabulary? How do you plan on using it?
11. Do you plan to use ChatGPT as an English teacher to teach and/or evaluate grammar? How do you plan on using it?
12. Do you plan to use ChatGPT as an English teacher to teach and/or evaluate pronunciation? How do you plan on using it?
13. Do you plan to use ChatGPT as an English teacher to teach and/or evaluate spelling? How do you plan on using it?
14. Do you plan on using ChatGPT as a tool for professional development as a teacher? Please explain.
15. Do you plan on using ChatGPT for lesson plan design? Please explain.
16. Do you plan on using ChatGPT for material design? If yes, for what skills? In what ways do you plan on using ChatGPT for this purpose?
17. Do you plan on grading students using ChatGPT? Please explain.
18. Do you plan on using ChatGPT to develop tests? Please explain.
19. Do you plan on using ChatGPT to develop assignments/homework? Please explain.
20. What are your opinions about students using ChatGPT for their language classes? What is your stance on it as a future teacher? Please explain in terms of in-class use, for studying, and for help with assignments/homework.
21. Overall, what do you think is the main advantage and disadvantage of ChatGPT?

Appendix V – Preliminary Questionnaire

1. Cinsiyet:

- a.** Erkek
- b.** Kadın
- c.** Diğer

2. Yaş: _____

3. Üniversite Seviyesi:

- a.** 1.Sınıf
- b.** 2. Sınıf
- c.** 3. Sınıf
- d.** 4. Sınıf
- e.** Diğer: _____

4. İngilizce Öğretmenliği uzmanlık dersleri aldınız mı (Method, evaluation, classroom management, vb.)?

- a.** Evet
- b.** Hayır

5. ChatGPT’yi daha önce duydunuz mu ve kullandınız mı?

- a.** ChatGPT’yi daha önce duydum ve kullandım.
- b.** ChatGPT’yi daha önce duydum ama kullanmadım.
- c.** ChatGPT’yi daha önce duymadım.

Appendix VI – Written Consent Form

ARAŞTIRMA GÖNÜLLÜ KATILIM FORMU

Bu çalışma, **Türkiye’de İngilizce Öğretmenliği Öğrencilerinin ChatGPT’ye karşı Yeterlilikleri ve Tutumları** başlıklı bir araştırma çalışması olup ChatGPT gibi yapay zekâ dil modelleri hakkında İngilizce öğretmenliği üniversite öğrencilerinin yeterliliklerini ve tutumlarını öğrenmek, öğrencilerin akademik hayatlarında ne sıklıkla bu modellere başvurdıklarını, nasıl kullandıklarını, hakkındaki fikirleri ve bu kuramın geleceği hakkındaki düşüncelerini öğrenme amacını taşımaktadır. Çalışma, Umut Ceylan tarafından yürütülmektedir ve sonuçları ile öğrencilerin ChatGPT gibi yapay zekâ dil modelleri hakkında yeterlilikleri ve tutumları ortaya konacaktır.

- Bu çalışmaya katılımınız gönüllülük esasına dayanmaktadır.
- Çalışmanın amacı doğrultusunda, Nitel ve Nicel araştırma ile anket ve görüşme yapılarak sizden veriler toplanacaktır.
- Toplanan nitel verilerin değerlendirilmesi için aynı zamanda görüşmeler sırasında ses kaydı alınacaktır. Bu ses kaydı sadece araştırmacı tarafından dinlenilecek ve sadece veri kaydı ve değerlendirilmesi amacı için kullanılacaktır.
- İsminizi yazmak ya da kimliğinizi açığa çıkaracak bir bilgi vermek zorunda değilsiniz. Gönüllü katılım formunda isminizi belirtmeniz durumunda isminiz gizli tutulacaktır.
- Araştırma kapsamında toplanan veriler, sadece bilimsel amaçlar doğrultusunda kullanılacak, araştırmacının amacı dışında ya da bir başka araştırmada kullanılmayacak ve gerekmesi halinde, sizin (yazılı) izniniz olmadan başkalarıyla paylaşılmayacaktır.
- İstemeniz halinde sizden toplanan verileri inceleme hakkınız bulunmaktadır.
- Sizden toplanan veriler gizlilik ile korunacak ve araştırma bitiminde arşivlenecek veya imha edilecektir.
- Veri toplama sürecinde size rahatsızlık verebilecek herhangi bir soru/talep olmayacaktır. Yine de katılımınız sırasında herhangi bir sebepten rahatsızlık hissederseniz çalışmadan istediğiniz zamanda ayrılabilirsiniz. Çalışmadan ayrılmanız durumunda sizden toplanan veriler çalışmadan çıkarılacak ve imha edilecektir.

Gönüllü katılım formunu okumak ve değerlendirmek üzere ayırdığınız zaman için teşekkür ederiz. Çalışma hakkındaki sorularınızı araştırmacı Umut Ceylan’a yöneltebilirsiniz.

Araştırmacı Adı : Umut Ceylan
Adres :
İş Tel :
Cep Tel :

Bu çalışmaya tamamen kendi rızamla katılıyorum, istediğim takdirde çalışmadan ayrılabileceğimi bilerek verdiğim bilgilerin bilimsel amaçlarla kullanılmasını kabul ediyorum.

Katılımcı Ad ve Soyadı: _____

Tarih: __/__/__

İmza: _____

Appendix VII – Permission from the Ethical Committee

Evrak Kayıt Tarihi: 10.06.2024 Protokol No: 744442

Tarih: 02.07.2024



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

ÇALIŞMANIN TÜRÜ:	Yüksek Lisans Tez Çalışması
KONU:	Eğitim Bilimleri
BAŞLIK:	Türkiye'deki İngilizce Öğretmenliği Öğrencilerinin Yapay Zeka Bazlı Büyük Dil Modellerine Karşı Yeterlilikleri ve Tutumları
PROJE/TEZ YÜRÜTÜCÜSÜ:	Prof. Dr. Özgür YILDIRIM
TEZ YAZARI:	Umut CEYLAN
ALT KOMİSYON GÖRÜŞÜ:	-
KARAR:	Olumlu

ÖZ GEÇMİŞ

Adı Soyadı : Umut Ceylan

Yabancı Dil : İngilizce

Eğitim Bilgileri:

2022, Başkent Üniversitesi, Eğitim Fakültesi, İngilizce Öğretmenliği Bölümü

Mesleki Deneyimi:

2024, Araştırma Görevlisi, Dokuz Eylül Üniversitesi, Yabancı Diller Eğitimi Anabilim Dalı, İngiliz Dili Eğitim Bölümü

Yayınları:

Ceylan, U. (2025). A Systematic Review of Recent Studies on Second Language and Foreign Language Listening Strategies. *Contemporary English Language Studies*, 1(1), 72-88.

Bilimsel Faaliyetleri (Kongre, Sempozyum, Çalıştay vb.):

Sanatsal Faaliyetleri (Sergi, Konser, Gösteri vb.):

Görev Aldığı Projeler:

Ödülleri:

Patentleri/Telif Hakları/Faydalı Model:

Mesleki Birlik/Dernek/Kuruluş Üyelikleri: