A COMPARISON OF AI-BASED CHATBOT FEEDBACK AND TEACHER-WRITTEN FEEDBACK IN EFL WRITING CLASS

Sema YILDIZ HUSEYNOV
221113104

MASTER'S THESIS

Department of Foreign Languages Education
English Language Education with Master's Thesis
Danışman: Prof. Dr. Aysu Aryel Erden

İstanbul
T.C. Maltepe University
Graduate School
July, 2025

A COMPARISON OF AI-BASED CHATBOT FEEDBACK AND TEACHER-WRITTEN FEEDBACK IN EFL WRITING CLASS

Sema YILDIZ HUSEYNOV
221113104
ORCID: 0000-0003-3479-3195

MASTER'S THESIS

Department of Foreign Languages Education
English Language Education with Master's Thesis
Danışman: Prof. Dr. Aysu Aryel Erden

İstanbul
T.C. Maltepe University
Graduate School
July, 2025



JÜRİ VE ENSTİTÜ ONAYI

SEMA YILDIZ HUSEYNOV'ın "A COMPARISON OF AI-BASED CHATBOT FEEDBACK AND TEACHER-WRITTEN FEEDBACK IN EFL WRITING CLASS " başlıklı tezi 01.07.2025 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Maltepe Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği" nin ilgili maddeleri uyarınca Yabancı Diller Eğitimi Anabilim Dalı Yüksek Lisans/Doktora/Sanatta Yeterlik tezi oy birliğiyle/oy çokluğuyla, başarılı/başarısız olarak kabul edilmiştir.

	Unvanı. Adı ve Soyadı	İmza
Üye (Tez Danışmanı)	Prof. Dr. Aysu Aryel ERDEN Maltepe Üniversitesi	.....
Üye	Dr. Öğr. Üyesi Özlem PERVAN Maltepe Üniversitesi	.....
Üye	Dr. Öğr. Üyesi Aybars ERÖZDEN Okan Üniversitesi	.....

Enstitü Müdürü

ETİK İLKE VE KURALLARA UYUM BEYANI

Bu belge, Yükseköğretim Kurulu tarafından 19.01.2021 tarihli "Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge" ile bildirilen 6698 Sayılı Kişisel Verilerin Korunması Kanunu kapsamında gizlenmiştir.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to Professor Aysu Aryel Erden**,** my thesis advisor, for her invaluable guidance, encouragement, and unwavering support throughout the entire process. Her expertise and mentorship have been instrumental in shaping this study.

I would also like to extend my sincere thanks to Assistant Professor Özlem Pervan**,** whose support and constructive feedback have significantly contributed to the development of this thesis.

I would like to express my appreciation to the members of my thesis committee for their valuable time, insightful comments, and academic guidance, all of which have contributed to the refinement and completion of this work.

I also wish to acknowledge my colleagues for their occasional support and kind encouragement during this journey.

My heartfelt thanks go to my husband, Associate Professor Dr. Farid Huseynov**,** for his endless patience, understanding, and belief in me. His support has been indispensable.

To my beloved children, Emir and Karya**,** thank you for bringing light into my life and for your love and patience. You are my greatest motivation.

Lastly, I would like to express my love and appreciation to my entire family for their unwavering support, encouragement, and faith in me. This journey would not have been possible without them.

Sema YILDIZ HUSEYNOV

June, 2025

ÖZET

İNGİLİZCE YAZMA DERSİNDE YAPAY ZEKA DESTEKLİ CHATBOT GERİ BİLDİRİM İLE ÖĞRETMENİN YAZILI GERİ BİLDİRİMİNİN KARŞILAŞTIRILMASI

Sema Yıldız Huseynov
Yüksek Lisans Tezi
Yabancı Diller Eğitimi Anabilim Dalı
İngiliz Dili Eğitimi Tezli Yüksek Lisans Programı
Danışman: Prof. Dr. Aysu Aryel Erden
Maltepe Üniversitesi Lisansüstü Eğitim Enstitüsü, 2025

Bu çalışma, İngilizceyi yabancı dil olarak öğrenen (EFL) öğrencilerin yazılı anlatım becerilerinin geliştirilmesinde yapay zekâ tabanlı sohbet robotu geri bildirimleri ile öğretmen yazılı geri bildirimlerinin karşılaştırmalı etkililiğini incelemektedir. Araştırma, yazılı geri bildirimlerin söylem çözümlemesi ile öğrenci ve öğretmen görüşlerinin tematik analizine dayanmaktadır. Öğrenci paragraf metinlerine ChatGPT, Grammarly ve üç farklı öğretmenden oluşan beş farklı kaynaktan yazılı geri bildirim uygulanmıştır. Ayrıca 15 öğrenciyle görüşmeler yapılıp ardından 15 İngilizce öğretmeniyle de verilen geri bildirimlerle alakalı görüşmeler yürütülmüştür. Çalışmaya toplamda 15 öğrenci ve 15 İngilizce öğretmeni katılmıştır.

Söylem çözümlemesi bulgularına göre, yapay zekâ araçları özellikle dil bilgisi ve kelime düzeyinde daha yüksek sayıda ve doğru düzeltme sağlayabilmektedir. Ancak öğretmen geri bildirimleri, kişiselleştirilmiş, yönlendirici ve motive edici açıklamalar sunma açısından daha etkilidir. Öğrenci görüşleri, yapay zekâ geri bildirimlerinin hız ve netlik bakımından faydalı olduğunu kabul etmekle birlikte, anlam geliştirme ve metin organizasyonu gibi alanlarda öğretmen geri bildirimlerine öncelik verdiklerini göstermektedir. Öğretmen görüşleri ise yapay zekânın verimliliğini takdir etmekle birlikte, pedagojik kontrol, bağlamsal farkındalık ve etik boyutlara dikkat çekmektedir.

Sonuç olarak, yapay zekâ destekli yazılı geri bildirimler, öğretmen geri bildirimlerini tamamlayıcı nitelikte kullanılabilir; ancak insani etkileşimlerin duygusal, bağlamsal ve gelişimsel boyutlarını tam anlamıyla yerine getirememektedir. EFL bağlamlarında geri bildirim uygulamaları için bütüncül yaklaşımlar önerilmiş ve ileriye dönük araştırma alanlarına dair önerilerde bulunulmuştur.

Anahtar Kelimeler: Yapay zekâ geri bildirimi, öğretmen geri bildirimi, ChatGPT, Grammarly, EFL yazma, söylem çözümlemesi, öğrenci görüşleri, öğretmen görüşleri, geri bildirim okuryazarlığı

ABSTRACT

A COMPARISON OF AI-BASED CHATBOT FEEDBACK AND TEACHER-WRITTEN FEEDBACK IN EFL WRITING CLASS

Sema Yıldız Huseynov

Master's Thesis

Department of Foreign Languages Education

English Language Teaching Program with Thesis

Thesis Advisor: Prof. Dr. Aysu Aryel Erden

Maltepe University Graduate School, 2025

This study investigates the comparative effectiveness of AI-based chatbot feedback and teacher-written feedback in the context of English as a Foreign Language (EFL) writing instruction. The research adopts a qualitative approach, combining discourse analysis of written feedback with thematic analysis of student and teacher interviews. Feedback was collected from five sources,ChatGPT, Grammarly, and three human teachers,and applied to student opinion paragraphs. A total of 15 students participated in the study, along with 15 English language instructors who contributed to the evaluation of feedback effectiveness.

Findings from the discourse analysis reveal that AI tools provide higher volume and precision in surface-level corrections, particularly grammar and vocabulary. However, teacher feedback demonstrated greater strength in offering personalized, motivational, and content-oriented responses. Student interviews indicated a general appreciation for the immediacy and clarity of AI feedback but emphasized a preference for teacher input on deeper revision. Teacher interviews reflected cautious optimism, with educators valuing AI's efficiency while highlighting the need for pedagogical oversight and contextual sensitivity.

The study concludes that AI-originated feedback can serve as a valuable supplement to teacher-written feedback, but it cannot fully replicate the affective, contextual, and developmental dimensions of human interaction. Implications are offered for blended

feedback practices in EFL contexts, with suggestions for future research into long-term effects and ethical considerations.

Keywords: AI feedback, teacher feedback, ChatGPT, Grammarly, EFL writing, discourse analysis, student perceptions, teacher perceptions, feedback literacy

GENİŞLETİLMİŞ ÖZET

İNGİLİZCE YAZMA DERSİNDE YAPAY ZEKA DESTEKLİ CHATBOT GERİ BİLDİRİM İLE ÖĞRETMENİN YAZILI GERİ BİLDİRİMİNİN KARŞILAŞTIRILMASI

Yazılı geri bildirim, yabancı dil olarak İngilizce (EFL) öğretiminde öğrencilerin dilsel ve bilişsel gelişimlerini destekleyen temel bir öğretim aracıdır. Geleneksel olarak öğretmenler tarafından sağlanan yazılı geri bildirimler; dil bilgisel doğruluğun sağlanması, içerik geliştirme, organizasyon ve öğrenci motivasyonunu artırma gibi çeşitli işlevlere sahiptir. Ancak son yıllarda teknolojik gelişmeler, özellikle yapay zekâ (YZ) tabanlı araçların eğitimde kullanımı ile birlikte, yazılı geri bildirim süreçlerinde yeni yaklaşımların benimsenmesine yol açmıştır. Bu çalışmanın amacı, EFL yazma sınıfında öğretmen yazılı geri bildirimleri ile yapay zekâ destekli sohbet robotu (ChatGPT) ve otomatik yazım kontrol aracı (Grammarly) tarafından sağlanan yazılı geri bildirimlerin etkililiğini karşılaştırmalı olarak incelemektir.

Araştırma, karma yöntem desenine göre yapılandırılmıştır. Nicel veriler, 15 öğrencinin görüş paragraflarına beş farklı kaynaktan sağlanan yazılı geri bildirimlerin söylem çözümlemesi ile elde edilmiştir. Söylem çözümlemesinde Ferris (2006) ve Hyland (2003) gibi temel kaynaklardan yararlanılarak geri bildirimlerin işlevsel dağılımı (dil bilgisi, kelime bilgisi, içerik, organizasyon vb.) belirlenmiş, her kaynak tarafından sunulan geri bildirimlerin sayısı ve niteliği karşılaştırılmıştır. Ayrıca 15 öğrencinin ve 15 öğretmenin yarı yapılandırılmış görüşme kayıtlarının tematik analizine dayanmaktadır. Braun ve Clarke'ın (2006) tematik analiz yaklaşımı izlenmiş, öğrencilerin ve öğretmenlerin geri bildirim algıları beş ana tema altında toplanmıştır.

Söylem çözümlemesi sonuçlarına göre, yapay zekâ araçları dil bilgisi ve yüzeysel hataların düzeltilmesinde yüksek hacimli, hızlı ve doğru geri bildirimler sağlamaktadır. ChatGPT, dilsel doğruluğa odaklanmakta; Grammarly ise biçimsel hataları sistematik biçimde işaretlemektedir. Buna karşın öğretmen geri bildirimleri içerik geliştirme, organizasyon, öğrenciye özel geribildirim sağlama ve duygusal destek gibi alanlarda daha güçlü bulunmuştur.

Öğrenci görüşleri, yapay zekâ geri bildirimlerinin hızlı ve görsel olarak net olduğunu kabul etmekte, ancak önerilerin bazen bağlama uygun olmadığını ya da çok mekanik bulunduğunu ifade etmektedir. Öğrenciler, öğretmen geri bildirimlerinin özellikle anlam geliştirme, yönlendirme ve motive edici dil açısından daha değerli olduğunu vurgulamıştır. Öğretmen görüşleri ise yapay zekânın öğretmen iş yükünü azaltıcı potansiyelini kabul etmekle birlikte, pedagojik kontrolün kaybı, öğrencilerin geri bildirimi yanlış yorumlama riski ve yapay zekâya aşırı bağımlılık gibi endişeleri gündeme getirmiştir. Öğretmenler ayrıca geri bildirimin sadece düzeltme değil, öğrenme sürecini şekillendiren bir iletişim biçimi olduğunu belirtmişlerdir.

Çalışma sonucunda; yapay zekâ destekli yazılı geri bildirimlerin özellikle biçimsel doğruluk sağlamada etkili olduğu, ancak öğretmen yazılı geri bildirimlerinin içeriksel derinlik, kişiselleştirme, duygusal destek ve bağlam duyarlılığı açısından daha güçlü olduğu ortaya konulmuştur. Her iki geri bildirim türü farklı güçlü yönlere sahiptir ve bu nedenle birbirini tamamlayıcı şekilde kullanılması önerilmektedir. Öğrencilerin geri bildirim okuryazarlığının geliştirilmesi ve yapay zekâ geri bildirimlerinin eleştirel biçimde değerlendirilmesi için öğretim ortamlarında rehberli uygulamalara ihtiyaç duyulmaktadır.

Bu çalışma, EFL bağlamında hem yapay zekâ hem de öğretmen geri bildirimlerinin işlevselliğini karşılaştırmalı olarak analiz eden az sayıdaki çalışmadan biridir. Bulgular, öğretmenlerin geri bildirim sürecindeki yerinin korunması gerektiğini, yapay zekânın ise öğretimi destekleyici bir araç olarak yapılandırılması gerektiğini ortaya koymaktadır. Gelecek araştırmaların, farklı dil becerileri, uzun dönemli yazılı ürünler ve farklı yapay zekâ modelleri ile benzer karşılaştırmaları yapması önerilmektedir.

Anahtar Kelimeler: İngilizce yazma, öğretmen geri bildirimi, yapay zekâ geri bildirimi, ChatGPT, Grammarly, öğrenci algısı, söylem çözümlemesi, tematik analiz, geri bildirim okuryazarlığı

TABLE OF CONTENTS

JÜRİ VE ENSTİTÜ ONAYI ii

ETİK İLKE VE KURALLARA UYUM BEYANI iii

ACKNOWLEDGEMENTS iv

ÖZET v

ABSTRACT vii

GENİŞLETİLMİŞ ÖZET ix

TABLE OF CONTENTS xi

LIST OF TABLES xv

LIST OF ABBREVIATIONS xvi

1. INTRODUCTION 1

1.1 Statement of the Problem 1

1.2 Aim of the Study 3

1.3 Significance of the Study 4

1.4 Assumptions of the Study 6

1.5 Limitations of the Study 7

1.6 Keywords and Identifications 9

2. LITERATURE REVIEW 13

2.1 Teacher-Written Feedback in EFL Writing 14

2.1.1 Types and focus of teacher feedback 14

2.1.2 Effectiveness of teacher feedback: current consensus 16

2.2 Theoretical Frameworks Informing Feedback Practices 18

2.2.1 Sociocultural theory: feedback as scaffolded learning 19

2.2.2 Metacognitive theory: feedback and self-regulation 21

2.2.3 Dialogic feedback: feedback as ongoing dialogue 24

2.3 Tools for Automated Writing Feedback: Current State of AI in EFL Writing ……27

2.3.1 Grammarly: AI-powered grammar and style assistance 27

2.3.2 Turnitin feedback studio: plagiarism detection and teacher-annotated feedback…….. 30

2.3.3 ETS CRITERION: automated writing evaluation system 32

2.3.4 AI chatbots (e.g., ChatGPT): the new frontier in writing feedback..... 34

2.4 Comparative Perspectives: AI-Based Feedback vs. Teacher Feedback in EFL Writing 39

2.4.1 Focus and scope of feedback: form-focused vs content-focused 39

2.4.2 Volume and specificity of feedback 40

2.4.3 Interaction and dialogue..... 41

2.4.4 Student perceptions and preferences..... 42

2.4.5 Writing improvement outcomes 43

3. METHODOLOGY 45

3.1 Research Questions..... 45

3.2 Research Design 46

3.3 Participants 48

3.4 Data Collection 51

3.5 Data Analysis Procedure 58

3.5.1 analysis of written feedback (discourse analysis framework) 59

3.5.2 Thematic analysis of interviews (qualitative analysis) 62

4. FINDINGS AND DISCUSSION 69

4.1 Discourse Analysis of Feedback 69

4.1.2 analysis of feedback functions 75

4.1.3 summary of discourse analysis 78

4.2 Thematic Analysis of Student Interviews 79

4.2.1 Clarity and understandability 79

4.2.2 Usefulness and relevance 80

4.2.3 Motivation and affect 81

4.2.4 Personalization and context 81

4.2.5 Trust and credibility 82

4.3 Thematic Analysis of Teacher Interviews 83

4.3.1 Theme 1: Perceived usefulness of ai feedback 84

4.3.2 Theme 2: Concerns about clarity and misinterpretation 85

4.3.3 Theme 3: Pedagogical trust and control 86

4.3.4 Theme 4: Time and workload efficiency 86

4.3.5 Theme 5: Personalization and emotional value 87
4.3.6 Summary of teacher interview findings 87
4.4 Cross-Case Discussion and Interpretation 88
4.4.1 Agreement on ai's strength in surface-level feedback 88
4.4.2 Divergence on higher-order feedback and personalization 88
4.4.3 Emotional tone and motivation: human touch vs. Machine neutrality ... 89
4.4.4 Trust, authority, and pedagogical control 89
4.4.5 Toward an integrated feedback model 90
4.4.6 Summary of the cross-case discussion 90
5. CONCLUSION AND SUGGESTIONS 92
REFERENCES 96
APPENDICIES 107
Appendix:1: Semi Structered Interview Forms with the Consent Form 107
Appendix:2: Maltepe University Ethical Clearance 112
Appendix:3: Tables Showing Frequencies of Feedback Focus For Each Student 113
CURRICULUM VITAE 121

LIST OF TABLES

Table 1. Frequency of Feedback by Focus Across Five Sources 70

Table 2. Distribution of Feedback Functions Across Five Sources 75

Table 3. Themes from Students' Interviews 79

Table 4. Summary Table: Themes and Teacher Contributors 84

LIST OF ABBREVIATIONS

AI : Artificial Intelligence

EFL : English as a Foreign Language

ELT : English Language Teaching

ESL : English as a Second Language

FDT1 : Feedback from Teacher 1

FDT2 : Feedback from Teacher 2

FDT3 : Feedback from Teacher 3

L1 : First Language (Mother Tongue)

L2 : Second Language

SLA : Second Language Acquisition

ZPD : Zone of Proximal Development (Vygotsky's Theory)

TESOL : Teaching English to Speakers of Other Languages

CEFR : Common European Framework of Reference for Languages

St1-St15 : Student Participants(e.g., St1 = Student 1, St2 = Student 2, etc.)

T1–T15 : Teacher Interview Participants (e.g., T1 = Teacher 1, T2 = Teacher 2, etc.)

GPT : Generative Pre-trained Transformer (AI model used in ChatGPT)

TL : Target Language

AWE : Automated Writing Evaluation

WCF : Written Corrective Feedback

LLM : Large Language Models

RQ : Research Question

1. INTRODUCTION

1.1 Statement of the Problem

Within the field of English Language Teaching (ELT), the provision of efficient feedback on students' writing assignments has been widely acknowledged as a crucial factor in promoting language proficiency and improving communicative abilities (Ferris, 2003; Hyland & Hyland, 2006). The complex relationship between feedback, language acquisition, and technology has raised an intriguing inquiry: what is the influence of integrating Artificial Intelligence (AI) on the delivery of feedback in English as a Foreign Language (EFL) writing courses?

The English language, serving as a widely used means of communication across the globe, possesses exceptional significance in a progressively interconnected world (Crystal, 2003). Within the context of English as a Foreign Language (EFL) education, students encounter the complex aspects of English writing, putting effort to overcome linguistic obstacles and enhance their written communication skills. Throughout history, the pursuit of this undertaking has been under the guidance of dedicated educators, who, through their specialized knowledge, offer valuable critique that directs students towards achieving linguistic competence and successful communication (Hyland & Hyland, 2006).

However, with the advent of the digital age, we find ourselves on the verge of a profound shift in the educational landscape. The incorporation of artificial intelligence (AI) technology within educational settings has introduced novel perspectives, presenting unparalleled prospects to enhance and redefine the process of providing feedback (Dikli, 2006; Warschauer & Ware, 2006). Artificial Intelligence (AI) holds the potential to enhance efficiency and consistency in addressing linguistic challenges through its ability to quickly identify errors and provide feedback (Dodigovic & Tovmasyan, 2021, p. 82).

Even though the possible benefits of AI-generated feedback in EFL writing instruction are starting to become clear, it is still important to learn more about how well it works

compared to traditional feedback from a teacher. Existing studies give hints about how this changing landscape looks, but it is necessary to have a fuller look (Wang, 2025) that takes into account both students' and teachers' points of view as well as the complexities of language learning and the nuances of pedagogical practice.

This research aims to close this knowledge gap by comparing the feedback given by AI and feedback given by teachers in EFL writing classes in a very thorough way. The study hypothesizes that AI-generated feedback might offer benefits such as consistency and efficiency in addressing linguistic challenges, but its effectiveness compared to teacher-written feedback needs comprehensive exploration.

In the realm of English as a Foreign Language (EFL) instruction, feedback on student writing is a critical component for language development and skill enhancement. Traditionally, this feedback has been provided by human instructors, who, despite their expertise, may face challenges such as:

1. Time Constraints: Providing detailed feedback on student writings can be time-consuming, especially for instructors handling large classes or multiple batches.

2. Consistency: Human feedback, while valuable, can sometimes vary in consistency due to factors like fatigue, bias, or subjective interpretation of errors.

3. Scalability: In contexts where there are limited qualified EFL instructors available, it becomes challenging to ensure that all students receive quality feedback in a timely manner.

4. Adaptability: Human feedback might not always cater to the individual needs of each student, especially in larger classes where personalized attention is challenging.

With the rapid advancements in artificial intelligence (AI) and its increasing integration into educational settings, there's a potential solution to these challenges. However, the efficacy, accuracy, and acceptability of AI-generated feedback in EFL writing contexts remain largely unexplored.

It is essential to examine how AI-generated feedback compares to human feedback in terms of accuracy, relevance, and overall effectiveness. Equally important is

understanding how both students and teachers perceive AI-generated feedback whether they find it trustworthy, clear, and genuinely helpful. Furthermore, exploring the potential of AI feedback as a scalable solution without compromising quality is a critical area of inquiry.

1.2 Aim of the Study

The aim of this study is to explore how EFL students and teachers perceive and respond to written feedback provided by AI-based tools (ChatGPT and Grammarly) and by human instructors in a university-level writing context. Specifically, the study investigates how different feedback types are delivered and how they are interpreted, trusted, and acted upon by students, and how teachers evaluate the clarity, usefulness, and pedagogical value of AI-generated feedback in comparison to their own. The goal is to uncover the affordances and limitations of both feedback sources from the perspective of their users, contributing to a deeper understanding of how feedback functions as a mediating tool in second language writing instruction.

By offering a comprehensive comparison grounded in actual classroom practice, this study aims to contribute to the growing body of literature on technology-enhanced language learning. It also provides practical insights for EFL educators seeking to integrate AI tools into their teaching without compromising the instructional value of human feedback.

This study is guided by the following research questions:

RQ1: In what ways do AI-originated and teacher-written feedback differ in how they identify and address errors in EFL writing?

RQ2: How do students perceive and evaluate the quality, clarity, and usefulness of feedback from both AI-originated and teacher-written sources?

RQ3: How do teachers perceive and evaluate the quality, clarity, and usefulness of feedback from both AI-originated and teacher-written sources?

1.3 Significance of the Study

The significance of this study in the field of English Language Teaching (ELT) is underscored by the pressing need to address a critical gap in the existing literature and to provide practical solutions to both educators and learners. This chapter elucidates the paramount importance of conducting research on the topic of AI-generated feedback versus teacher-written feedback in EFL writing courses, articulating how the study will contribute to the literature and address the needs of practitioners and learners.

The motivation for this research stems from the recognition of a conspicuous gap in the literature. While feedback in English writing instruction has been a subject of extensive study, there remains a scarcity of comprehensive investigations specifically focused on the comparative analysis of AI-generated feedback and teacher-written feedback within the context of EFL writing courses. Existing research often lacks contextual specificity, and the nuances of AI-generated feedback, its advantages, and potential drawbacks warrant in-depth exploration.

The literature review revealed that although feedback modalities have been extensively examined, few studies have undertaken a thorough examination of AI-generated feedback in comparison to teacher-written feedback, particularly in EFL settings. This gap in the literature underscores the need for a study that delves into the effectiveness, perceptions, and contextual nuances of AI-generated feedback, thus contributing a unique perspective to the ongoing discourse in ELT.

This study seeks to make several noteworthy contributions to the literature and offers practical solutions to the challenges faced by both educators and EFL learners:

1. Enhanced Understanding of Feedback Modalities: By conducting a rigorous comparative analysis, this research aims to provide a nuanced understanding of the strengths and limitations of AI-generated feedback and teacher-written feedback in EFL writing courses. It will contribute to the evolving discourse on technology-enhanced language learning.

2. Pedagogical Insights: The study's findings will offer pedagogical insights into the integration of AI technology in EFL writing instruction. It will inform educators about

the potential benefits and drawbacks of AI-generated feedback, enabling them to make informed decisions about its incorporation into their teaching practices.

3. Student and Teacher Perceptions: Through interviews with both students and educators, this research will shed light on their perceptions, preferences, and attitudes toward feedback modalities. Such insights will inform educators about the receptivity of students to AI-generated feedback and provide a basis for addressing any concerns or misconceptions.

4. Contextual Considerations: Recognizing that contextual factors play a crucial role in the effectiveness of feedback methods, this study takes into account the contextual nuances that may influence the outcomes. It will provide valuable guidance for tailoring feedback practices to specific EFL writing contexts.

5. Practical Implementation Strategies: The research findings will guide educators in developing effective strategies for incorporating AI-generated feedback alongside teacher-written feedback, allowing for a balanced and responsive approach to feedback provision.

While the topic of feedback in English writing courses has been explored in various contexts, the unique setting of this study, focusing on EFL learners, adds contextual significance. EFL learners often face distinct challenges and requirements compared to their native English-speaking counterparts. This research acknowledges these differences and examines how AI-generated feedback can be contextualized to address the specific needs of EFL learners, thereby contributing to a more comprehensive understanding of feedback modalities in diverse linguistic and educational settings.

In conclusion, this study's significance lies in its potential to bridge a critical gap in the literature, offering practical solutions and insights to educators and learners in the field of English Language Teaching. By addressing the research questions, the study seeks to contribute meaningfully to the body of knowledge in ELT and offer valuable guidance for educators navigating the integration of AI technology into their writing instruction practices.

1.4 Assumptions of the Study

Every research study is based on certain underlying assumptions that are considered to be true for the purposes of inquiry. This study makes the following assumptions:

1. Participants Provided Honest Responses

   It is assumed that both the students and teachers who participated in interviews responded sincerely and shared their genuine opinions and perceptions regarding the feedback they received or provided. The credibility of thematic analysis depends on the authenticity of these insights.

2. Feedback Was Read and Considered Seriously by Students

   It is assumed that students are engaged with the feedback (whether from AI tools or teachers) with a sincere intention to understand and learn from it, rather than ignoring or dismissing it. The effectiveness of feedback depends significantly on the learner's receptiveness.

3. AI Tools (ChatGPT and Grammarly) Functioned Consistently

   It is assumed that the AI-based feedback tools (ChatGPT and Grammarly) provided consistent, unbiased, and accurate feedback across all student writing samples. Any variation in responses is presumed to result from the content of the texts rather than the tools themselves.

4. Teachers Provided Feedback Based on Their Professional Expertise

   It is assumed that the teacher-written feedback (FDT1, FDT2, FDT3) reflected the professional knowledge, pedagogical reasoning, and experience of the instructors. Their feedback is considered to be representative of common practices in EFL writing classrooms.

5. Students Had Comparable Writing Proficiency

Although student participants may have had minor variations in writing ability, it is assumed that all students possessed a similar level of English proficiency, allowing for a fair comparison of how different types of feedback impacted their writing.

6. Feedback Was Interpreted According to Intended Purpose

   It is assumed that both AI-generated and teacher-written feedback were interpreted by students in alignment with the feedback's intended purpose. Misinterpretations, if any, are assumed to be minimal or non-impactful on the general findings.

These assumptions form the foundation for the analysis and interpretation of both qualitative and quantitative data. Acknowledging them helps define the scope of the study and clarifies the boundaries within which the results are considered valid.

1.5 Limitations of the Study

While this research provides meaningful insights into the comparative effectiveness of AI-generated and teacher-written feedback in EFL writing instruction, several limitations must be acknowledged to contextualize the scope and generalizability of the findings.

First, the study was conducted with a sample of 15 students from a single institutional context, specifically, a university preparatory program in Türkiye. While thematic saturation was achieved in interviews and feedback analysis was thorough, the results may not be fully generalizable to all EFL learners or educational settings, particularly outside similar academic or geographical contexts.

Another limitation concerns data accessibility. The availability of datasets containing both AI-generated and teacher-written feedback with adequate contextual detail was limited. As a result, the study relied on a carefully selected sample, which, while rich in detail, may constrain broader applicability.

Although a growing number of AI-powered feedback tools exist (e.g., ProWritingAid, Quillbot, Turnitin Revision Assistant), this study focused specifically on ChatGPT and Grammarly as representative systems of two core categories: generative AI and automated writing evaluation (AWE). These tools were selected due to their high accessibility, widespread usage in educational contexts, and contrasting feedback styles.

Including a broader range of tools was not feasible due to time and resource constraints. Moreover, limiting the scope allowed for deeper comparative analysis. Future research could expand the investigation to include additional platforms and explore how tool-specific algorithms shape feedback quality and learner uptake.

Teacher variance introduced another constraint. Feedback was collected from three teachers with varying levels of experience and differing degrees of familiarity with the students. While this diversity reflects authentic classroom conditions, it also introduces variability that may influence feedback style and consistency.

All participants in this study were affiliated with the same university's English preparatory program. While this facilitated logistical coordination and ensured curriculum consistency, it also introduced a limitation in terms of generalizability. The shared institutional context likely influenced participants' perceptions of feedback, access to digital tools, and familiarity with AI systems. As such, the findings should be interpreted within this bounded setting. Future research should incorporate a more diverse range of institutions and learner populations to validate and expand upon the present conclusions.

Additionally, resource limitations including time and personnel posed constraints on the depth and breadth of data collection and analysis. Time limitations inherent in the academic research calendar influenced several aspects of this study. First, they restricted the scope of feedback sources and participant numbers. Although a more longitudinal investigation could have included follow-up writing tasks or delayed post-tests to measure uptake and revision outcomes, this was not feasible within the available time frame. Second, the analysis was limited to a single written assignment per student, meaning feedback impact was interpreted without observing actual revisions. Lastly, the number of interview participants was shaped by both scheduling challenges and time boundaries, potentially limiting the depth of emergent themes. These constraints should be considered when interpreting the generalizability and longitudinal implications of the findings.

Transcription accuracy and participant bias represent further limitations. Although interview transcripts were reviewed and refined, minor errors may persist due to the use

of automatic transcription tools. Student responses may also have been influenced by social desirability bias or pre-existing attitudes toward teachers and AI technologies.

While "effectiveness" in feedback studies can encompass a wide range of cognitive, linguistic, and longitudinal learning outcomes, this study adopts a narrower definition aligned with its qualitative scope. Specifically, effectiveness is interpreted through participants' descriptions of whether the feedback helped them notice their errors, provided actionable strategies for revision, improved their understanding of writing conventions, and encouraged engagement with the revision process. Broader measures such as post-test writing scores, grammatical accuracy rates, or delayed retention effects were not within the scope of this study. As such, effectiveness was analyzed thematically based on perceived clarity, depth, and applicability of feedback, as reflected in participant narratives.

Lastly, the study's findings reflect the functionality of AI tools during a specific window in 2025. Given the rapid evolution of these systems, future versions may differ significantly in terms of accuracy, responsiveness, and pedagogical value.

Despite these limitations, measures were taken to ensure methodological rigor, including collaboration with instructors, detailed documentation, and strategic planning to maximize available resources. These constraints also open up fruitful avenues for future research to build upon the current study's foundation.

1.6 Keywords and Identifications

AI-generated feedback: AI-generated feedback is feedback on written assignments that is made by artificial intelligence systems, such as automated grading tools, language processing algorithms, or computer-based writing evaluation software. This idea is in line with what is usually said about the term in the fields of educational technology and language assessment (Dodigovic & Tovmasyan, 2021; Warschauer & Ware, 2006).

Teacher-written feedback: This refers to the feedback that teachers or educators give to students on their written assignments. It includes the comments, suggestions, and changes that teachers make to help students get better at writing. The idea of written feedback

from a teacher is well-known in the writing assessment and teaching literature (Ferris, 2003; Carless, 2006).

EFL writing: EFL (English as a Foreign Language) writing is when people who do not speak English as their first language practice writing in English. In this study, "EFL writing" refers to the writing assignments and activities that happen in EFL classrooms, which are often part of programs to help people learn a new language. This idea is in line with what language education and EFL/ESL (English as a Second Language) pedagogy (Hyland & Hyland, 2006; Hyland, 2003) say about how to teach a language.

English Language Teaching (ELT): ELT is a more general term for teaching and learning the English language. It includes teaching writing as well as other parts of learning a language. In the field of language education, the idea of ELT is well-known (Richards & Rodgers, 2001).

Comparative analysis: This is the process of systematically looking at and comparing two or more variables, like AI-generated feedback and feedback written by a teacher, to find similarities, differences, and patterns. This idea comes from research methods and techniques for analysing data (Bryman, 2016).

Feedback effectiveness: Effectiveness refers to the extent to which written feedback supports student improvement by helping them notice, understand, and revise errors. It involves qualities such as accuracy, specificity, clarity, pedagogical depth, and long-term developmental value (Bitchener & Ferris, 2012; Carless & Boud, 2018).

Feedback Usefulness: Usefulness, in the context of written feedback, reflects distinct yet complementary interpretations for students and teachers.

For students, usefulness is defined as the perceived clarity, applicability, and motivational value of the feedback. It encompasses whether the feedback helps learners understand what to revise, how to revise it, and whether it enhances their confidence or causes anxiety. These perceptions are shaped by affective and cognitive dimensions of feedback reception (Carless & Boud, 2018; Hyland & Hyland, 2006).

For teachers, usefulness refers to the anticipated pedagogical effectiveness of the feedback,whether it promotes student uptake, supports writing development, and aligns with instructional goals. It also involves concerns about efficiency, appropriateness, and feedback literacy. As Winstone and Carless (2019) highlight, teachers often evaluate usefulness based on how likely the feedback is to foster learner autonomy and engagement in revision processes.

In this study, these dual perspectives on usefulness informed the development of thematic codes and guided the interpretation of interview data from both groups.

Student perceptions: Students' attitudes, beliefs, and personal experiences with feedback methods, such as AI-generated feedback and teacher-written feedback, are part of their perceptions. Wang (2025) says that this idea is based on the field of educational psychology and on qualitative research.

Teacher perspectives: The views and ideas of teachers about how AI-generated feedback is used and how well it works compared to teacher-written feedback in EFL writing instruction. This idea fits with qualitative research methods and studies of what teachers say (Carless, 2006).

Writing instruction: It includes all of the practices, strategies, and methods that teachers use to teach and improve students' writing skills. This idea is an important part of language education and teaching (Hyland & Hyland, 2006).

Technology-enhanced feedback: This is feedback that is given through digital tools and platforms, such as AI-driven systems and software, that are meant to improve the feedback process. This idea fits with what the research says about educational technology and how it can be used to teach writing (Dodigovic & Tovmasyan, 2021; Warschauer & Ware, 2006).

This thesis is organized into five chapters. Following this introduction, Chapter 2 reviews the existing literature on written feedback in language education, including studies related to teacher feedback, AI-assisted feedback, and learner/teacher perceptions. Chapter 3 outlines the methodology employed in the study, detailing the research design, participants, data collection tools, and analytical procedures. Chapter 4 presents the

findings of both the discourse analysis and thematic interview analysis. Finally, Chapter 5 concludes the thesis by summarizing the key findings, discussing the limitations of the study, and offering pedagogical implications and suggestions for future research.

2. LITERATURE REVIEW

Feedback in second language (L2) writing instruction has long been recognized as a critical component for improving learner outcomes. In English as a Foreign Language (EFL) writing classes, teachers traditionally invest significant time in reviewing students' texts and providing corrective feedback on language use, content development, and organization. Such feedback is considered *"a prominent and non-negotiable component in writing instruction"*, forming an integral part of the process approach to writing where students iteratively improve their drafts. Over the past several decades, changes in writing pedagogy have expanded feedback practices beyond teacher-written comments to include peer feedback, writing conferences, and computer-mediated feedback. However, while feedback is widely regarded as crucial for consolidating learning, research findings have not been unequivocally positive about its impact on writing development. Teachers often sense that conventional feedback practices may not be realizing their full potential in fostering student improvement. This chapter reviews the extensive literature on feedback in EFL writing, comparing traditional teacher-provided feedback with emerging AI-based feedback mechanisms. Foundational theories and studies are discussed alongside recent research (2020–2024) to establish a comprehensive understanding of how feedback influences EFL writing. In particular, the review covers (a) the role and effectiveness of teacher-written feedback in L2 writing, (b) key theoretical frameworks underlying feedback processes – including sociocultural theory, metacognitive theory, and dialogic feedback concepts, (c) current tools for writing feedback such as Grammarly, Turnitin Feedback Studio, ETS's Criterion, and AI chatbots like ChatGPT, and (d) empirical findings comparing AI-based feedback with teacher feedback, especially in EFL contexts. Where relevant, studies from Türkiye and similar EFL settings are highlighted to provide a contextual perspective. By synthesising this literature, the review will identify how AI-based chatbot feedback and teacher feedback differ or complement each other and underscore best practices for integrating technology with pedagogy to enhance EFL writing outcomes.

2.1 Teacher-Written Feedback in EFL Writing

Teacher-written feedback has been a cornerstone of second language writing instruction, grounded in the belief that learners benefit from expert guidance on their written production. In L2 writing courses, *"grammar correction is something of an institution"* as noted by Truscott (1996, p.327) – a practice so ingrained that students and instructors alike take it for granted as essential. Students generally expect and desire extensive feedback on their writing. Early research by Leki (1991) found that ESL students believed good writing should be error-free and *wanted all their written errors to be corrected*. Even decades later, surveys confirm that most teachers feel written corrective feedback (WCF) is needed and effective, and they continue to *"slave over students' errors"* despite uncertainty about its long-term impact. As Ferris (1999) acknowledged, the endeavor of marking L2 writing is *"fraught with uncertainty about its long-term effectiveness"* (p.1), yet teachers persist in providing feedback because of strong pedagogical convictions and student expectations. Indeed, a study of teachers' perspectives by Evans et al. (2010) confirmed that an overwhelming majority see WCF as a beneficial practice overall.

2.1.1 Types and focus of teacher feedback

Teacher feedback on writing can address a range of aspects, commonly categorized into feedback on form (grammar, vocabulary, mechanics) and feedback on content (ideas, organization, argumentation). Within form-focused feedback, a major line of inquiry has been *written corrective feedback (WCF)* on grammatical errors. Research has distinguished various strategies, such as direct vs. indirect feedback – directly providing the correct form versus indicating an error's presence and expecting the student to self-correct and focused vs. unfocused feedback – targeting a subset of error types versus marking all errors. Studies have also explored *metalinguistic feedback*, where teachers give explanations or codes about the error, and *feedback mode* (written comments on paper, electronic annotations, face-to-face conferences, etc.). Each approach has its advantages: direct feedback provides explicit guidance which may benefit lower-proficiency writers, whereas indirect feedback can stimulate student reflection and self-editing skills. A recent meta-analysis by Lim & Renandya (2020) found that overall WCF has a moderate positive effect on L2 grammatical accuracy, with direct feedback showing

a slightly larger (though not statistically significant) effect size than indirect feedback. This suggests that while students can improve accuracy through feedback, the difference between feedback strategies is nuanced and other factors mediate effectiveness. Notably, learner proficiency emerged as a key moderator – lower-level students may struggle to benefit from indirect hints and often require more explicit guidance.

In terms of content-focused feedback, teachers often comment on the clarity of ideas, coherence and cohesion of arguments, organization of paragraphs, and relevance of support. Such commentary aligns with a process-oriented approach, aiming to guide students through revisions that improve the substance of their writing, not just surface correctness. Hyland (2003) observed that effective feedback plays multiple roles: it not only corrects linguistic errors but also functions as a *"constructive critique of the text"* that can point students forward to future writing development. Teacher comments might praise strong points, ask questions for clarification, or suggest ways to reorganize or elaborate content. This comprehensive feedback approach is time-intensive but resonates with students' desire for detailed input. For instance, Zhang (2020) found that students valued feedback that addressed both language accuracy and content issues, perceiving it as more helpful for improving their drafts.

Despite teachers' best efforts, the efficacy of written feedback has been debated. Some early research questioned whether extensive teacher feedback actually leads to improved writing. A seminal controversy was sparked by Truscott (1996), who argued that grammar correction in L2 writing is ineffective, even counterproductive, and should be abandoned. Truscott's position was that red-inked corrections do not produce durable gains in accuracy and might discourage learners or lead to *"pseudo-learning"* – superficial short-term fixes without genuine language development. Ferris (1999), in a direct response, disputed this claim and advocated for a more nuanced view: if error correction is "selective, prioritized, and clear", it *can* benefit at least some learners. She pointed out that many earlier studies had methodological flaws (e.g., lack of control groups) and that it was premature to write off corrective feedback entirely. Over the next decade, researchers carried out more rigorous experiments and longitudinal studies to settle this "grammar correction debate." The results have been mixed but generally more favorable to the use of WCF under the right conditions. Meta-analyses in the 2010s concluded that

WCF can indeed lead to greater grammatical accuracy in new pieces of writing. For example, a meta-analysis by Kang & Han (2015) reported overall positive effects of WCF, though moderated by factors such as feedback timing, the specificity of feedback, and learner variables. Likewise, a synthesis by Russell and Spada (2006) noted that while oral CF in speaking has strong support, evidence for WCF's role in L2 acquisition was initially inconclusive, gradually moving toward a consensus that it *can* aid accuracy given adequate practice and consistency.

It is important to note that teacher feedback is not limited to correcting errors. Teachers often provide praise and encouragement, make suggestions for further development, and tailor their feedback tone to motivate students. According to Hyland & Hyland (2006), teacher-written feedback frequently uses hedges, questions, and personal comments to soften criticism and maintain rapport with students, an aspect that automated feedback typically lacks. Such *interpersonal dimension* of feedback can affect how students uptake and apply the comments. Research indicates that student engagement with feedback – their cognitive, emotional, and behavioral investment in using the feedback – is critical to whether improvements occur. Zhang and Hyland (2018) demonstrated through case studies that two Chinese EFL students' degree of engagement (e.g. reflecting on comments, seeking clarification, revising diligently) mediated the impact of both teacher and automated feedback on their writing. In other words, even high-quality feedback may have little effect if a student is unengaged or overwhelmed, whereas a motivated student can benefit from even minimal feedback by actively processing it. Teachers thus face the dual challenge of delivering useful feedback and fostering students' feedback literacy – the ability to understand and use feedback constructively.

2.1.2 Effectiveness of teacher feedback: current consensus

The current state of research suggests a guarded consensus that teacher-written feedback, while not a panacea, does contribute to improved writing accuracy and quality over time, especially when delivered thoughtfully. A comprehensive meta-analytic review by Lim & Renandya (2020) aggregated 35 studies and found a positive, moderate effect size for WCF on L2 writing accuracy. This indicates that learners who receive systematic feedback tend to outperform those who receive none or only minimal feedback,

confirming teachers' intuitive belief that feedback is beneficial. However, the effectiveness is mediated by many factors: the nature of the feedback (direct vs indirect, amount of information given, etc.), the focus (many studies advocate focusing on a subset of error types rather than every error to avoid overload), the timing (immediate feedback on drafts vs delayed feedback on final products), and individual learner differences (e.g., proficiency level, motivation, anxiety). For instance, learners at beginner levels might benefit more from direct correction since they lack the knowledge to self-correct, whereas advanced learners might gain more from comments that prompt reflection and self-editing. Additionally, the writing genre and task type can influence what kind of feedback is most relevant – an argumentative essay might need more feedback on logic and evidence use, while a narrative might invite feedback on coherence and description.

Another aspect of effectiveness is long-term retention and transfer of improvements. Ferris (2010) and others have raised the question: do students internalize corrections such that they avoid the same errors in future writings? Some longitudinal studies (e.g., Bitchener & Knoch, 2010) found that students who received consistent feedback on certain error categories (like verb tense or article use) showed significant reduction in those errors over several months, implying lasting learning. Others have found more transient improvements, where accuracy increases in revisions of the same piece (when directly responding to feedback) but does not carry over to entirely new essays (Truscott, 2007; Chandler, 2003). Overall, the literature supports that iterative feedback loops – where students apply feedback on multiple drafts or in subsequent assignments – are key to cementing writing gains (Ferris, 2011; Hyland & Hyland, 2006). This aligns with process-writing pedagogy where writing is seen as a cycle of drafting, feedback, and revising, rather than a one-shot product.

While teacher feedback remains central, researchers have increasingly acknowledged its limitations. It is labor-intensive for instructors, sometimes leading to delays in returning papers by which time students may have moved on (Lee, 2008). There is also the risk that students become overly reliant on teacher corrections, rather than developing self-editing skills (Ferris, 2002). Such concerns have spurred interest in alternative and supplementary feedback sources (peers, automated systems) and in strategies to make feedback more learner-centered. For example, teachers might employ peer review sessions or self-

assessment checklists to engage students in the feedback process (Min, 2006; Andrade & Du, 2007). Nonetheless, surveys in various contexts continue to show that students prefer teacher feedback over other forms. In a recent study in Saudi Arabia, university students rated teacher feedback as the most preferred feedback type, far above peer feedback (which was least preferred). These students responded positively to the idea of automated writing evaluation tools but expressed dissatisfaction with those tools' limitations in handling content and organization feedback (Alsowat, 2022). Such findings illustrate that while students value the authority and comprehensiveness of teacher feedback, they are also open to new feedback modalities if those can address their needs.

In summary, teacher-written feedback in EFL writing is a time-honored practice supported by both tradition and a substantial body of research. It plays an irreplaceable role in guiding learners to higher levels of writing proficiency, offering not just linguistic correction but also mentorship in how to communicate effectively in written English. However, it is most effective when integrated into a broader ecosystem of feedback and revision activities. The next sections will delve into theoretical frameworks that explain *why* feedback works (when it does) by looking at cognitive, social, and dialogic dimensions of the feedback process. This theoretical lens will provide a foundation for understanding how new forms of AI-based feedback might emulate or differ from human feedback in facilitating learning.

2.2 Theoretical Frameworks Informing Feedback Practices

Multiple theoretical perspectives shed light on how feedback operates in language learning and why it can lead to improved writing performance. This section will highlight three key frameworks relevant to feedback in EFL writing: sociocultural theory, metacognitive theory, and dialogic feedback. Each offers a lens to interpret the teacher-student feedback dynamic (and by extension, student-AI feedback interactions), underlining different mechanisms – social interaction, self-regulation, and dialogue – that make feedback effective.

2.2.1 Sociocultural theory: feedback as scaffolded learning

Sociocultural theory, based on the work of Vygotsky (1978), emphasizes that learning is fundamentally a socially mediated process. Knowledge is constructed through interaction, and with appropriate support, learners can perform beyond their independent ability. A central concept is the Zone of Proximal Development (ZPD) – the gap between what a learner can do alone and what they can do with guidance from a more knowledgeable other. Feedback in writing can be viewed as a form of scaffolding provided by the teacher (or a capable peer or system) to assist the learner in bridging that gap. From a sociocultural perspective, teacher comments and corrections are not just information about errors; they are *mediating tools* that help students internalize new linguistic and rhetorical knowledge through guided practice.

Vygotsky's principle that *"the secret of effective learning lies in the nature of the social interaction between two or more people with different levels of skills and knowledge"* directly applies to teacher-student feedback. The teacher, possessing greater expertise in English writing, provides strategic assistance,be it pointing out a grammatical mistake or questioning an argument's logic,that the student uses to improve the text and, ideally, learns from to improve future writing. Over time, what begins as an external scaffold (e.g., a teacher's margin comment) can become part of the learner's internal repertoire (the student learns to notice that type of error or lapse in coherence by themselves). In this way, feedback enables a form of guided participation in the target language discourse, gradually transferring responsibility to the learner. Although ChatGPT can simulate interactive conversation, it lacks awareness of learners' prior knowledge, emotional state, or developmental readiness,key elements in scaffolding within the ZPD.

Empirical studies grounded in sociocultural theory have demonstrated these concepts in action. For example, Aljaafreh and Lantolf's (1994) classic study of tutor-learner interactions showed how graduated, contingent feedback (starting with implicit clues and moving to explicit correction only as needed) helped L2 writers progress in accuracy, illustrating the sensitive calibration of support within the learner's ZPD. More recent work by Rahmah Fithriani (2019) in an Indonesian EFL context explored students' perceptions of written feedback benefits through a sociocultural lens. The study found

that students overwhelmingly felt teacher feedback improved their writing quality and skills, and it also *"encourages critical reasoning, and promotes learner autonomy"*. Notably, Fithriani reported that social interactions during feedback activities helped students extend their ZPDs. In practice, this meant that class sessions where students and the instructor discussed feedback (for instance, going over common errors or exemplary texts) enabled students to handle slightly more complex writing issues than they could independently, a tangible manifestation of ZPD expansion.

Sociocultural theory also informs *how* feedback is delivered. It suggests that feedback should ideally be dialogic and contingent: initially, more explicit guidance can be given, but as the learner gains competence, the support is dialed back (fading of scaffolding) to encourage independence. Teachers intuitively do this when they decide to mark only the first few instances of a repeated error and then prompt the student to find the rest, or when they shift from providing corrections to asking the student to reflect on a problem. In addition, Vygotskyan theory underscores the role of affect and encouragement in interaction. A supportive social context can lower anxiety and build the confidence learners need to attempt revisions and take risks with language. This aligns with the common practice of balancing criticism with positive remarks in feedback to maintain a constructive tone.

Peer feedback and collaborative writing activities can also be viewed through a sociocultural lens. Although peer reviewers are not "experts" in the same way teachers are, peer interaction offers a different kind of scaffold, often in more accessible language, and through a process of mutual negotiation of meaning. Some studies (e.g., Villamil & De Guerrero, 1996) have documented how pairs of students working together on writing tasks provide scaffolding to each other, collectively noticing problems and resolving them. The sociocultural perspective values these interactions for creating a community of practice where learners co-construct knowledge.

In summary, sociocultural theory contributes the insight that feedback works not merely by transferring information, but by *engaging learners in socially mediated problem-solving*. Effective feedback, from this viewpoint, is that which is optimally tuned to the learner's current developmental level – enough help to advance their skills, but not so

much as to eliminate the productive struggle. The theory thus provides a strong rationale for teacher involvement in feedback: a skilled teacher can diagnose a student's needs and provide the right level of scaffolded feedback, something that automated systems strive to emulate but may not always achieve. It also implies that the interactive aspect of feedback (back-and-forth clarification, dialoguing about comments) is central to its efficacy – a theme which will be revisited under dialogic feedback. Notably, research in a variety of L2 contexts, including Türkiye, reinforces these ideas. For instance, a study by Uysal (2020) in a Turkish university context found that students considered teacher feedback more trustworthy and motivating than automated alternatives. It is found that teacher scaffolding in feedback sessions was key to students' improvements, and students reported that guided revision exercises in class (a sociocultural activity) helped them understand and apply the feedback better than when they worked in isolation.

2.2.2 Metacognitive theory: feedback and self-regulation

Writing in a second language is not only a linguistic endeavor but also a cognitive one that involves planning, monitoring, and evaluating one's work. Metacognitive theory deals with awareness and control of one's cognitive processes – essentially, "thinking about thinking." In the context of L2 writing, metacognitive skills include the ability to plan one's writing (e.g., brainstorming, outlining), monitor progress (noticing errors or logical gaps as one writes), and evaluate and revise the text. Feedback can play a pivotal role in developing these metacognitive skills by prompting learners to reflect on their writing strategies and outcomes (Teng & Zhang, 2016, p. 20).

When a teacher provides feedback, at a basic level it draws the student's attention to something that went wrong (or could be improved) – this is a form of external regulation guiding the learner. Over time, if the student internalizes this guidance, they may begin to *self-regulate*, catching similar issues on their own. For example, a teacher's repeated comments about vague topic sentences might lead a student to start double-checking the clarity of their topic sentences as a matter of habit. Metacognitively, the student has developed a strategy to monitor a particular aspect of their writing. Thus, feedback can act as a catalyst for metacognitive awareness, helping students become more conscious of their errors, knowledge gaps, or the effectiveness of their arguments.

Research on feedback and metacognition suggests that different feedback sources may influence students' strategy use in distinct ways. A recent experimental study by Zhang & Zhang (2022) specifically examined how teacher feedback, peer feedback, and automated feedback affect students' use of metacognitive writing strategies. In this study, EFL university students were divided into three groups over a two-month period: one group received extensive teacher feedback on their essays, another engaged in structured peer feedback, and the third used an Automatic Writing Evaluation (AWE) system (Criterion) for feedback. The students' metacognitive strategy use (planning, monitoring, and self-evaluating strategies) was measured before and after the intervention. The results were telling: overall, the teacher feedback group showed the greatest enhancement in metacognitive strategy use, followed by the peer feedback group, whereas the automated feedback group showed comparatively less improvement. In fact, for certain strategies, automated feedback had a hindering effect. The authors found that automated feedback significantly hindered students' use of monitoring strategies, while teacher and peer feedback had a promotive effect on all types of metacognitive strategies. One interpretation is that teacher and peer feedback, being more interactive and explanatory, encouraged students to think deeply about the feedback, selectively attend to important issues, and reflect on how to revise – all metacognitive activities. In contrast, the automated feedback (likely presented as a list of errors or suggestions) may have led some students to follow suggestions uncritically or become overly dependent on the tool, thereby reducing their own monitoring and decision-making. This finding underscores a potential weakness of AI feedback: while it is immediate and comprehensive in flagging issues, it might not engage students in the same reflective process that human feedback can stimulate.

Metacognitive theory would suggest that the most valuable feedback is that which prompts learners to engage in self-evaluation. One approach consistent with this is having students respond to feedback – for instance, writing a revision plan or a short reflection on what they learned from the comments. Such practices force students into a metacognitive stance, processing the feedback at a deeper level. Teachers can scaffold this by asking questions in their comments (e.g., "What is your main evidence for this point?") that the student must think about, rather than just giving directives. Studies have shown that when students actively use feedback to set goals or criteria for their next draft,

they tend to improve more than those who passively receive corrections (Nelson & Schunn, 2009).

Another area where feedback intersects with metacognition is in reducing writing anxiety and building confidence. Students who perceive feedback as useful and fair develop more self-efficacy in writing,a belief that they can improve with effort,which in turn makes them more likely to employ metacognitive strategies (like planning more carefully or double-checking their work). A systematic literature review on Grammarly (an AWE tool) found that it not only reduces errors but also "enhances students' writing confidence, and fosters autonomous learning" (Llausas, Ruiz, Ayucan, & Evardo Jr., 2024, p. 3508). The implication is that when students trust a feedback source (be it a tool or a teacher) and see improvements, they become more willing to take charge of their learning process,a key metacognitive outcome.

In EFL contexts, promoting metacognitive awareness is particularly crucial because learners often need to become their own editors outside the classroom. Metacognitive theory supports practices such as self-editing checklists, training students to systematically review their work for common issues before submission. Teachers who incorporate such strategy instruction along with their feedback often find students become less reliant on teacher correction over time. This is aligned with the goal of developing *"autonomous writers"*. In connection with technology, one interesting emerging research direction is how students use tools like ChatGPT to reflect on their writing. A study by Teng (2025) examined EFL students' metacognitive experiences using ChatGPT for writing feedback. While detailed results are forthcoming, the study premise suggests that when students query ChatGPT about their text, they might develop a sense of how to evaluate and question their own writing – essentially using the AI as a mirror to trigger self-reflection.

To sum up, metacognitive theory highlights that feedback is most powerful when it does not just *fix* the immediate problem in a piece of writing, but when it helps learners *learn how to learn* – to plan better, to monitor their writing process, and to evaluate their output critically. Teacher feedback that encourages this (through strategies like asking students to find patterns in their errors, or to articulate what they will do differently next time) can

have lasting benefits. In contrast, there is a concern that overly convenient feedback (as sometimes provided by AI) might short-circuit this process if students simply accept suggestions without understanding. The challenge and opportunity in leveraging AI tools is to use them in a way that supports metacognitive development – for example, by having students compare the AI's feedback with their own judgments, or discuss AI suggestions in class, thus turning a potentially surface-level fix into a learning experience.

2.2.3 Dialogic feedback: feedback as ongoing dialogue

Traditional models of feedback often implied a one-way transmission: the teacher "delivers" feedback to the student, who then applies it. However, contemporary perspectives advocate for dialogic feedback, which conceptualizes feedback as a two-way conversational process between students and teachers (or among peers). A dialogic approach is rooted in theories of learning that emphasize interaction and the co-construction of meaning (some of which overlaps with sociocultural theory) and has been highlighted in higher education research on assessment (e.g., Nicol & Macfarlane-Dick, 2006; Carless et al., 2011). In a dialogic feedback framework, feedback is not an end point (a teacher's final say on a piece of writing) but part of an ongoing *dialogue* where students clarify the feedback, express their interpretations, and possibly negotiate revisions.

Nicol (2010) proposed that feedback should be *"conceptualised as a dialogue between teacher and student and/or peer-to-peer where meaning is constructed"*, rather than as a simple delivery of information. The underlying idea is that students need to actively make sense of feedback messages – and this sense-making is greatly enhanced by opportunities to discuss or respond to the feedback. For example, rather than just handing back an essay with comments, a teacher might hold a short conference with each student to talk through the comments, or have students write back a reply or plan. In doing so, any misunderstandings can be immediately addressed (e.g., the student might misinterpret a written comment, but in conversation, the teacher can clarify what was meant). Dialogic feedback thus helps close the loop of communication that is often left open in written feedback alone.

Carless et al. (2011) describe feedback as inherently *multi-directional and socially constructed*, involving active learner engagement. The student is not a passive recipient but a crucial agent in the process – they ask questions, seek clarification, and connect feedback to their own goals. In this sense, feedback involves dialogic processes whereby learners make sense of information from various sources and use it to enhance their work or learning strategies. The inclusion of various sources is notable – dialogic feedback can encompass teacher comments, peer comments, self-reflection, and even automated feedback outputs, all as part of a broader dialogue about quality and improvement. What makes it dialogic is the interplay and reflection, not the medium of feedback per se.

Implementing dialogic feedback can take several forms in writing classes. Teacher–student conferences are a classic dialogic method: sitting down one-on-one (or in a small group) to discuss a draft. This format encourages students to voice their intentions ("What I was trying to say here was…"), and teachers can tailor their advice more personally. Research has found that many students appreciate oral feedback for its immediacy and clarity. For instance, in an EFL writing context, Ene & Upton (2014) observed that synchronous online discussions about teacher feedback led to higher uptake of revision suggestions than written comments alone, presumably because of the interactive nature. Peer workshops can also be dialogic. When well-structured, students not only give each other comments but also ask questions like "What did you mean by this sentence?", prompting writers to articulate and thereby refine their ideas.

Dialogic feedback links closely with the concept of feedback literacy – students' ability to understand and act on feedback. Carless and Boud (2018) argue that part of developing students' feedback literacy is engaging them in dialogue, so they internalize standards and quality criteria. In writing, this might involve collectively discussing evaluation criteria or having students assess sample essays and talk about them. This way, when they receive feedback, they better grasp the underlying expectations and can more readily see how to improve.

The dialogic approach also addresses the emotional aspect of feedback. Written comments can sometimes be taken harshly or cause confusion, whereas a conversation can soften criticisms and ensure they are constructive. A dialogic space allows students

to express if they feel a comment is unclear or even unfair, giving the teacher a chance to explain or reconsider. This dynamic helps maintain a supportive learning relationship, which in turn can increase student willingness to take feedback on board (Carless, 2015, p. 90).

There is emerging empirical support for the effectiveness of dialogic feedback. For instance, recent work by Wu & Schunn (2020) showed that when students engaged in back-and-forth discussion on feedback (through threaded online comments in their study), they revised more effectively and produced higher quality final texts than those who only received one-way comments. The dialogues often revealed students' thought processes, enabling instructors to provide more targeted guidance or correct misinterpretations in real time.

In the context of the thesis topic – AI-based chatbot feedback vs. teacher feedback – dialogic theory raises interesting questions. Can AI chatbots provide a form of dialogic feedback? A chatbot like ChatGPT is inherently conversational: a student can input their essay and then have a conversation about it, asking for clarification or further advice, much as they might with a tutor. In principle, this could simulate the dialogic process, with the student probing the AI for more details ("Can you explain why my thesis statement is weak?") and the AI responding. This is a step beyond older AWE tools, which typically just output a static set of comments. Some scholars are optimistic that AI tools can indeed support a more dialogic form of feedback interaction, given their interactive nature. For example, Hwang and Chang (2021) suggest that *"It has the potential to provide students with instant, accessible, and personalised feedback anytime"* in a dialogic manner, though they caution that guidance is needed to ensure quality. ChatGPT offers interaction on command, but lacks genuine responsiveness and memory of past drafts, which limits its capacity for sustained dialogic feedback.

Overall, dialogic feedback theory encourages educators to foster a culture of conversation around writing improvement. It aligns with sociocultural ideas of learning through interaction and with metacognitive ideas of students taking an active role. As we consider AI tools, one measure of their success might be how well they engage students in a dialogue *about* their writing as opposed to just delivering a monologue *to* the student. The

following section will examine current AI-based tools and their capabilities, setting the stage for comparing how these tools' feedback aligns or contrasts with traditional teacher feedback along dimensions like content, focus, and the potential for dialogue.

2.3 Tools for Automated Writing Feedback: Current State of AI in EFL Writing

Advancements in artificial intelligence and natural language processing have led to the emergence of various tools designed to provide feedback on writing. These range from grammar checking software to fully-fledged conversational agents. In EFL writing instruction, such tools are increasingly used to supplement teacher feedback, with the promise of immediacy and scalability – especially useful in contexts where teachers face large class sizes and heavy workloads. This section reviews four prominent tools for writing feedback: Grammarly, Turnitin's Feedback Studio, ETS Criterion, and AI chatbots (exemplified by ChatGPT). For each tool, its features and affordances are outlined and summarize academic findings on its effectiveness and limitations as reported in recent studies. Understanding these tools will provide context for how AI-based chatbot feedback compares to teacher feedback, as these are the technologies often at the forefront of AI-assisted writing evaluation.

2.3.1 Grammarly: AI-powered grammar and style assistance

Grammarly is one of the most widely used automated writing evaluation (AWE) tools globally, with millions of users including ESL/EFL students. It functions as a real-time writing assistant, available via browser extensions or as a standalone app. Grammarly automatically checks text for grammatical errors, spelling mistakes, punctuation, and stylistic issues, underlining potential problems and offering suggestions for correction. It also provides some feedback on clarity, conciseness, vocabulary usage, and tone. Essentially, Grammarly leverages a combination of rule-based NLP and machine learning to detect deviations from Standard English norms and suggests edits.

According to a tech review by Barrot (2023, p.33), *"Grammarly identifies duplicate content and errors in grammar, vocabulary, mechanics, and language style"*, and offers an accessible means for students to get computer-mediated corrective feedback outside of class. Its popularity in EFL contexts stems from the immediacy of feedback – students

can get error alerts as they write or upon uploading a draft, without waiting for a teacher's red pen. This aligns with students' desire for quick feedback and frees up some teacher time on basic corrections. Importantly, Grammarly can act as a learning tool: by seeing repeated alerts (e.g., consistently flagged verb tense errors), students may become more aware of their common mistakes.

Research on Grammarly's effectiveness in EFL writing has generally reported positive outcomes, albeit with caveats. A systematic literature review (Dembsey, 2017; updated by O'Neill & Russell, 2019) indicated that using Grammarly can significantly reduce grammatical errors in students' writing, especially lower-order concerns like articles, prepositions, and subject-verb agreement. More recent sources echo that Grammarly "significantly reduces grammatical errors, enhances students' writing confidence, and fosters autonomous learning" (Llausas, Ruiz, Ayucan, & Evardo Jr., 2024, p. 3512). The confidence boost is an interesting finding – possibly students feel more assured in their writing when an AI has vetted it, which may reduce anxiety. It might also encourage autonomy because students can self-correct with Grammarly's help before turning in their work, rather than relying solely on the teacher to catch mistakes.

However, studies also highlight limitations of Grammarly. The tool may miss more nuanced errors or context-specific issues (for example, it might not catch logical flaws or subtle coherence problems). It sometimes flags correct sentences as errors (false positives) or suggests stylistic changes that are debatable. For ESL students, blindly accepting all Grammarly suggestions can even introduce new errors if the student does not fully understand them. Mohammadi, Borna, and Zoghi (2023, p.196) found that while Grammarly helped Iranian EFL students correct surface errors, the students sometimes applied suggestions inappropriately due to lack of grammatical knowledge. Another limitation is that Grammarly's feedback, while extensive on micro-level issues, is comparatively weak on global writing aspects like organization and content development. It might comment on sentence clarity, but it will not tell a student if their argument is unconvincing or their examples are off topic. This is where teacher feedback still far outshines an automated tool.

From a pedagogical viewpoint, researchers suggest using Grammarly as a complement to teacher feedback, not a replacement. Sarıcaoğlu and Bilki (2021), in a Turkish EFL context, argue that such technologies *"should be used as supportive mediums – helping students enhance their language proficiency by providing feedback"*. In practice, a teacher might ask students to run their drafts through Grammarly to catch easy-to-fix errors, so that the drafts they submit are cleaner and the teacher can focus on higher-level feedback. This approach not only saves teacher time but potentially makes students more responsible for self-editing. Some classroom interventions have students reflect on Grammarly's corrections, turning it into a metacognitive exercise ("Why did Grammarly underline this phrase? Do I agree with the change?"). This can help develop their grammatical understanding.

In terms of attitudes, EFL students often report favorable perceptions of Grammarly. They appreciate its instant feedback and the privacy of learning from mistakes without public embarrassment. A study by Raihan et al. (2023) found that EFL postgraduate students viewed Grammarly feedback as useful and generally incorporated many of its suggestions in revisions. Nonetheless, students also recognized that Grammarly cannot fully substitute for teacher input, especially for content guidance and nuanced language use like appropriate academic tone.

Finally, it's worth noting that Grammarly (as of 2023) has introduced an AI-powered beta feature called Grammarly GO, aiming to provide more contextual suggestions (even ideas or rephrasings for clarity upon request), edging slightly into the territory of AI chatbots. The landscape of such tools is evolving. But in summary, Grammarly's core strength lies in grammar and mechanics feedback delivered quickly and in an easy-to-use format, making it a valuable tool for improving linguistic accuracy in EFL writing. Its weaknesses in higher-order feedback mean it's best used alongside human feedback. Teachers integrating Grammarly into their courses have reported that it offloads the "red ink" duty of marking every minor error, allowing them to concentrate their feedback on content and organization where human insight is irreplaceable.

2.3.2 Turnitin feedback studio: plagiarism detection and teacher-annotated feedback

Turnitin is primarily known as a plagiarism detection platform, but its Feedback Studio component provides a suite of tools for instructors to give feedback on student writing. When a student submits a paper to Turnitin, the system generates a Similarity Report highlighting any matched text from its vast database (internet sources, publications, and other student papers). This originality checking is a form of feedback in itself – often used to teach students about proper citation and to deter copy-paste plagiarism. However, beyond plagiarism detection, Turnitin Feedback Studio includes features for online grading and feedback, such as in-line comments, QuickMarks (pre-made comment templates for common errors), rubrics, and voice comments.

From a technological standpoint, Turnitin's automated feedback to students is mostly in the plagiarism domain (highlighting unoriginal text). It does not automatically correct grammar or give writing tips in the same way Grammarly does, but Turnitin has integrated ETS's e-rater engine to flag grammar and mechanics issues as annotations if enabled. Still, the heavy lifting of feedback in Turnitin is typically done by the teacher using the interface to place comments on the paper. In effect, Turnitin acts as a delivery system for teacher feedback, streamlining the process by having everything digital, trackable, and easily accessible to students.

Studies on Turnitin's impact often focus on how the Similarity Report influences student behavior. Many have found that students become more aware of paraphrasing and citation due to Turnitin, as they can see what percentage of their text is matched. Some instructors allow (or even encourage) students to pre-check their drafts via Turnitin, then revise to lower the similarity index – which can be a useful learning exercise in source use. Beyond plagiarism, a 2021 study by Hattie *et al.* (with involvement from Turnitin staff) examined which forms of feedback in Turnitin Feedback Studio were most predictive of student improvement. They analyzed thousands of essays with both teacher comments and automated comments, correlating them with the degree of improvement from first to final draft. One key finding was that "Where to next" feedback – comments that gave forward-looking suggestions – led to the greatest gains between drafts. In other words, feedback

that was oriented toward *how to improve in the next version* (e.g., "Clarify your main point in the next draft by adding an example in paragraph 2") was more powerful than feedback that was merely corrective or evaluative of the current draft. This aligns with broader feedback research emphasizing feedforward. The study also concluded that there is strong support for the *combination* of teacher and computer-generated feedback: *"There is support for the worthwhileness of computer-moderated feedback systems that include both teacher- and computer-generated feedback."*. This suggests that using Turnitin to provide automated checks (like similarity and basic grammar alerts) alongside detailed teacher comments can be effective, presumably because the automated part covers routine issues and the teacher can focus on substantive feedback.

Turnitin's influence in EFL settings, such as Türkiye, has also been noted in relation to academic writing and integrity. Turkish university students, especially at graduate level, often use Turnitin to ensure their theses and assignments do not inadvertently plagiarize. This fosters a habit of more careful paraphrasing and citation – an important aspect of writing quality.

One limitation of Turnitin is that students might fixate on the similarity percentage as a measure of success, possibly at the expense of deeper revision. It's not uncommon for a student to say "I got my similarity down to 5%" but still have a poorly structured essay. Therefore, teachers must guide students to use Turnitin's reports judiciously – focusing on the qualitative insights (e.g., which sections need reworking) rather than just the quantitative score.

Another concern is that Turnitin is primarily an instructor's tool; students do not receive *automated* formative feedback on things like coherence or argument strength. They rely on the teacher's input in the system. Thus, Turnitin's effectiveness is as good as the teacher's feedback practice. A conscientious teacher can use it to provide very rich feedback (embedding rubric criteria, linking to examples, etc.), whereas a less engaged teacher might only give minimal comments, which Turnitin cannot compensate for.

In terms of integrating Turnitin with other tools: some institutions use Turnitin in tandem with an AWE like Grammarly – Turnitin for plagiarism and teacher grading, Grammarly (or others) for grammar self-check. Each tool covers different bases.

Overall, Turnitin Feedback Studio is best characterized as a platform that supports efficient teacher feedback and upholds academic writing standards. It is an important tool in the ecosystem for teaching writing, particularly to address plagiarism (which is a form of feedback: informing students about originality) and to expedite the feedback cycle. For the purpose of this literature review's focus, Turnitin exemplifies how technology can assist teachers rather than replace them: it flags issues and provides an environment for feedback, but the expertise and nuanced feedback largely still come from the instructor. This is distinct from tools like Grammarly or ChatGPT which attempt to *generate* feedback content on their own.

2.3.3 ETS CRITERION: automated writing evaluation system

Criterion, developed by Educational Testing Service (ETS), is a prominent Automated Writing Evaluation (AWE) system that has been used in both classroom and testing contexts. It uses the ETS e-rater engine and other NLP tools to automatically score essays and provide diagnostic feedback on various writing traits. When a student submits an essay to Criterion, within seconds it returns an overall holistic score (on a scale, say 1–6) and trait scores (such as organization, development, grammar, style), along with specific feedback on grammar, usage, mechanics, sentence structure, and essay-based prompts on how to improve. Essentially, Criterion aims to mimic some functions of a human grader or tutor by giving immediate evaluation and pointing out areas for improvement.

Criterion has been the subject of many validation and classroom studies over the past two decades, particularly in ESL/EFL contexts. One key advantage is the speed and opportunity for multiple revisions – students can write a draft, get instant feedback, revise and resubmit, iteratively improving their essay (at least in theory) until they are satisfied or reach a score goal. This makes writing a more formative, practice-oriented process rather than one-and-done.

Research on Criterion in EFL classrooms reveals mixed but generally positive findings about improvement in writing mechanics and structure. An early study by Chodorow and Burstein (2004) found that students using Criterion over a semester reduced certain grammar error rates more than a control group. More recently, Hoang (2022) conducted a study in a Vietnamese EFL writing class using Criterion over five months. The results

showed that Criterion's feedback had *"satisfactory precision"* in identifying errors and it did trigger various engagement strategies among learners (e.g., students consulting grammar resources or peers to fix issues). This indicates the tool was fairly good at pinpointing actual errors (not too many false alerts) and that students actively tried to respond to the automated feedback. However, the study also found limitations: students' *response accuracy was modest and their revisions lacked substance despite the feedback*. In other words, learners would address some of the surface errors flagged by Criterion, but they often failed to make deeper content changes, and sometimes they misunderstood the automated feedback leading to incorrect revisions. This resonates with a common critique: AWE systems can lead to a focus on form over content, with students "satisficing" – making minimal changes to appease the system and get a better score, rather than truly improving the essay's ideas or coherence.

Another study, from Türkiye, by Turgay Han and Elif Sari (2022) examined combined use of Criterion and teacher feedback versus teacher-only feedback in university EFL writing classes. In a quasi-experimental design, one group of students received feedback from both Criterion and their instructor, while a control group only received the instructor's feedback. Interestingly, the study found that both groups improved their writing to a similar extent in overall analytic scores (meaning using Criterion did not shortchange general improvement). However, the group with combined automated-teacher feedback outperformed the teacher-only group in reducing grammar and mechanics errors. This suggests a complementary effect: Criterion, with its tireless error marking, helped students clean up language issues more effectively, while the teacher feedback ensured overall content and structure were addressed – resulting in comparable holistic improvement. Students in the combined condition got the best of both worlds in terms of accuracy. Qualitative feedback from the Turkish students indicated they found the Criterion reports helpful for identifying errors they hadn't noticed, though some noted frustration when the system's feedback was vague or when it didn't explain *why* something was wrong. The study's implications were that AWE can be integrated into EFL writing pedagogy to handle the labor of basic corrections and give students additional practice opportunities, but it shouldn't eliminate the need for teacher guidance, especially for higher-level writing skills and clarifying misunderstandings.

Some challenges with Criterion (and similar AWE) include: lack of contextual understanding (it may misjudge the quality of content or give generic advice), potential to be "gamed" (students might learn tricks to raise their score without genuinely better writing, e.g., adding irrelevant text to hit word count), and variable student attitudes (some students trust the system, others are skeptical or feel it's too impersonal). Instructors often have to train students on how to interpret Criterion's feedback – for example, if Criterion says "organization: needs improvement," the teacher might need to discuss with the student what specifically could be reorganized, since the tool's explanation might be limited.

Overall, ETS Criterion represents a mature generation of AWE tools that has proven its utility in providing rapid feedback and additional writing practice. It excels in consistency (every essay is judged by the same criteria) and can lighten teachers' load by addressing repetitive grammar issues. Academic studies, including those in EFL contexts like Vietnam and Türkiye, generally indicate that using Criterion leads to improvement in measurable aspects of writing accuracy and can be as effective as traditional feedback when combined thoughtfully. However, Criterion on its own may not drive substantial improvements in content depth or rhetorical quality. The human element – whether from a teacher or through guided peer discussions – remains crucial to interpret the feedback and push revisions to a higher level. The latest research suggests that the sweet spot is in blending automated and human feedback, leveraging the strength of each. As one review noted, such AWE tools should be seen as *"supportive mediums"* to enhance proficiency with feedback, not stand-alone teachers.

2.3.4 AI chatbots (e.g., ChatGPT): the new frontier in writing feedback

The newest and most transformative entrant into the field of writing feedback is the AI chatbot, epitomized by OpenAI's ChatGPT (particularly the GPT-3.5 and GPT-4 models introduced in 2022–2023). Unlike traditional AWE systems that operate on fixed algorithms and provide static feedback, AI chatbots leverage advanced large language models (LLMs), enabling them to engage in dynamic, context-aware conversations with users about their writing. This capability marks a significant evolution: instead of just highlighting errors or scoring an essay, a chatbot can *discuss* an essay with the student,

answering questions, explaining corrections, and even providing examples or counter-suggestions. Essentially, an AI chatbot has the potential to function as a virtual writing tutor available on demand.

ChatGPT and similar LLM-based tools can perform a variety of feedback-related functions: they can check grammar and suggest rephrasings (like Grammarly), evaluate coherence and structure (e.g., "Does my introduction clearly state a thesis?" – the AI can give an opinion), provide ideas for development, and even role-play as an audience to give reader perspective. Importantly, because the interaction is conversational, the student can ask follow-up questions, making the feedback process iterative and tailored, aligning well with the dialogic feedback principles discussed earlier. For example, a student might prompt, "Provide feedback on this essay's content and organization," receive a detailed paragraph from ChatGPT, then follow up with, "What did you mean by unclear reasoning in paragraph 3?" to get clarification.

The emergence of ChatGPT in late 2022 triggered a flurry of research in 2023-2024 examining its potential in education, including writing instruction. Initial studies paint a picture of both great promise and areas of concern. On the positive side, ChatGPT can produce comprehensive feedback covering multiple dimensions of writing. Steiss et al. (2023) compared ChatGPT's feedback on student essays to feedback given by experienced human EFL teachers. They found that *ChatGPT produced significantly more feedback comments than the teachers*, and notably, while teachers tended to focus the bulk of their feedback on content issues, ChatGPT provided a more balanced distribution of comments on content, organization, and language. This suggests the AI was thorough in addressing different aspects of writing. For instance, a human teacher with limited time might primarily critique the argument logic and not comment much on grammar unless it was egregious, whereas ChatGPT would comment on a logical gap, a confusing sentence, and a grammar mistake all in one go. From a student's perspective, this breadth of feedback can be very useful – it's like getting a holistic review.

Teachers in that study were then asked to judge the quality of ChatGPT's feedback. They acknowledged some strengths: ChatGPT's comments were often valid and pointed out issues similar to those identified by teachers, indicating the AI can accurately diagnose

many writing problems. However, teachers also noted weaknesses. Sometimes ChatGPT's feedback lacked depth in explaining *why* something was an issue or *how* to improve it, whereas a teacher might provide a clearer rationale or a more concrete suggestion. There were also instances where the AI's feedback might be off-mark due to lack of true understanding (e.g., it might misinterpret a subtle argument or cultural reference).

Another line of inquiry is how students actually use ChatGPT's feedback. Zou et al. (2025) conducted a comparison of students' *uptake* of feedback from teachers versus ChatGPT. While that study is very recent (published in *Computer Assisted Language Learning* in 2024/25), preliminary descriptions indicate a *"high uptake rate of ChatGPT feedback"*, meaning students implemented a large portion of the chatbot's suggestions in their revisions. In fact, they reportedly found that students followed ChatGPT's suggestions at a comparable rate to teacher suggestions. This underscores that, at least for willing students, AI feedback can be as actionable as teacher feedback. Students might enjoy the clarity and prescriptive nature of ChatGPT's comments (e.g., it might directly propose a revised sentence, making it easy to adopt). However, a potential downside is if students accept AI feedback uncritically – there is a risk of introducing changes that sound fluent but may not reflect the student's authentic voice or intent.

The comparative effectiveness of ChatGPT vs. teacher feedback is not yet definitively established, but early evidence suggests each has different strengths. A study by Lin and Crosthwaite (2024) intriguingly titled "The grass is not always greener: Teacher vs. GPT-assisted WCF" compared quality and characteristics of WCF by teachers versus GPT. It noted that while human feedback was generally better in some qualitative aspects (more contextualized, more selective in targeting important issues), the *differences were not always significant* and ChatGPT feedback was surprisingly on par in many areas. One practical edge of ChatGPT is its speed and availability – it can generate feedback almost instantly, and students can use it anytime as a 24/7 tutor. This democratizes access to feedback beyond the classroom and office hours.

Researchers like Barrot (2023) have commented on this potential, noting *"quality feedback is not always available to students; ChatGPT addresses this need as it offers*

useful features linked to writing assessment", such as automatic scoring and specific comments on strengths and weaknesses. By prompting ChatGPT appropriately (for example, "Grade this essay and provide feedback" or "Give me feedback focusing on how to improve clarity"), students can get a form of evaluation and constructive pointers in seconds. Barrot further explains that ChatGPT can highlight what is done well and what can be improved, essentially mimicking some duties of a teacher or writing center tutor in providing *formative assessment*.

However, alongside the enthusiasm, cautionary findings have emerged. One concern is that ChatGPT's feedback may sometimes be superficial. For instance, Yoon et al. (2023) pointed out that older AWE systems struggled with content and coherence feedback, and while ChatGPT is better, it might still focus on *surface-level features*. A study using an English Learner corpus (ELLIPSE) to test ChatGPT's feedback on coherence found that the AI could identify use of cohesive devices (transitions, etc.) and there was a correlation with human coherence scores, but ChatGPT was *"not particularly effective in terms of coherence and cohesion"* when it came to giving actionable steps for improvement. In plainer terms, ChatGPT might say "your ideas could flow better" but not clearly pinpoint how to reorganize them for logical flow, whereas a teacher might give a specific suggestion ("Consider moving the second paragraph later and adding a topic sentence to link back to your thesis").

Accuracy of feedback is another issue. ChatGPT can sometimes be overly accommodating or even incorrect. If a student writes a flawed claim, a teacher would likely point it out as incorrect or unsupported. ChatGPT, depending on prompt, might hedge or not fully catch logical inaccuracies unless explicitly asked to check facts or logic. Conversely, if asked to be critical, ChatGPT might generate some criticism even if an essay is quite sound (since it tries to comply with the instruction to provide critique). Thus, the *calibration* of feedback can vary, whereas teacher feedback usually draws on actual understanding and priorities (for example, a teacher might ignore a trivial grammar mistake to focus on a major content flaw, something an AI might not know to do unless told).

Another challenge is student trust and interpretation of AI feedback. Some students might over-trust AI suggestions, as noted, implementing them without discernment. Others might under-trust, dismissing AI feedback because "it's just a bot." Developing students' ability to evaluate the AI's comments is vital. In one anecdotal account, university students initially doubted ChatGPT's grammar corrections until they checked credible sources to confirm the bot was right. This indicates a learning curve in accepting AI as a valid source of feedback.

Notably, ethical and academic integrity considerations also come into play. If students use ChatGPT not just for feedback but to generate content (a separate but related issue), it raises concerns. However, for feedback usage specifically, one must consider if heavy AI guidance might homogenize writing style or reduce the struggle that leads to learning. Teachers in some contexts might limit AI feedback usage for draft assignments meant to develop students' self-reliance.

In summary, AI chatbots like ChatGPT represent a game-changer in EFL writing feedback. They combine some strengths of earlier tools (speed, focus on form) with a new ability to discuss and explain, moving closer to human-like feedback interactions. Early research underscores: (1) ChatGPT can provide extensive and multi-faceted feedback, often aligning with teacher feedback on identifying issues ; (2) Students can and will use this feedback, sometimes as much as they use teacher feedback, indicating high uptake; (3) The best results might come from *integrating* ChatGPT with teacher input, rather than relying on it alone. For instance, a teacher might instruct students to get initial feedback from ChatGPT on a draft, implement some changes, and then submit to the teacher for final feedback – a layered approach that leverages AI for first-stage improvements and the teacher for advanced guidance. Indeed, a study on combining ChatGPT with teacher feedback found that such integration provided more individualized feedback and could enhance writing skills effectively.

As research from 2024 onward accumulates, how AI feedback and human feedback can complement each other, and what balance yields the best learning outcomes will better be understood. The next section will delve specifically into comparative findings between teacher feedback and AI-based feedback, drawing on both the tools reviewed and the

theoretical insights earlier, with attention to the EFL context and any evidence from Türkiye or similar settings.

2.4 Comparative Perspectives: AI-Based Feedback vs. Teacher Feedback in EFL Writing

Having examined the nature of teacher feedback, theoretical frameworks, and the capabilities of current AI feedback tools, we now turn to studies that directly compare AI-generated feedback with teacher-provided feedback. The goal is to highlight similarities and differences in their content, effectiveness, and reception by students, particularly in the context of EFL writing. Scenarios where they are used in combination will also be considered. Key themes emerge in the literature: differences in focus (language vs content), volume and specificity of feedback, the quality of interaction (static comments vs. dialogue), student preferences and trust, and outcomes in terms of writing improvement.

2.4.1 Focus and scope of feedback: form-focused vs content-focused

One clear difference noted earlier is that AI tools, especially AWE systems like Grammarly or Criterion, excel at form-focused feedback (grammar, spelling, mechanics), whereas teachers often put more emphasis on content and ideas. This was evidenced in Steiss et al. (2023) comparison: teachers devoted more of their feedback commentary to the substance and argumentation of the essays, while ChatGPT gave substantial attention to those but also evenly covered grammar and organization. In essence, AI feedback tends to be comprehensive in coverage, including minor language issues that a teacher might overlook or decide to ignore for the bigger picture. Teachers, constrained by time, frequently prioritize – they might leave a few comma splices unmentioned if the overarching thesis is weak and needs the commentary space. Therefore, a student receiving AI feedback might see a longer list of issues (some major, many minor) compared to a teacher's feedback which might highlight fewer issues but possibly in more depth.

From a student's perspective, the breadth of AI feedback can be both a blessing and a curse. On one hand, they become aware of all sorts of errors (catching those overlooked

by the teacher), potentially learning more. On the other hand, they could feel overwhelmed or may not know which feedback points are most crucial. Teachers often act as filters, deciding what is important for the student's level and assignment goals. AI currently lacks that pedagogical judgment unless specifically prompted. For instance, a teacher might think, "This student is still struggling with basic paragraph structure; I won't nitpick their citation format this time," whereas an AI checker will mark every missing comma in a reference list.

However, new research suggests AI's focus can be adjusted. For example, if a student asks ChatGPT, "Ignore grammar and just tell me if my arguments make sense," the AI can tailor its feedback accordingly. This implies that savvy users can overcome the one-size-fits-all nature of AI feedback by directing it, something not possible with static AWE tools in the pre-chatbot era (Barrot, 2023, p. 3).

2.4.2 Volume and specificity of feedback

Studies consistently find that AI (especially ChatGPT) can provide more feedback comments in quantity than human teachers. This sheer volume is partly due to AI not being subject to fatigue or time constraints,it can comment on every sentence if asked (Wilson & Roscoe, 2020; El-Dakhs & Alghizzi, 2022). Teachers, no matter how diligent, have limits; they often summarize issues to avoid marking every instance of a repeated error. AI might flag all 10 occurrences of a misused article, while a teacher underlines two and writes "article usage needs work" as a global comment (Barrot, 2023).

The specificity of AI feedback can sometimes be impressive: it might suggest a rephrasing for a sentence that is awkward, whereas a teacher might just note "awkward phrasing." In this sense, AI can give very concrete recommendations. For example, Grammarly might underline a clause and say "Consider revising to: …" giving a full suggestion, and ChatGPT can do similarly if prompted ("How can I rephrase this sentence more academically?") (Barrot, 2023; Farrokhnia et al., 2023). Teachers often expect the student to figure out the rephrase once an issue is pointed out, to promote learning (Hyland & Hyland, 2006).

On the flip side, teachers can be more context-specific in feedback. They know what was taught in class, the intended learning outcomes of the assignment, and the individual student's prior performance. So a teacher might say, "In this essay you needed to compare X and Y, but you've only described X. Remember how we practiced comparison in class,try to apply that here." AI would not know that context unless fed with it. Teacher feedback can also tie into future tasks ("When you do the next assignment, focus on integrating sources better as we discussed"), which AI will not do on its own (Truscott, 1996; Lee, 2017).

There is an interesting finding from a recent meta-analysis by Wisniewski et al. (2020) (spanning broad educational feedback, not just writing): feedback is more effective when it contains more information, which favors the detailed nature of AI feedback, but also when it avoids simply giving the answer or doing the task for the student. Good teacher feedback often provides hints or raises questions rather than full solutions, to stimulate thinking. AI's suggested rewrites might inadvertently short-circuit that learning opportunity if students just accept them. That said, one could also use AI Socratically – for instance, asking ChatGPT, "Is there another way to support my point in paragraph 2?" rather than "Rewrite paragraph 2 for me." The former fosters thinking with guidance, akin to a tutor's questioning, whereas the latter is more like outsourcing the solution.

2.4.3 Interaction and dialogue

One of the biggest distinctions is the interactive potential. Traditional teacher-written feedback is often one-directional; students rarely respond to each comment in a written dialogue (though they could in theory or in writing conferences). ChatGPT and similar chatbots enable real-time dialogue about the text. A study by Shao et al. (2023) observed that students who used an AI chatbot for iterative feedback engaged in a kind of self-dialogue: they would ask the bot to explain feedback, ask for examples, etc., leading to multiple turns of revision. In contrast, students who only had teacher-written feedback often did a single round of revisions without deeper inquiry, unless they actively sought the teacher for clarification (which many did not).

This dialogue capability is probably the single most promising feature of AI chatbots for learning. It effectively gives every student a personal tutor that they can "talk" to as much

as needed (Farrokhnia et al., 2023; Wilson & Roscoe, 2020). For example, a Turkish EFL student might be too shy to ask the teacher why their introduction is weak, but they might feel comfortable asking ChatGPT and getting a detailed explanation,or even ask it to "show me a good introduction on this topic." That said, the quality of this dialogic engagement depends on student skill in prompting and the reliability of the AI's responses. Teachers often have to train students in peer review or self-assessment; similarly, they may need to train students in how to query AI effectively and critically (Barrot, 2023; Hyland & Hyland, 2006).

2.4.4 Student perceptions and preferences

Student surveys so far present a nuanced picture. Many students appreciate AI feedback for its speed and availability – it reduces stress to know they can check their work anytime. In contexts like Türkiye, where teacher feedback might be constrained by large classes, having an AI assistant might alleviate student anxiety about not getting enough feedback. In the Saudi study mentioned, while teacher feedback was top-preferred, students showed *positive attitudes toward automated feedback* in general, viewing it as helpful extra input (aside from technical glitches or content limitations). Similarly, a recent survey in China found students valued ChatGPT's feedback especially for grammar and vocabulary, but still trusted teacher feedback more for content accuracy and overall evaluation (Zhang et al., 2023).

A key point is that students often do not want to choose either/or between teacher and AI feedback – they want both. Each meets different needs. Teacher feedback provides a human touch, encouragement, and credibility, as well as alignment with course goals (students know the teacher will grade them, so teacher feedback is seen as directly relevant to scoring). AI feedback provides immediacy, a chance to self-improve before facing evaluation, and as noted, can be used without fear of judgment. In some cases, weaker or more anxious writers actually prefer getting AI feedback first, to "save face" by fixing obvious errors, and then submitting to the teacher, a pattern noted by Huawei and Aryadoust (2024) in an experimental class.

However, some students express skepticism about AI feedback accuracy or depth. There's also a learning preference aspect: some learners find the impersonal nature of AI

uninspiring and may ignore those suggestions, whereas a teacher's comment carries weight. Conversely, a few might over-rely on AI, assuming it's always correct. In either case, managing expectations is important. Several studies recommended explicitly discussing the role of AI tools with students – positioning them as tools, not authorities, and integrating their use into the curriculum so students feel guided.

2.4.5 Writing improvement outcomes

The ultimate measure is: do students who use AI feedback improve their writing as much as, or perhaps even more than, those who rely solely on teacher feedback? The evidence is still emerging, but so far indicates that AI feedback can lead to tangible improvements in writing, especially for linguistic accuracy, on par with teacher feedback in certain scenarios. The Turkish study by Han & Sari (2022) showed no significant difference in overall writing score gains between the teacher-only group and the combined group, with the combined group excelling in error reduction. This implies AI did not hinder progress and even gave an edge in grammar. Another study in Taiwan (Apriani et al., 2024) found that an experimental group using an AI chatbot for feedback outperformed a control group in post-test writing complexity and organization (though not in content), suggesting AI input helped them develop more structurally sound essays.

That said, no study has shown AI feedback *replacing* teacher feedback entirely without any drop in quality of learning. Most positive reports involve blended feedback approaches: AI used to supplement or complement, rather than substitute, the human feedback. When researchers tried a fully automated feedback course (no teacher comments, just AWE), the outcomes were less impressive in terms of content development and student satisfaction (Wilson & Czik, 2016). It appears that teacher expertise is still crucial for higher-order writing skills and motivation.

One interesting insight comes from Hattie et al. (2021) who emphasized *feedforward (where to next)* comments. AI can be explicitly prompted to give feedforward (e.g., "What should I work on to improve this draft further?"). If used that way, it might approximate effective teacher coaching. The combination of teacher and AI can ensure that students get both the big-picture direction (often from the teacher) and the nitty-gritty fixes (from the AI).

In Turkish EFL classrooms specifically, writing instructors are exploring such combinations. Anecdotally, some universities in Türkiye have begun pilot programs where students draft essays in a platform that provides immediate AI feedback, then they revise and submit to the teacher. Preliminary feedback from instructors suggests students' second drafts (after AI input) are noticeably cleaner, allowing the teacher to focus comments on argumentation and style. This aligns with the findings of reduced language errors with AWE support in the region.

Finally, from the perspective of sociocultural and dialogic theory, teacher feedback offers real social interaction and mentorship, which is hard to fully replicate. However, AI can serve as a *mediational tool* too. It's not a human interlocutor, but it can mediate the student's thinking process. For instance, an AI might not *truly understand* the student's intent, but the act of explaining their writing to the AI in prompts could itself help students clarify their ideas – akin to the self-explanation effect in learning. Dialogic feedback ideally involves negotiation of meaning; with AI, the negotiation is artificial, but it still forces the student to put their thoughts into words to get better answers from the AI.

3. METHODOLOGY

This chapter outlines the methodological framework and procedures employed in the present study, which investigates the comparative effectiveness of AI-based chatbot feedback versus teacher-written feedback in an EFL writing class. It begins by presenting the research questions and design and rationale, followed by a detailed description of the participants and setting. Next, the data collection instruments and procedures are described, including how student writing samples were collected and how feedback was generated. The chapter then explains the data analysis procedures (discourse analysis of feedback and thematic analysis of interviews), along with steps taken to ensure trustworthiness. The methodology is aligned with the study's original research questions, which are preserved here for clarity.

3.1 Research Questions

The study is guided by three research questions formulated to compare AI-based chatbot feedback with teacher-written feedback in an EFL writing context. These questions address both the nature of the feedback provided by each source and the perceptions of those who use or give such feedback:

RQ1: How does the effectiveness of AI-originated written feedback compare to teacher-written feedback in terms of identifying and addressing errors in EFL writing?

RQ2: How do students perceive and evaluate the quality, clarity, and usefulness of feedback from both AI-originated and teacher-written sources?

RQ3: How do teachers perceive and evaluate the quality, clarity, and usefulness of feedback from both AI-originated and teacher-written sources?

These questions align with the study's aim of exploring both objective differences in feedback content and the subjective experiences of students and teachers. By addressing these questions, the research seeks a holistic understanding of how AI feedback and teacher feedback compare in an EFL writing class.

3.2 Research Design

This study employs a qualitative research design, which is appropriate for examining the contextual, interpretive, and affective dimensions of written feedback in second language writing. The qualitative paradigm allows for an in-depth understanding of how feedback is constructed, delivered, and interpreted within a real classroom setting (Dörnyei, 2007; Merriam, 2009). The research specifically focuses on meaning-making processes and subjective experiences, rather than quantifying feedback outcomes or generalizing findings statistically.

The study is structured around two core qualitative analysis methods: discourse analysis and thematic analysis, each aligned with specific research questions.

Discourse Analysis (For RQ1)

To address Research Question 1, which investigates the nature and communicative function of feedback provided by AI tools and teachers, a qualitative discourse analysis was conducted on the written comments given to student texts. Discourse analysis was selected for its ability to explore not just the content of language but also the intentions, power dynamics, tone, and pedagogical roles embedded in written discourse (Gee, 2014; Paltridge, 2012).

Feedback texts from five different sources (ChatGPT, Grammarly, and three teachers) were treated as discourse events situated in pedagogical practice. Each comment was analyzed at two levels:

1. Focus: What aspect of writing it targeted (e.g., grammar, content, organization, vocabulary).
2. Function: What communicative move it performed (e.g., direct correction, suggestion, praise).

While frequency counts of categories were presented descriptively to support interpretation (e.g., number of content-focused vs. grammar-focused comments), the analysis prioritized qualitative interpretation of tone, depth, personalization, and instructional intention. For example, comments like “Consider adding an example here”

(suggestion-function) were compared to AI-generated corrections such as “Change ‘has did’ to ‘has done’” (mechanical correction) to identify stylistic and pedagogical patterns across sources.

Thematic Analysis (For RQ2 and RQ3)

To address Research Questions 2 and 3, which explore how students and teachers perceive the feedback, a thematic analysis was conducted on interview transcripts. This method was chosen for its suitability in identifying recurring ideas, emotional responses, and subjective evaluations in participants' narratives (Braun & Clarke, 2006; Nowell et al., 2017). Interviews with students (addressing RQ2) capture their feelings about and responses to the feedback they received, while interviews with teachers (RQ3) capture professional reflections on the efficacy and clarity of AI vs. human feedback. The thematic anaysis is essential for understanding not just what differences exist in feedback, but why those differences matter in practice (e.g., how students act on AI feedback versus teacher feedback).

The thematic analysis followed Braun and Clarke’s six-phase model:

1. Familiarization with data: Interview transcripts were read repeatedly to immerse in the data.
2. Generating initial codes: Meaningful data chunks were systematically coded across all interviews.
3. Searching for themes: Codes were grouped into candidate themes based on conceptual similarity.
4. Reviewing themes: Themes were checked for coherence and distinctiveness.
5. Defining and naming themes: Each theme was clearly defined to reflect the core message.
6. Producing the report: Final themes were written up with supporting quotations from participants.

An inductive approach was used, allowing themes to emerge naturally from the data rather than imposing a pre-existing framework (Patton, 2002). Themes were not only based on repetition but also on emphasis, intensity, and emotional tone, reflecting the affective and contextual dimensions of feedback (Boyatzis, 1998).

For instance, students repeatedly emphasized the clarity and emotional impact of feedback, while teachers highlighted issues of trust, pedagogical control, and the limitations of AI tools. These themes were interpreted in light of existing literature, ensuring analytical depth and theoretical grounding.

By combining these two qualitative approaches, the study offers a multi-layered understanding of written feedback, both in terms of what it looks like on paper and how it is received, processed, and acted upon by learners and educators.

Ethical Considerations: The study design was approved by the university's ethics committee. Prior to data collection, all participants were informed about the research purpose and procedures, and written consent was obtained. Participants were assured of anonymity and confidentiality; pseudonyms or codes are used (students are coded as ST1, ST2, … ST15, and teachers as T1, T2, T3) to protect identities. They were also informed of their right to withdraw at any time without penalty (as stated in the informed consent form). These measures ensured compliance with institutional ethical standards and APA 7 guidelines for research with human participants.

3.3 Participants

The study involved two groups of participants: EFL students who received both AI and teacher feedback on their writing, and EFL teachers who provided or evaluated feedback. The sampling strategy was purposive, driven by the goal of including participants with direct experience of the feedback under comparison.

This study employed purposive sampling, a non-probability sampling technique commonly used in qualitative research to intentionally select participants who possess specific knowledge or experience relevant to the research aims (Palinkas et al., 2015). Both students and teachers were chosen based on their direct engagement with written

feedback, students who had received AI-based and teacher feedback on their writing, and teachers who had experience giving feedback and/or familiarity with AI tools. This criterion-based approach ensured that the data collected would be information-rich and closely aligned with the research questions. Purposive sampling was particularly appropriate given the exploratory nature of the study and its focus on detailed, context-dependent understanding rather than statistical generalization.

Participant Selection: Participants were selected based on defined inclusion criteria to ensure the relevance and richness of the data. For students, eligibility required:

1. current enrollment in the university's English preparatory program,
2. completion of at least one written assignment that received feedback from all five sources (three teachers, ChatGPT, and Grammarly),
3. willingness to participate in an interview reflecting on their feedback experience.

For teachers, the inclusion criteria were:

1. active involvement in teaching academic writing within the program,
2. experience in providing written feedback to EFL learners,
3. at least basic familiarity with AI-based tools used in writing instruction.

These criteria ensured that participants could provide informed, experience-based responses relevant to the study's focus. Individuals who did not meet these conditions, such as students without exposure to all feedback types or teachers unfamiliar with AI feedback, were excluded from the study.

Students (n = 15): The student participants were 15 adult EFL learners in an upper-intermediate English writing course at a university preparatory program in Istanbul. There were 9 female and 6 male students, with ages ranging from 18 to 22 years. All students were non-native English speakers (L1 Turkish), studying English for academic purposes. Their English proficiency levels ranged from B1+ to B2 (intermediate to upper-intermediate), based on placement tests in the program. These students were chosen because they regularly produce academic essays in English and thus benefit from feedback to improve their writing. All 15 students had experience receiving teacher-written feedback as part of their coursework. Prior to this study, only a few had used AI-

based writing tools; most were only vaguely familiar with Grammarly or ChatGPT and had not systematically used them for writing feedback. This context ensured that the comparison between AI and teacher feedback would be novel and salient for nearly all student participants.

Sampling and Recruitment: Participants were recruited via convenience sampling from the researcher's institution. An announcement was made in writing classes inviting volunteers for a study on writing feedback. The first 15 students who volunteered and met the criteria (intermediate+ English level, completed at least one essay assignment) were selected. Participation was voluntary with no grade incentives. Students were informed that the study would involve having an extra feedback session (via AI tools) on their paragraph and a follow-up interview. They were assured that the AI feedback exercise was for research and would not affect their course grade.

Participant Anonymity: In data presentation, each student is referred to by a code (ST1 through ST15) rather than name. The ordering (1–15) was assigned based on the sequence of paragraph submissions and does not reflect any performance ranking. No identifying personal details are included in the thesis, and individual quotes from interviews are attributed only by these codes.

Each student's paragraph received feedback from five distinct sources:

ChatGPT (AI-based chatbot using OpenAI's GPT-4)

FDT1: Teacher 1 – 35-year-old female with over 10 years of EFL teaching experience, not currently teaching the students

FDT2: Teacher 2 – 24-year-old male, early-career EFL teacher with 2 years of experience, also not teaching the students

FDT3: Teacher 3 – 34-year-old female with over 10 years of experience and the current course instructor of the student participants

Grammarly (commercial grammar-checking tool)

Teachers (N = 15): 15 teachers participated in the interview phase, making a total of 15 instructors coded as T1 to T15. These instructors varied in terms of teaching experience, institutional context, and familiarity with AI feedback tools. The teachers giving feedback aren't involved in interview process. The teacher providing feedback (FDT3) was the students' regular course instructor and the other interviewed teachers did not know the students at all.

The study involved two distinct participant groups: 15 EFL students and 15 EFL instructors. The decision to include 15 participants in each group was guided by the principle of thematic saturation, the point at which further data collection no longer yields new themes (Guest et al., 2006). Saturation was observed after approximately 12 interviews in each subgroup; however, 3 additional interviews were conducted to confirm thematic consistency. Given the relative homogeneity within each group (students enrolled in the same preparatory program; teachers with comparable professional roles), and the focused nature of the interview topics, this sample size was sufficient to generate rich, layered data. Including both groups allowed for triangulation of perspectives, supporting a more robust interpretation of how AI and teacher feedback are experienced and perceived in EFL writing contexts. This is consistent with qualitative recommendations that suggest 12–20 interviews per homogeneous group are generally adequate to reach saturation (Saunders et al., 2018; Braun & Clarke, 2021).

3.4 Data Collection

In this part, data collection tools and procedure have been clearly explained.

3.4.1 Instruments and Materials

Multiple data collection instruments were employed, corresponding to each phase of the study:

Student Writing Samples: The primary material for analysis was a set of 15 student writing samples. Each student provided one piece of writing – specifically, a short paragraph (approximately 150 words) written as part of their EFL writing class. The topic of these paragraphs is the same, social media's psychological effect on young adults. These paragraphs served as the common text on which different feedback sources would

be applied. The rationale for using the same student texts to generate both AI and teacher feedback was to enable a direct content comparison: each paragraph received feedback from three sources (teacher, ChatGPT, Grammarly), making it possible to compare how each source responded to the *exact same writing*. The student paragraphs were collected in electronic form (Word documents) to facilitate sharing with the teacher and input into the AI tools.

AI Feedback Tools: Two AI-based feedback providers were used:

ChatGPT: an AI chatbot (based on GPT-4) was used to generate feedback in a conversational style. An OpenAI ChatGPT interface was employed. The researcher input each student's paragraph into ChatGPT with a prompt such as: *"Provide feedback on this student's English paragraph. Identify any errors or weaknesses and suggest improvements in content, organization, and language."* The prompt was kept consistent for all paragraphs to ensure comparability. ChatGPT then produced a written feedback report for each paragraph, typically one paragraph long. These outputs were saved for analysis. Each ChatGPT feedback generally included an initial positive remark and several suggestions addressing different aspects (e.g., pointing out grammar mistakes, suggesting clearer transitions, commenting on argument strength).

Grammarly: a popular AI-driven writing evaluation tool focused on grammar and style. The premium version of Grammarly was used by uploading each paragraph into the Grammarly editor. Grammarly automatically generates suggestions inline, highlighting issues in grammar, spelling, punctuation, conciseness, clarity, and tone. For each paragraph, the researcher compiled Grammarly's feedback by extracting all the suggestions and explanations it provided (e.g., "Replace 'explications' with 'explanations' for clarity" or "Consider using present tense for consistency"). These were collated into a list of feedback points for analysis. Unlike ChatGPT's narrative feedback, Grammarly's output is a set of specific corrective annotations on the text. Nonetheless, for comparison, Grammarly's output was treated as a collection of feedback comments.

Both AI tools were used in July 2024, ensuring that the versions were up-to-date (ChatGPT using the GPT-4 July 2024 model; Grammarly's latest algorithm as of 2024). The AI-generated feedback was not edited by the researcher, except to remove any

identifying information (if, for instance, a student's name appeared in text). This preserved the authenticity of the AI feedback.

Teacher Feedback: 3 teachers as specified before provided written feedback on each of the 15 paragraphs. All three teachers provided handwritten feedback on printed versions of student paragraphs. They underlined errors, wrote suggestions in margins, and added overall comments. These physical papers were later scanned for analysis. No electronic marking tools were used. The teacher had approximately one week to provide feedback on all pragraphs. To maintain consistency, a feedback rubric (the same criteria given to ChatGPT in the prompt: content, organization, language accuracy) was shared with the teacher, but the teacher was free to prioritize issues as they saw fit for each individual paragraph. The resulting teacher feedback documents were collected by the researcher. These documents served as the "teacher feedback" samples for analysis and for showing to students.

Interview Protocols: Two semi-structured interview forms were developed: one for students and one for teachers. Each form contained a series of open-ended questions designed to address RQ2 and RQ3 respectively. The Student Interview Form (see Appendix 1) included questions such as: "Which feedback (AI or teacher) helped you more with your writing errors? Why?", "Did you find the AI feedback clear and easy to understand?", "How helpful do you think the AI feedback will be for your future writing?", "Were there any feedback comments that were particularly confusing or particularly helpful? Can you give examples?", and an invitation for any additional thoughts on the feedback process (see sample questions in the Structured Interview Form). These questions were carefully worded to prompt students to reflect on both positive and negative aspects of each feedback source. The Teacher Interview Form (Appendix 1) contained questions mirroring those asked of students but from a teacher's perspective: e.g., "After reviewing the feedback on the student writings, which source (AI or teacher) do you believe was more effective in identifying and addressing errors? Why?", "How would you rate the clarity and comprehensibility of the AI feedback? … of the teacher feedback?", "Which source provided more relevant comments for the student's writing?", "How valuable do you believe AI-generated feedback is for students' improvement? … and teacher feedback for improvement?", and "Were there specific

comments that stood out as effective or ineffective? Please specify source." These questions were formed by issues identified in the literature (such as clarity of feedback, relevance, higher-order vs. lower-order concerns) and by the research questions of the study. Both interview forms were reviewed by two experts in ELT research (the thesis advisor and one faculty member) for content validity and clarity. Minor adjustments were made (e.g. simplifying wording for student interviews) based on their feedback.

Each question was carefully formulated to elicit detailed reflections corresponding to the study's research questions. For example, questions such as *"How did you feel about the feedback you received?"* and *"Was the feedback easy to understand or apply?"* were designed to address RQ2, focusing on the perceived usefulness and emotional response of students. Similarly, teacher interviews included prompts like *"How would you compare AI and human feedback in terms of clarity or instructional value?"*, which mapped onto RQ3, exploring teacher judgments about feedback effectiveness and pedagogical impact.

In constructing these questions, reference was made to existing literature on written corrective feedback (e.g., Hyland & Hyland, 2006; Carless & Boud, 2018), ensuring that dimensions such as clarity, uptake, trust, and depth were incorporated into the protocol. Open-ended formats were used to allow participants to elaborate freely, while thematic links ensured that responses could later be systematically coded and compared across data sources.

3.4.2 Data Collection Procedure

Data collection took place in three stages corresponding to the instruments above: obtaining the writing samples and generating feedback (Stage 1), distributing feedback to participants (Stage 2), and conducting interviews (Stage 3).

Stage 1: Collection of Paragraphs and Generation of Feedback (June – July 2024)

Students provided their written paragraphs as part of a scheduled class assignment. The researcher compiled the 15 paragraphs for analysis. For each paragraph, two parallel feedback processes were initiated:

Teacher feedback: The paragraphs (anonymized copies) were sent to all three teacher FDT1, FDT2 and FDT3 who agreed to provide written feedback for research purposes. The teachers were given one week to mark up all paragraphs. The researcher received the commented documents and scanned them.

AI feedback: The researcher input each paragraph into ChatGPT and Grammarly as described. ChatGPT's conversational feedback was copied from the chat interface into a document. Grammarly's suggested edits were downloaded by saving the edited version of the paragraph with all suggestion highlights visible, and simultaneously extracting a list of suggestions. In practice, for clarity, the researcher summarized Grammarly's output as a list of discrete feedback points (since Grammarly's interface does not output one narrative text). This summary was carefully checked against the Grammarly interface to ensure no suggestion was missed or misrepresented.

By the end of Stage 1, for each of the 15 student paragraphs, there were five sets of feedback prepared: three versions of the teacher-written feedback (embedded in the student's paragraph as comments/edits), the ChatGPT feedback , and the Grammarly feedback.

Stage 2: Distribution of Feedback to Students (July 2024)

In order for students to compare and reflect on the feedback, each student was provided with their paragraph accompanied by both the teacher feedback and the AI-generated feedback. The researcher organized individual feedback review sessions. In these sessions, the student sat with the researcher to go over the feedback:

First, the student was given the teacher feedback document (their paragraph with teacher comments). The researcher provided the students with a printed copy with comments in the margins. The student read through the teacher's feedback. The researcher answered any clarification questions if the student was confused by a comment (though this was rarely needed as teacher comments were designed to be clear).

Next, the student was given the ChatGPT feedback text and the Grammarly feedback list for the same paragraph. For convenience, the AI feedback was compiled into a single document: the ChatGPT feedback was listed first, followed by Grammarly's key

corrections and suggestions. The student had time to read through these. Again, the researcher clarified any points as needed (for example, if a student was unsure why Grammarly suggested a certain change, the researcher could explain the grammar rule).

The order of presentation was counter-balanced to avoid order bias: roughly half the students (8 of 15) saw the teacher feedback first then AI feedback, as described, while the other half were first shown the AI feedback and then the teacher's comments. This was randomized to ensure one type was not always given primacy in their impressions. Students were not told upfront which feedback came from AI and which from a human teacher, to observe their initial reactions without bias. They were labeled neutrally as "Feedback 1, 2,3,4 and 5" during the review. After reading all, the researcher revealed which was which and then proceeded to the interview questions. (All students correctly guessed which was teacher vs. AI feedback after reading them, likely due to differences in style and format, but thi was confirmed explicitly before interviewing.)

It is important to note that the students did *not* revise their paragraphs between receiving the feedback and the interview. The focus here was on their perception of the feedback itself, not on measuring actual improvement. However, many students (n=10) naturally commented on how they would use (or did use) the feedback for potential revisions or future writing during the interview.

Stage 3: Semi-Structured Interviews (August 2024)

After reviewing the feedback as described, each participant underwent a one-on-one interview with the researcher. Student interviews took place within two days of receiving the feedback (to ensure impressions were fresh). Each interview session lasted approximately 30–45 minutes (including students' examining all feedback files) and was conducted in a quiet office on campus.

During the interview, the researcher used the Student Interview Form as a guideline, but followed a semi-structured approach: questions were not read verbatim in rigid order if the conversation naturally flowed through the topics. Probing questions were used to delve deeper (e.g., "Can you tell me more about why that comment was confusing?" or "How did you feel when you read that suggestion from ChatGPT?"). The atmosphere was

kept conversational to encourage honest and detailed responses. Students were reminded that there were no right or wrong answers and that criticism of any feedback (teacher or AI) was welcome,the aim was to learn from their experience. All interviews were in English but when needed L1 was also used to support students.

Similarly, the fifteen teachers participated in interviews, each approximately 10-15 minutes. For teacher interviews, the researcher provided them in advance with the set of sample feedback on two student paragraphs (to make the discussion concrete). Specifically, they were given anonymised copies of two students' paragraphs along with: the teacher feedback provided on those paragraphs and the corresponding ChatGPT and Grammarly feedback. The two sample cases were chosen because they represented typical scenarios (one paragraph with many language errors and coherence issues, another paragraph that was overall stronger but with some argumentation weaknesses). The teachers reviewed these materials before the interview. During the interview, they answered the questions on the Teacher Interview Form, often referencing the specific examples from those paragraphs to justify their answers. For instance, a teacher might say, "In Student A's writing, the teacher's comment about the topic sentence was very effective, whereas the Grammarly feedback only caught the grammar issue 'does not'." This approach ensured teachers had a concrete frame of reference, making the interview discussion more focused and evidence-backed. All teacher interviews were conducted in English, as all participants were English-proficient.

The interviews were audio-recorded with permission. Immediately after each interview, the researcher made brief reflective notes about key points or striking quotes, to aid later analysis. The audio recordings were transcribed word-for-word using an orthographic style (i.e., including filler words, pauses, laughter where relevant for meaning). Transcription was done by the researcher, which also served as an initial familiarization with the data. Each transcript was labeled with the participant code (e.g., "ST7 Interview Transcript") and stored securely.

Throughout the data collection, measures were taken to ensure consistency and reliability: All students received the same set of feedback elements and were asked the same core interview questions. All interviews were conducted by the same person (the researcher),

following the same protocol, which enhances consistency in how questions were posed. The multi-step procedure (feedback review then interview) was standardized as much as possible across participants. This careful design of instruments and procedure contributes to the trustworthiness of the data gathered.

3.5 Data Analysis Procedure

Data analysis was conducted separately for each type of data (feedback texts and interview transcripts) and then the findings were integrated. The analysis was driven by the research questions, with RQ1 addressed by analysis of the feedback content and RQ2/RQ3 addressed by thematic analysis of interview data.

Each data collection instrument in this study was purposefully selected to address a specific research question. To answer RQ1, a qualitative discourse analysis was conducted on feedback comments given to student writings by five feedback providers. The comments were categorized by both focus (e.g., grammar, content, organization) and function (e.g., correction, suggestion, praise), allowing a multi-dimensional comparison of feedback characteristics across sources.

For RQ2, semi-structured interviews were conducted with 15 students whose texts received feedback from all five sources. Interview questions explored students' emotional reactions, perceived clarity, usefulness, and revision behavior in relation to each type of feedback.

Similarly, RQ3 was explored through semi-structured interviews with EFL instructors who had experience with both AI-supported and traditional feedback practices. Teachers reflected on pedagogical depth, accuracy, and impact on student learning, providing interpretive insights that complemented the textual analysis.

This design ensured that each research question was addressed through a distinct but complementary data source, enhancing analytical coherence and enabling triangulation of findings.

Thematic analysis in this study was conducted through an abductive approach, which integrates both inductive and deductive reasoning. In qualitative research, an inductive

approach refers to developing themes directly from raw data without prior assumptions, while a deductive approach uses existing theoretical frameworks to guide coding. Abduction bridges the two by iteratively moving between data and theory (Timmermans & Tavory, 2012).

In the first phase of analysis, feedback texts were examined inductively through open coding. Categories such as feedback focus (e.g., grammar, content, organization) and function (e.g., correction, suggestion, praise) were derived from recurring linguistic patterns and discourse strategies observed across providers. No pre-set codebook was imposed at this stage.

In contrast, the interpretation of interview transcripts followed a partially deductive path. Initial codes were informed by prior research on learner perceptions of feedback (e.g., Hyland & Hyland, 2006; Carless & Boud, 2018), focusing on aspects like perceived usefulness, emotional impact, and clarity. However, space was left for unexpected themes to emerge inductively during coding.

Throughout the process, the analysis moved back and forth between theory and data. Emerging codes from feedback texts were re-examined in light of participants' verbalized experiences, and vice versa. This iterative engagement, characteristic of abductive analysis, allowed for a flexible yet theory-informed construction of themes that remained grounded in participants' lived experiences.

3.5.1 analysis of written feedback (discourse analysis framework)

To analyze the feedback comments given by ChatGPT, Grammarly, and the teachers, a content-based discourse analysis was employed. This involved both quantitative coding of feedback features and qualitative examination of examples. The goal was to characterize the focus, form, and style of feedback from each source and identify differences and similarities.

Given the literature on WCF, Two primary coding dimensions were decided on: (a) Feedback Focus (target of the comment) and (b) Feedback Function (illocutionary act). Drawing on known frameworks (e.g., Ferris, 1997; Hyland & Hyland, 2001), an analytic scheme was developed as follows:

Feedback Focus Categories: Each feedback comment was categorized by the aspect of writing it addresses. The main categories emerged both from our data and prior studies:

1. Language/Grammar: corrections or comments addressing grammar, vocabulary, spelling, or punctuation (surface-level/mechanical issues).
2. Content/Ideas: comments addressing the development of ideas, clarity of arguments, relevance of content, or presence/absence of necessary information (higher-level content issues).
3. Organization/Coherence: comments addressing text structure, paragraph unity, transitions, and overall coherence or flow of the writing.
4. Style/Tone: suggestions related to the style, formality, or tone of the writing, as well as clarity/conciseness of expression.
5. Praise: any explicit positive feedback or praise about what was done well.
6. Other: any comments not fitting above (e.g., about formatting or references, if any, though in our dataset none of the feedback concerned formatting).

Each piece of feedback (whether a comment in teacher's document, a suggestion from Grammarly, or a sentence from ChatGPT's response) was coded into one (or more, if needed) of these categories. In practice, most comments had a clear primary focus. For example, "Correct 'does not' for third person singular" is Language; "Your paragraph presents good ideas" is Praise; "Add a transition to connect these paragraphs" is Organization.

Feedback Function Categories: Comments were also classified by their nature as speech acts, following Hyland & Hyland's (2001) broad types of written feedback: Praise, Criticism, or Suggestion. This classification helps understand the tone and intent of feedback.

1. Praise: positive remarks or reinforcement (e.g., "Good use of examples here.").
2. Criticism: direct or indirect indication of a problem without an explicit solution (e.g., "This paragraph is unclear.").
3. Suggestion (Constructive Feedback): advice or request for change, including corrections or how to improve (e.g., "Consider defining this term for clarity." or an

explicit correction “change X to Y”). Suggestions often were couched with hedging (in the case of polite feedback) or imperatives.

In our data, teacher and ChatGPT feedback sometimes combined criticism with suggestions (e.g., “This idea is not clear [criticism] – try explaining it with an example [suggestion].”). Grammarly’s feedback was almost exclusively direct suggestions (it identifies errors and suggests fixes, rarely pure praise). Each feedback point was coded accordingly. Notably, praise was expected to appear mostly in teacher or ChatGPT feedback (ChatGPT often includes a polite compliment in its opening).

To ensure coding reliability, a second rater independently coded 20% of the data. Inter-coder agreement was calculated based on simple percentage agreement due to the categorical and interpretive nature of thematic codes used in this study. Inter-coder agreement was approximately 90% on focus categories and 85% on function types which are considered acceptable for qualitative research contexts (Miles, Huberman, & Saldaña, 2014). Discrepancies were resolved through discussion, and final themes were refined collaboratively to enhance analytical coherence (for example, if “clarity” comments count as Style or as Content - clarity of idea expression was decided to go under Content if it affects meaning, or Style if it’s about wording). The first researcher then proceeded to code the remaining data. Frequency counts were obtained for each category by source.

In addition to categorical coding, the length and specificity of feedback were examined (for instance, how many words was a typical comment from each source, and did it pinpoint the issue clearly or speak in general terms). Usage of hedging language (like “maybe consider…”) vs. Directives was noted as an aspect of feedback tone. This qualitative noting helped supplement the categorical data.

The outcome of this analysis is a profile of each feedback source. Chapter 4 will present tables and charts (e.g., distribution of feedback by focus for teacher vs. AI) to answer RQ1. An example of how data were tabulated: how many total feedback comments each source gave across all paragraphs, and how those comments were broken down by focus category were counted. This allowed to say, for instance, what percentage of ChatGPT’s feedback dealt with content as opposed to grammar, and compare that to the teacher’s proportions. illustrative example excerpts from each source’s feedback were extracted to

demonstrate typical feedback styles (see Findings). These excerpts were chosen to represent common patterns (such as a teacher giving a specific suggestion, vs. ChatGPT giving a broad comment, vs. Grammarly giving a direct correction). By analyzing both quantitatively and qualitatively, it was aimed to comprehensively characterize the discourse of feedback from each source. This addresses not only whether differences exist, but also provides insight into the nature of those differences (e.g., teacher feedback might contain more suggestions related to content development, whereas Grammarly focuses on grammar corrections).

3.5.2 Thematic analysis of interviews (qualitative analysis)

The interview data (15 student transcripts and 15 teacher transcripts) were analyzed using thematic analysis, following the six-phase approach outlined by Braun & Clarke (2006). Thematic analysis was deemed appropriate because it offers "a method for identifying, analyzing, and reporting patterns (themes) within data". This flexible method allowed to distill the rich qualitative data into key themes that directly address RQ2 and RQ3 regarding perceptions of feedback. A primarily inductive approach (letting themes emerge from the data) was adopted, but with some deductive influence from the interview questions (since questions revolved around effectiveness, clarity, helpfulness, etc., it is natural that themes would form around these concepts).

The analysis proceeded as follows (per Braun & Clarke's phases):

Phase 1: Familiarization with data: The researcher transcribed all interviews and then read through each transcript multiple times. Initial notes were jotted in the margins – for example, noting whenever a student expressed a strong opinion or an emotional reaction (surprise, frustration, relief) toward some feedback, or when a teacher made an evaluative comment about AI or teacher feedback. Through this immersive reading, the researcher became deeply familiar with the content and began noticing preliminary patterns (e.g., many students were saying in different ways that teacher feedback helped them understand their mistakes better, whereas Grammarly was "just fixes").

Phase 2: Generating initial codes: The researcher systematically coded the transcripts manually. A code in this context is a label attached to a segment of text that identifies a

meaningful feature related to the research question. Semantic coding was employed (staying close to the explicit content of what was said) for concrete statements, and latent coding for underlying ideas or attitudes. Some examples of initial codes for student data included: *"Prefers teacher feedback," "AI feedback – clarity issue," "Grammarly useful for grammar," "Teacher feedback – learning something new," "Emotional impact – motivated by teacher praise," "AI feedback lack trust," "Suggestion for using both sources,"* etc. Teacher data initial codes included: *"Teacher feedback addresses higher-order," "AI feedback limited scope," "Clarity of teacher vs AI," "Student autonomy – AI could help?," "Concern about AI specificity," "Teacher role remains important,"* among others. Altogether, around 40–50 initial codes were generated across the student and teacher datasets. Coding was done at the sentence or paragraph level as appropriate – whenever a participant expressed a distinct idea or experience, it was coded. Segments often got multiple codes if they touched on several ideas (e.g., a student might say, "Grammarly was helpful for quick fixes, but I still prefer the teacher's comments because they explained why my sentence was wrong" – this got coded as both *Quick fixes (AI advantage)* and *Explanations (teacher advantage)*).

Phase 3: Searching for themes: After coding all transcripts, the researcher examined the codes to collate them into broader themes. A theme is a coherent pattern that captures something important about the data in relation to the questions. A primarily inductive approach was used by being mindful of the interview topics (effectiveness, clarity, helpfulness, etc.) which guided some theme development. Codes that were related were grouped. For example, codes like "prefers teacher feedback," "trust in teacher," "teacher seen as expert" all clustered under a candidate theme about Preference for Teacher Feedback/Trust in Teacher. Codes such as "Grammarly for grammar," "AI as quick tool," "use both feedback sources" clustered under an idea of Complementary Use of AI and Teacher Feedback. Negative perceptions of AI feedback (codes on lack of clarity, vague suggestions, lack of depth) formed a theme around Limitations of AI Feedback. It was done separately for student interviews and teacher interviews initially, to see themes within each group.

In this study, the concept of effectiveness was thematically analyzed through multiple qualitative dimensions rather than measured in quantitative terms. Drawing on

frameworks by Bitchener & Ferris (2012) and Hyland (2019), feedback effectiveness was interpreted based on its:

1. Clarity – whether the feedback was understandable and free from ambiguity,
2. Accuracy – whether the feedback correctly diagnosed the issue,
3. Depth – whether it explained why the change was necessary,
4. Actionability – whether the feedback provided clear guidance for revision,
5. Relevance – whether the feedback addressed the central aims of the writing task.

These dimensions were used as analytical lenses during coding, allowing for systematic interpretation of both student and teacher reflections on feedback. Rather than treating effectiveness as a binary or scalar outcome, it was approached as a multifaceted construct grounded in participants' evaluative language and revision experiences.

For the student interviews, the following candidate themes emerged:

Effectiveness and Learning Value: encompassing student judgments on which feedback helped them improve (usually favoring teacher feedback for content and understanding). Effectiveness refers to the extent to which written feedback supports student improvement by helping them notice, understand, and revise errors. It involves qualities such as accuracy, specificity, clarity, pedagogical depth, and long-term developmental value (Bitchener & Ferris, 2012; Carless & Boud, 2018).

Clarity and Specificity: capturing comparisons of how clear or vague the comments were (teacher feedback often cited as more specific, AI – especially ChatGPT – sometimes too general; Grammarly clear but only on surface errors).

Trust and Confidence: relating to students' trust in the source of feedback (e.g., seeing the teacher as more reliable or knowledgeable, versus wariness about AI correctness), and how feedback affected their confidence.

Future Usefulness: perceptions of how each feedback source would help in future writing or exam preparation (some noted teacher feedback teaches underlying skills, AI is immediate but not instructive).

Suggested Hybrid Approach: interestingly, several students mentioned that both sources together could be useful – e.g., use Grammarly/ChatGPT for initial corrections and teacher for deeper guidance.

Emotional/Motivational Impact: a smaller theme where students mentioned how feedback made them feel (e.g., feeling encouraged by teacher praise or feeling overwhelmed by too many corrections). This was mentioned by a few and was noted as a potential theme.

For the teacher interviews, candidate themes included:

Coverage of Higher-Order vs Lower-Order Issues: teachers consistently observed that teacher feedback covered content and coherence (higher-order), whereas AI (especially Grammarly) stuck to lower-order (grammar). This theme encapsulates perceptions of *scope* of feedback.

Clarity and Actionability: teachers compared how clear and actionable the feedback comments were (teacher feedback deemed detailed and actionable; AI feedback sometimes vague or lacking follow-up).

Perceived Effectiveness for Student Improvement: overall judgments on which feedback is more effective and why (all teachers felt their own feedback was more effective for genuine improvement, citing examples).

Role of AI as Supplement, Not Replacement: teachers acknowledged the value of AI tools (especially for catching mechanical errors), but emphasized that AI cannot fully replace human feedback. This theme includes their views on how AI might be integrated (e.g., as a preliminary checker) and the importance of teacher expertise.

Concerns and Cautions: any concerns raised about AI feedback – such as potential misinformation, over-reliance by students, or AI's inability to tailor to student needs.

(One teacher mentioned concern that students might not understand AI's comments without teacher help.)

Positive Aspects of AI Feedback: any mentions of what AI did well (e.g., Grammarly's correctness, ChatGPT's breadth in pointing out issues) was also noted to present a balanced view.

Phase 4: Reviewing themes: The candidate themes was then reviewed against the data to ensure they were valid representations. This involved two levels of review: (1) checking the collated quotes under each theme to see if they formed a coherent pattern, and (2) checking the themes against the entire data set to see if any important data was overlooked or if themes overlapped excessively. Adjustments were made accordingly. For instance, it was found that "Effectiveness and Learning Value" and "Future Usefulness" in student data were closely related and could be merged into a broader theme about Learning Impact of Feedback. Conversely, it was ensured that the theme "Clarity and Specificity" was distinctly evidenced and not redundant given the Learning Impact theme (it was kept separate because even if something was effective, the clarity aspect was a distinct point frequently made). Some themes were renamed for clarity. In teacher data, "Coverage of Issues" and "Effectiveness" were separated although they relate because teachers not only commented on what issues were covered but also explicitly on which feedback leads to improvement (their rationale often being the coverage breadth). At this stage, any theme that was not sufficiently supported by data was also dropped or demoted. For example, the emotional impact for students was mentioned but not by many, so instead of a main theme, it became a minor point under other themes (like students feeling encouraged by clarity or discouraged by confusion). By the end of this phase, a set of well-supported themes for each group appeared.

Phase 5: Defining and naming themes: Clear definitions for each theme, detailing what aspect of the data it captures were written. For example, Theme 1 (Students): "Teacher feedback as more comprehensive and instructive than AI feedback" – defined as students' overall preference for and trust in teacher feedback due to its perceived greater effectiveness in improving their writing and providing learning. Each theme was given a concise name used in reporting (e.g., "Preference for Teacher Feedback and Trust in

Teacher's Expertise," "AI Feedback: Speedy but Superficial," "Clarity and Specificity of Feedback," "Combining AI and Teacher Feedback," etc.). It was also decided on how to present sub-themes if any. In student data, for instance, under the broad theme of "Preference for Teacher Feedback," sub-components included *perceived deeper learning*, *personalized guidance*, and *trust*. These sub-points are discussed in narrative, but they all support the overarching idea of that theme. It was ensured that theme definitions avoided overlap – each theme addresses a distinct facet of the research question responses.

Phase 6: Producing the report (integration into findings): Finally, in Chapter 4 (Findings & Discussion), the themes are reported with evidence. This involves presenting thematic tables and narrative explanations, including representative quotes from participants that exemplify each theme. The participant codes are used for attribution. For instance, if a theme is that *teacher feedback is clearer and more specific*, a quote might be included like: *"The teacher wrote exactly what was wrong and how to fix it, which was easy to follow, but the AI comments were a bit vague,"* (ST3). It was also quantified where appropriate (e.g., "10 out of 15 students explicitly stated they found teacher feedback more helpful overall, while none stated a clear preference for AI alone; a few (n=3) said both were helpful in different ways"). This quantification of qualitative data helps indicate prevalence of themes.

Throughout the thematic analysis, it was adhered to qualitative rigor criteria. Credibility was enhanced by member checking: after initial analysis, a summary of key findings was shared with 5 student participants and 2 teacher participants to see if they agreed that it reflected their views. All generally agreed; minor clarifications from their feedback were incorporated (for example, a student clarified that when they said AI feedback was "fast," they meant in terms of receiving it instantly, not necessarily that they *preferred* quick feedback over detailed feedback, a nuance which was ensured to capture). Dependability was addressed by keeping an audit trail of coding decisions, and having an external audit by a peer who reviewed a sample of coded transcripts and the theme list to see if it made sense. Transferability is supported by rich description of context and participant quotes, allowing readers to judge applicability to other settings.

In qualitative research, the role of the researcher is integral to data interpretation and thematic construction. In this study, all interviews were conducted, transcribed, and analyzed by the researcher. Familiarity with the institutional context offered valuable contextual understanding but also posed a potential for interpretive bias. To address this, analytic reflexivity was maintained through detailed memoing and documentation of analytical decisions throughout the coding process. Themes were derived inductively through iterative engagement with the data, and triangulation was achieved by comparing multiple sources (e.g., textual feedback, student interviews, teacher interviews). Member checking was conducted with selected participants to ensure the trustworthiness of interpretations. Although subjectivity is inherent in qualitative inquiry, all efforts were made to ensure credibility, dependability, and confirmability in line with qualitative research standards (Lincoln & Guba, 1985).

Finally, once both the feedback analysis and the thematic analysis were completed, it was engaged in a triangulation step: comparing and contrasting what was found in the actual feedback content with what participants said. For instance, if the content analysis showed that Grammarly gave exclusively grammar corrections, it was interesting to see students comment that *"Grammarly only fixes grammar, it does not tell me about my ideas,"* which reinforces nicely. In the discussion, these perspectives are integrated and also connected them to existing research findings (to be presented in Chapter 4). This comprehensive analysis procedure ensures that the research questions are answered with a robust combination of objective evidence and human perspectives.

4. FINDINGS AND DISCUSSION

This chapter presents and discusses the findings of the study through two main analytical lenses: (1) discourse analysis of written feedback provided to students, and (2) thematic analysis of student and teacher interviews. The goal is to answer the research questions by identifying both patterns in the feedback content and form, and by understanding the perceptions of those who give and receive it. This approach aligns with Hyland and Hyland's (2006) framework for analyzing WCF and Braun and Clarke's (2006) model for thematic analysis.

The findings are organized into the following major sections:

1. Analysis of the written feedback across five feedback sources
2. Students' perceptions of the feedback (ST1–ST15)
3. Teachers' evaluations and reflections on AI and teacher-generated feedback (T1–T15)

In order to maintain clarity and objectivity, all feedback providers except AI tools and participants have been anonymized. The three teachers are referred as FDT1 (experienced, external), FDT2 (novice, external), and FDT3 (experienced, course teacher). Students are labeled ST1–ST15. Teacher interviews are coded as T1–T15.

4.1 Discourse Analysis of Feedback

This section presents a detailed analysis of feedback provided to each student (ST1–ST15) by five different feedback providers: ChatGPT, Teacher 1 (FDT1), Teacher 2 (FDT2), Teacher 3 (FDT4), and Grammarly. Each student's writing was reviewed by all five feedback sources, and the findings were summarized through tables based on each student which can be found in Appendix 3. Here, rather than listing each feedback set separately, this analysis synthesizes the data using a unified feedback frequency table (Table 1), and supports each interpretation with sample excerpts from actual student texts. The aim is to highlight key differences in the focus, style, and communicative intent of each feedback type.

This section also highlights trends, consistency, and anomalies among feedback sources, framed within academic discourse analysis principles (Hyland & Hyland, 2006; Ferris, 2003).

Each student received feedback from five different sources. The feedback was analyzed and categorized according to two dimensions:

1. Feedback Focus: What linguistic or content aspect the comment addresses (e.g., grammar, content, organization)

2. Feedback Function: How the comment is structured and what purpose it serves (e.g., direct correction, suggestion, praise)

These categories were derived from Hyland and Hyland (2006), Ferris (2003), and Lee (2008), whose work identifies functional and rhetorical strategies in L2 writing feedback. The analysis provides insight into the degree of attention different feedback sources give to different writing aspects.

4.1.1 Finding on Focus of Feedback

Table 1. Frequency of Feedback by Focus Across Five Sources

Feedback Type	ChatGPT	FDT1	FDT2	FDT3	Grammarly	Total Feedback
Grammar & Mechanics	91	39	75	117	178	500
Vocabulary	15	6	15	21	18	75
Organization	26	11	25	40	2	104
Content & Ideas	23	9	21	47	1	101
Style & Tone	5	0	4	11	0	20
Total Feedback	160	65	140	236	199	800

1. Grammar and Mechanics

This dimension refers to surface-level language correctness, including verb tense, subject-verb agreement, punctuation, spelling, and sentence-level syntax. It is the most frequently addressed category, with 91 comments from ChatGPT, 39 from FDT1, 75 from FDT2, 117 from FDT3, and 178 from Grammarly (n=500). Grammarly issued the highest number of grammar corrections, often without providing rules or reasoning. ChatGPT also frequently flagged errors, though its suggestions were more hedged and less directive. In contrast, FDT3 (teacher) embedded grammatical corrections within mini-lessons or reminders of rules, increasing their instructional value and retention.

Sample Feedback Excerpts:

"Change 'have did' to 'have done" (Grammarly to ST1).

"Incorrect verb form" (FDT1 to ST6).

"Incorrect use of 'their'- should be 'there'" (Grammarly to ST7).

"Typo: 'freind' → 'friend'" (Grammarly to ST10).

"Tense mistake here" (FDT1 to ST14).

"Wrong tense used here" (FDT1 to ST15).

"Incorrect preposition: 'interested on' → 'interested in'" (Grammarly to ST4).

2. Vocabulary

Vocabulary-related feedback refers to suggestions about word choice, lexical variety, appropriateness of register, collocation, and repetition. In total, 15 comments were made by ChatGPT, 6 by FDT1, 15 by FDT2, 21 by FDT3, and 18 by Grammarly (n=75). FDT3 offered the most context-sensitive vocabulary feedback, often pointing out word connotation, register, or overuse. AI tools tended to suggest simple synonyms without evaluating appropriateness or nuance, limiting their usefulness in developing lexical sophistication.

Sample Feedback Excerpts:

"Use a more formal word choice"(FDT2 to ST14).

"The word 'huge' sounds exaggerated here. Try 'significant' if you mean important"(FDT3 to ST9).

3. Organization

This dimension includes the internal structure of the text, such as paragraph unity, sequencing of ideas, coherence, and transitions. Teachers were more active in this category: ChatGPT (26), FDT1 (11), FDT2 (25), FDT3 (40), Grammarly (2), totaling n=104. FDT3 consistently guided students to improve paragraphing, topic sentence clarity, and overall flow. ChatGPT and FDT2 made vague references to organization, often lacking clear direction. Grammarly offered almost no feedback in this area, confirming its design focus on sentence-level issues.

Sample Feedback Excerpts:

"You could improve coherence by adding transitional phrases between your ideas" (ChatGPT to ST1).

"Your body paragraph lacks a clear topic sentence. Consider adding one" (FDT3 to ST6).

"Try reordering this paragraph to make your argument clearer" (ChatGPT to ST7).

"Your conclusion is missing a final remark. Try ending more strongly" (FDT2 to ST109).

"Paragraph 2 does not relate back to your thesis. Consider restructuring"(FDT3 to ST14).

4. Content and Ideas

This category refers to higher-order writing concerns such as idea development, argument support, reasoning, critical thinking, and relevance. ChatGPT (23), FDT1 (9), FDT2 (21), FDT3 (47), Grammarly (1), total n=101. FDT3 was most active in prompting elaboration, asking for examples, or questioning unsupported claims. ChatGPT and FDT2 made occasional content-related suggestions, but these were often general and lacked scaffolding. Grammarly showed minimal engagement with idea development.

Sample Feedback Excerpts:

"Nice job with your example. You could make it even more specific"(FDT2 to ST3).

"Avoid generalisations like 'everyone thinks',be more precise" (FDT3 to ST5).

"Your third paragraph lacks coherence. Use transitions to improve flow" (FDT3 to ST9).

"Elaborate on this point,add an example from daily life" (FDT3 to ST13).

"Well-argued paragraph. Now provide a source or example" (FDT3 to ST15).

5. Style and Tone

Style and tone involve rhetorical appropriateness, formality, hedging, and reader awareness. This was the least addressed category with ChatGPT (5), FDT1 (0), FDT2 (4), FDT3 (11), and Grammarly (0), total n=20. Effective feedback in this dimension helps students align their voice with academic discourse conventions. FDT3 and FDT2 gave feedback to soften overly strong claims or rephrase informal expressions. ChatGPT made rare references to tone, often without contextual justification. Grammarly did not address tone unless tied to a mechanical issue.

Sample Feedback Excerpts:

"Use a more formal tone in academic writing"(ChatGPT to ST10).

"Nice argument but rephrase the sentence to make it more formal" (FDT3 to ST11).

"Consider giving an example to support this argument" (ChatGPT to ST14).

"This part is a bit wordy try simplifying your phrasing" (ChatGPT to ST15).

Overall, FDT3 (Teacher 3), the students' current teacher, provided the most comprehensive and balanced feedback across all five feedback categories (n = 236), particularly excelling in higher-order concerns such as content, organization, and tone. This aligns with findings from Hyland & Hyland (2006), who emphasized that teachers with contextual familiarity are more likely to offer individualized, in-depth, and rhetorically supportive comments.

In contrast, FDT1 (Experienced External Teacher) offered the least amount of feedback (n = 65), with a heavy focus on grammar and minimal attention to higher-order issues. This may suggest a more conservative or detached evaluative stance, echoing what Ferris (2003) describes as "minimalist feedback" commonly observed in time-constrained or unfamiliar teaching contexts.

ChatGPT produced a moderately high quantity of feedback (n = 160), displaying a balance between surface-level and higher-order issues. However, the nature of ChatGPT's feedback leaned toward structural support and generic improvement suggestions, which, as Ranalli (2018) suggests, reflects the limitations of AI in providing deeply contextualized or pedagogically scaffolded input.

Grammarly outperformed others in sheer volume of grammatical corrections (n = 178), reinforcing prior literature (e.g., Li, Link, & Hegelheimer, 2015) that characterizes Grammarly as a rule-based, form-focused tool with limited relevance to discourse or content-level development. Notably, its near absence of content or style feedback highlights its transactional rather than instructional nature.

As shown in Table 1, the majority of feedback comments (n=500 out of 800) targeted Grammar & Mechanics**,** suggesting a predominant emphasis on surface-level language issues. This trend was especially pronounced in the case of AI-based tools, particularly Grammarly, which provided 178 grammar-related comments alone. In contrast, feedback addressing Style & Tone was notably scarce (n=20), indicating that both AI and some teachers offered limited guidance on rhetorical appropriateness and voice. Content/Ideas

(n=101) and Organization (n=104) were moderately addressed, with FDT3 (Teacher 3) standing out as the most active provider in these higher-order areas. This distribution suggests that while AI tools efficiently identify form-based issues, they are less equipped to support deeper aspects of academic writing such as coherence, development, and style.

4.1.2 analysis of feedback functions

Table 2. Distribution of Feedback Functions Across Five Sources

Feedback Function	ChatGPT	FDT1	FDT2	FDT3	Grammarly	Total Feedback
Direct Correction	66	39	55	88	165	413
Suggestions	58	18	48	89	12	225
Prompts/Questions	24	6	20	34	0	84
Praise	12	2	17	25	0	56
Total Feedback	160	65	140	236	177	778

Beyond what feedback addresses, how it is delivered also reveals important pedagogical intent. Based on discourse analysis principles (Hyland & Hyland, 2006), four main feedback functions were identified:

1. Direct Correction
2. Suggestions
3. Prompts/Questions
4. Praise

The cumulative data presented in 17 provide a comprehensive overview of communicative style of feedback delivered by five different sources. These results offer critical insights into the relative effectiveness and depth of each provider.

1. Direct Correction

This function identifies an error and provides the correct form. Out of 778 total feedback moves, 413 (53%) were direct corrections. Grammarly led with 165 such comments, reflecting its mechanistic focus on surface-level errors. FDT3 (n=88) and ChatGPT (n=66) also provided correction-based feedback, but with different pedagogical tones. While Grammarly was efficient, it often lacked clarity or explanation. In contrast, teacher-led corrections, particularly by FDT3, were often accompanied by mini grammar lessons that encouraged student learning.

Example:

"Incorrect preposition: 'interested on' → 'interested in'" (Grammarly to ST4).

"You used 'have wrote',this should be 'have written' because it's present perfect" (FDT3 to ST6).

2. Suggestions

Suggestions were the second most common feedback type, comprising 225 out of 778 total comments (29%). This function offers advice for improvement while leaving space for student autonomy. FDT3 (n=89) and ChatGPT (n=58) dominated this category, often suggesting improvements in argument clarity or organization. FDT2 gave 48 suggestions, while FDT1 (n=18) and Grammarly (n=12) rarely used this feedback function. The low number from Grammarly underscores its rigid, corrective design.

Example:

"Try reordering this paragraph to make your argument clearer" (ChatGPT to ST7).

"Consider restructuring your second paragraph to better link with your thesis" (FDT3 to ST14).

3. Prompts and Questions

Prompting and questioning comprised 84 instances (11% of total feedback). FDT3 (n=34) and FDT2 (n=20) frequently used this dialogic strategy to invite elaboration or

clarification. ChatGPT also contributed 24 prompting moves, albeit less pedagogically grounded. FDT1 used this function only 6 times, and Grammarly provided none, highlighting AI's current limitation in engaging students in reflective dialogue.

Example:

"Can you be more specific about who you mean by 'some people'?" (FDT2 to ST3).

"What evidence can you add to support this claim?" (FDT3 to ST13).

4. Praise

Praise occurred least frequently, with only 56 comments (7%) recognizing student strengths. FDT3 (n=25) and FDT2 (n=17) were the most frequent users of praise, providing affirmation for clarity, structure, or critical thinking. ChatGPT gave 12 instances of praise, while FDT1 offered just 2. Grammarly provided no praise at all, reflecting a lack of socio-affective engagement in its feedback logic.

Example:

"Well-argued paragraph,your reasoning is clear" (FDT3 to ST15)

"Nice use of contrast here between the two examples" (FDT2 to ST8)

From a functional perspective, Grammarly heavily relied on direct corrections (n = 165), with almost no use of suggestions, prompts, or praise,reinforcing its non-human, mechanical feedback profile. This corresponds with Winstone et al. (2016), who argued that effective feedback should foster dialogic engagement rather than one-way correction.

In contrast, FDT3 (Teacher 3) not only provided the highest number of suggestions (n = 89) but also demonstrated strong use of prompts (n = 34) and praise (n = 25), indicating a constructivist feedback orientation,encouraging students to reflect, elaborate, and take ownership of their revisions (Nicol & Macfarlane-Dick, 2006).

ChatGPT displayed relatively high numbers across all functions except praise, suggesting a more facilitative tone, though still generic. This is in line with recent work by Marzuki et al. (2023), who found that AI feedback tools offer hedged, polite suggestions but lack the human insight necessary for emotionally supportive commentary.

FDT2 (Novice Teacher) presented a promising use of varied functions, especially praise and suggestions, suggesting an engagement-oriented style, though lacking the structural and rhetorical precision of FDT3. FDT1, despite experience, showed minimal use of prompts or praise, which may shape students' perception of feedback by FDT1 as less useful or engaging.

4.1.3 summary of discourse analysis

In sum, the analysis reveals that quantity alone does not determine quality. FDT3's feedback stands out for its breadth, personalization, and pedagogical depth, likely due to active teaching relationship with the students. ChatGPT and FDT2 offered mid-range support with varying effectiveness, while FDT1 underperformed despite experience. Grammarly proved useful for form-level correction but lacked any discourse or engagement value.

Table 2 also illustrates that Direct Correction constituted over half of all feedback instances (n=413 out of 778), with Grammarly contributing the most in this category (n=165). While such corrections enhance surface accuracy, they rarely engage students in reflective revision. Suggestions (n=225) and Prompts (n=84) were more commonly used by human teachers, particularly FDT3, offering guidance without prescribing exact answers thus supporting learner autonomy. Praise was the least used function (n=56), mainly provided by FDT2 and FDT3. The limited presence of praise and prompting indicates a possible underutilization of affective and motivational dimensions in feedback, especially among AI tools.

These findings substantiate previous research advocating for human, dialogic feedback that addresses both local and global aspects of writing (Ferris, 2014; Hyland, 2019) and raise important questions about the role of AI tools in supporting (rather than replacing) teacher intervention.

4.2 Thematic Analysis of Student Interviews

Table 3. Themes from Students' Interviews

Theme	Student contributors
Clarity and Understandability	ST1, ST4, ST7, ST9, ST12, ST13
Usefulness and Relevance	ST1, ST3, ST6, ST10, ST13
Motivation and Affect	ST3, ST6, ST9, ST11, ST13
Personalization and Context	ST2, ST5, ST8, ST12, ST15
Trust and Credibility	ST2, ST5, ST8, ST10, ST14, ST15

Research Question 2: How do students perceive and evaluate the quality, clarity, and usefulness of feedback from both AI-originated and teacher-written feedback?

Thematic analysis of the 15 student interviews (ST1–ST15) yielded five key themes reflecting how students compare AI-generated feedback (ChatGPT, Grammarly) and teacher-written feedback. Table 3 summarizes which students commented under each theme. Below, each theme is elaborated with illustrative quotes and interpreted in light of feedback literature.

4.2.1 Clarity and understandability

Many students noted differences in how clearly AI and teacher feedback were explained. For example, ST1 observed, *"ChatGPT showed my grammar errors in red, which was clear, but it didn't always explain why I needed to change them; the teacher usually writes a note, so I understand the reason."*

Similarly, ST12 said *"Grammarly caught every mistake with highlights, but it was hard to tell which suggestions mattered the most. My teacher's comments, though fewer, made sense immediately."*

These comments suggest AI tools often give explicit corrections yet lack the elaboration teachers provide. The ambiguity in AI feedback (especially from Grammarly) made some students feel less sure of how to apply it. This aligns with feedback research emphasizing that understandability is crucial for feedback uptake (Hattie & Timperley, 2007; Hyland & Hyland, 2006). Hattie and Timperley (2007) note that feedback must be presented in a clear, digestible form for students to use it. In this study, students appreciated the visual clarity of AI feedback (e.g. error highlights) but valued the teacher's explanatory clarity – a finding echoed in prior L2 writing research (Ferris, 2003; Hyland & Hyland, 2006) that learners often need personalized explanation to interpret corrections.

4.2.2 Usefulness and relevance

Students also compared how helpful each feedback type was for improving their writing.

ST3 commented, *"When I look at ChatGPT's feedback, I see grammar fixes but I do not know if my ideas got better. My teacher points out what needs to be added or changed in the content, which actually improves my essay."*

In contrast, ST6 said, *"Grammarly was useful for fixing typos and basic grammar; I could immediately make corrections. But my teacher's feedback on organization and examples helped me understand the bigger picture of my writing."*

Here students recognize that AI tools tend to target low-level issues (grammar, spelling), which some found immediately actionable, while teacher feedback often addressed higher-order concerns (content, coherence). This pattern reflects existing findings that automated feedback is precise on surface errors but limited on deep content (Roscoe et al. 2017; Roscoe & McNamara, 2013) – a limitation students noted. The perceived utility of teacher feedback for learning was higher in many cases, as it was seen as more directly related to improving their writing ability. These observations are consistent with Hyland and Hyland's (2006) theory that effective feedback should guide revision and learning; students here felt teacher comments better served those goals. In short, while AI feedback was seen as useful for immediate corrections, teacher feedback was seen as more instructive and relevant to overall writing improvement (cf. Ferris, 2003; Hattie & Timperley, 2007).

4.2.3 Motivation and affect

Affective responses emerged as an important theme. Some students reported that teacher feedback felt more encouraging or engaging than AI feedback.

For instance, ST11 remarked, *"My teacher often says 'good job' or writes encouraging words, which makes me want to keep writing. ChatGPT just lists issues without praising anything, so I do not feel the same motivation."*

Similarly, ST9 noted, *"When ChatGPT corrects me, it's just text on a screen. It does not feel personal. But when the teacher writes a smiley face or a note, I feel proud."* These comments indicate that students value the human warmth and encouragement in teacher feedback – a factor known to affect learner motivation (Hyland & Hyland, 2006). Conversely, the AI feedback's impersonal tone (e.g. sterile corrections) was sometimes seen as demotivating or anxiety-inducing.

One student (ST6) explained, *"Grammarly underlines everything in red and it feels scary, like I'm doing everything wrong. The teacher usually highlights good parts too, so it balances."* This resonates with findings that non-judgmental, confidence-building feedback (Hattie & Timperley, 2007; Brindley, 1989) promotes motivation. In this analysis, teacher feedback often included positive language or contextualized critique, which students found motivating; AI feedback tended to lack positive reinforcement. Notably, Teng (2025) similarly found that EFL learners valued chatbots' instant feedback but missed the affective support of a teacher.

4.2.4 Personalization and context

Students were keenly aware of how personalized the feedback was. Many felt that teachers tailor their comments to the individual and the specific text.

As ST2 put it, *"My teacher knows what mistakes I usually make and corrects those. AI tools like ChatGPT or Grammarly do not know me – they just apply general rules."*

ST12 similarly said, *"When I made the same mistake twice, the teacher reminded me only once and I remembered it. Grammarly would underline it every time, which got annoying."*

These remarks suggest that teacher feedback was seen as adaptive to the student's history and context, whereas AI feedback was generic. In the literature, personalized feedback is recognized as more effective (Hyland & Hyland, 2006; Bitchener & Knoch, 2008). The students' comments mirror the idea that human instructors can distinguish student-specific learning needs, whereas AI algorithms apply standard criteria. For example, Kumaravadivelu (2006) and Hyland (2003) have argued that the interpersonal dimension of teacher feedback makes it more attuned to individual learners. In short, the learners in this study expected feedback to "know" them; they felt teachers satisfied this expectation more than AI tools did. This finding also parallels Çiftçi's (2024) theme of personalization in AI-enhanced learning, highlighting that students notice when feedback seems tailored versus one-size-fits-all.

4.2.5 Trust and credibility

Finally, students frequently commented on the trustworthiness of feedback sources. Several students said they trusted teacher feedback more.

ST5 explained, *"I know my teacher is experienced, so when she corrects something I believe it's right. Sometimes ChatGPT or Grammarly suggests changes that even my teacher disagrees with, so I'm skeptical."*

Similarly, ST8 noted, *"When ChatGPT was not sure about something, it said 'maybe use this word instead.' That made me doubt if I should do it; I'd rather ask the teacher."*

These examples show that students were aware of possible errors or uncertainty in AI feedback. The skepticism toward automated feedback aligns with findings by Bixler and Williams (2019) (in qualitative feedback studies) that learners view AI suggestions as fallible. In contrast, teacher feedback carried institutional authority and expertise in students' eyes.

One student (ST10) succinctly stated, *"I always trust the teacher's corrections more. Even if ChatGPT is usually right about grammar, I double-check with the teacher if I'm unsure."*

This reflects an important aspect of feedback quality: credibility (Carless, 2016). The literature warns that feedback is ineffective if learners do not accept it (Shute, 2008), and here students implicitly assessed credibility by source. Thus, the theme of trust complements the others, indicating that even clear or useful AI feedback may be discounted if not trusted (cf. Teng, 2025; Bixler & Williams, 2019).

In sum, students' remarks reveal a nuanced evaluation of AI-based versus teacher feedback. They appreciated AI for clarity on surface errors and immediacy but valued teacher feedback for contextual depth, encouragement, and trusted guidance. These themes – clarity, usefulness, motivation, personalization, and trust – echo patterns reported in ELT feedback research (e.g. Hyland & Hyland, 2006; Ferris, 2003; Hattie & Timperley, 2007). The thematic analysis parallels Çiftçi's (2024) approach, which similarly identified benefits (e.g. efficiency, personalization) and concerns (e.g. loss of human touch) in AI-supported ELT. As with Çiftçi's findings, the students saw the best outcomes coming from a balanced use of AI tools alongside traditional feedback. The themes identified here provide direct insight into RQ2 by highlighting how students judge the quality, clarity, and value of feedback from both AI and teacher sources.

4.3 Thematic Analysis of Teacher Interviews

Research Question 3: How do teachers perceive and evaluate the quality, clarity, and usefulness of feedback from both AI-originated and teacher-written feedback?

Fifteen semi-structured interviews with English language instructors (T1–T15) were analyzed thematically using Braun and Clarke's (2006) method. The goal was to identify how teachers perceive the feedback generated by AI tools (ChatGPT and Grammarly) in comparison to their own teacher-written feedback. Teachers were varied in experience, ranging from novice to highly experienced, and included those who had and had not taught the students directly.

The following five themes emerged:

1. Perceived Usefulness of AI Feedback

2. Concerns about Clarity and Misinterpretation

3. Pedagogical Trust and Control

4. Time and Workload Efficiency

5. Personalization and Emotional Value

Each theme is introduced with a definition and literature support, followed by illustrative quotes and interpretation. The themes and the teachers contributing to these themes are shown in Table 4 below.

Table 4. Summary Table: Themes and Teacher Contributors

Theme	Teacher Contributors
Perceived Usefulness of AI Feedback	T1, T2, T3, T5, T8, T10, T13
Concerns about Clarity and Misinterpretation	T2, T4, T6, T9, T12, T14
Pedagogical Trust and Control	T3, T5, T7, T8, T10, T15
Time and Workload Efficiency	T1, T2, T4, T6, T11, T13
Personalization and Emotional Value	T3, T6, T7, T9, T12, T15

4.3.1 Theme 1: Perceived usefulness of ai feedback

Usefulness in written feedback refers to how well it supports student learning and improves writing quality (Bitchener & Ferris, 2012). Feedback that is specific, accurate, and developmentally appropriate is typically rated as useful by educators (Hyland, 2019). Many teachers acknowledged the practical utility of AI tools, particularly for surface-level issues. T1 noted, *"ChatGPT and Grammarly catch a lot of grammar issues that students often miss, so they're useful in that way."* T5 echoed this, saying, *"I think Grammarly is effective for what it does,it cleans up the text. But it stops there."* T13 added, *"ChatGPT sometimes surprises me with its ability to rephrase awkward sentences. It can be a good revision companion."*

Teachers found AI feedback particularly useful for accuracy and fluency, consistent with Roscoe and McNamara's (2013) findings on intelligent tutoring systems. However, most saw this utility as limited to lower-order concerns, with little depth for content or structure. This mirrors Ferris's (2014) position that effective feedback should go beyond error correction to support the development of ideas and argumentation. As one teacher explained, "*AI can be very helpful with grammar and surface-level mistakes. I use it when I don't have time to go through every line. But for ideas, organization, or deeper coherence, it really doesn't replace what I can offer as a teacher" (T3).* In this light, AI was perceived more as a support tool than a replacement for human feedback.

4.3.2 Theme 2: Concerns about clarity and misinterpretation

Clarity in feedback ensures that learners understand what is being corrected and why (Hattie & Timperley, 2007). Misinterpretation occurs when students apply feedback incorrectly or feel confused by vague, overly generic, or hedged comments.

Three instructors expressed concern that AI feedback was not always clear or easy to interpret. T4 explained, *"Students sometimes ask me what Grammarly meant by 'improve conciseness',they do not always get it."* T9 shared, *"ChatGPT uses polite but vague language. It says things like 'consider revising,' but does not always say how."* T6 remarked, *"It looks good at first glance, but when you read closely, it's often not actionable."*

Teachers observed that AI feedback lacks specificity and pedagogical scaffolding. This reinforces Carless and Boud's (2018) idea that actionable feedback is essential for uptake. When suggestions are unclear, they may increase student anxiety or result in mechanical revisions. Instructors noted that while AI feedback is linguistically polished, it often fails to communicate intent effectively,a concern also noted by Marzuki et al. (2023) in studies of AI in academic writing. As one teacher remarked, *"It gives feedback that sounds impressive, but sometimes I don't even understand what it's asking the student to do. Imagine how confused the student must feel" (T6).* This highlights the need for clarity and instructional alignment in feedback design, which AI systems currently struggle to achieve.

4.3.3 Theme 3: Pedagogical trust and control

Trust in feedback is essential for its acceptance and use (Carless, 2016). Teachers often express concerns about maintaining pedagogical control when external tools suggest changes they did not approve or might not agree with (Winstone et al., 2016).

Several teachers expressed discomfort with the idea of students relying solely on AI tools. T3 stated, *"Sometimes ChatGPT tells students to reword a sentence in a way I wouldn't agree with. They bring me those changes, and I have to undo them."* T8 added, *"It feels like giving part of our job to a machine that does not know the context."* T15 emphasized, *"As educators, we want students to learn from our feedback, not just fix things blindly based on suggestions."*

These comments highlight a perceived threat to instructional authority. Teachers emphasized that while AI tools can assist, they should not replace teacher guidance. This aligns with Carless (2016), who argues that trust in the feedback process must be two-way,teachers must trust that tools support learning rather than undermine pedagogical intent. When students prioritize AI feedback over their teachers', it can disrupt the learning relationship and dilute the value of human instruction.

4.3.4 Theme 4: Time and workload efficiency

Feedback is often described as one of the most time-consuming teaching tasks (Carless et al., 2011). Tools that reduce this workload while maintaining quality are typically welcomed by instructors, provided they do not compromise pedagogical aims.

Many teachers recognized the efficiency offered by AI. T1 shared, *"When I have 30 essays to mark, Grammarly at least helps clean up the mechanics so I can focus on the ideas."* T4 agreed, *"It shortens my workload. I do not have to comment on every comma splice anymore."* However, some noted trade-offs. T11 commented, *"It saves time, but I still have to double-check its suggestions. That's extra work sometimes."*

AI tools were appreciated as time-saving assistants for surface-level corrections. However, several teachers emphasized that human oversight is still necessary, particularly for nuanced issues. This aligns with Hyland (2019), who highlights the

balance between efficiency and instructional quality. Teachers were open to hybrid models but insisted that automation should serve,rather than drive,the pedagogical process.

4.3.5 Theme 5: Personalization and emotional value

Personalized feedback reflects an understanding of the individual learner's needs, history, and affective state (Bitchener & Knoch, 2008). Teachers often include affective support,praise, encouragement, and emotional connection,which is critical for learner motivation and confidence (Hyland & Hyland, 2006).

Teachers emphasized the role of emotional presence in feedback. T6 shared, *"Students need more than corrections,they need encouragement."* T9 said, *"Sometimes I write a smiley or a comment like 'great improvement',these little things matter. ChatGPT can't do that."* T12 added, *"Feedback is also about connection. It's about the student feeling seen."*

AI feedback, while efficient and scalable, was seen as emotionally flat. Teachers viewed their own feedback as more empathetic and relational. This emotional value supports deeper engagement and reinforces learner identity,a point emphasized by Winstone et al. (2016). The loss of personalization in automated feedback systems was seen as a potential risk, particularly in emotionally sensitive or demotivated learners.

4.3.6 Summary of teacher interview findings

In sum, teachers acknowledged the potential of AI tools for increasing efficiency and surface-level accuracy but consistently emphasized the irreplaceable value of human feedback in terms of clarity, trust, emotional connection, and pedagogical integrity. AI feedback was seen as a useful supplement but not a substitute for teacher expertise.

These findings answer RQ3 by revealing nuanced teacher attitudes: generally open to integration, but cautious about overreliance. The themes here mirror those in student responses (Section 4.4) and echo patterns found in similar studies (e.g., Çiftçi, 2024; Marzuki et al., 2023; Carless & Boud, 2018), reinforcing the need for blended, thoughtful integration of AI tools in EFL writing instruction.

4.4 Cross-Case Discussion and Interpretation

This section synthesizes findings from the discourse analysis of feedback, student interviews (ST1–ST15), and teacher interviews (T1–T15) to provide a holistic answer to the research questions. Through triangulation of these three data sources, the discussion identifies converging patterns, conflicting perceptions, and critical implications for the implementation of AI-originated feedback in EFL writing instruction.

4.4.1 Agreement on ai's strength in surface-level feedback

Across all three data sources, there was clear consensus that AI tools,particularly Grammarly and ChatGPT,are effective in detecting and correcting lower-order issues such as grammar, spelling, and mechanics. This was evidenced in the discourse analysis, where Grammarly provided the highest volume of grammar corrections (n = 178), followed closely by ChatGPT and FDT3.

Students frequently mentioned appreciating AI tools for this purpose. For example, ST6 remarked, *"Grammarly is useful for fixing typos and simple mistakes."* Teachers (T1, T5, T13) echoed this view, acknowledging that AI tools offload the burden of micro-level corrections, freeing them to focus on higher-order concerns.

This pattern supports existing findings in the literature. As Ferris (2014) notes, students benefit from clear and immediate feedback on accuracy issues, and AI tools appear to fulfill this role efficiently. Similarly, Roscoe and McNamara (2013) found that intelligent feedback systems increase surface-level accuracy, though they are less effective in fostering deeper writing development.

4.4.2 Divergence on higher-order feedback and personalization

The most notable divergence between AI and teacher feedback emerged in relation to higher-order concerns such as content, coherence, organization, and argument development. Teachers consistently emphasized their capacity to provide individualized, pedagogically grounded, and context-aware comments, while students reported that AI feedback on these aspects often lacked depth or relevance.

The discourse analysis showed that FDT3 (the current classroom teacher) gave the highest number of comments related to content (n = 47) and organization (n = 40), while Grammarly barely addressed these categories.

Student perceptions mirrored this. ST3 said, *"My teacher helps me develop ideas. ChatGPT helps with grammar, but not so much with content."* Similarly, in the teacher interviews, T4 and T8 emphasized that AI lacks contextual awareness, making it "helpful but shallow" for deeper feedback.

This contrast aligns with Hyland and Hyland's (2006) argument that higher-order feedback is both instructional and relational. Teacher-written feedback benefits from familiarity with the student's developmental trajectory and classroom context. The impersonality of AI tools limits their ability to adapt to individual writing goals, a weakness emphasized in both student and teacher accounts.

4.4.3 Emotional tone and motivation: human touch vs. Machine neutrality

Another key theme that cut across data sets was the affective dimension of feedback. Students repeatedly expressed that teacher feedback provided encouragement, praise, and a sense of emotional connection, whereas AI feedback was perceived as neutral or even demotivating.

ST11 noted, *"When my teacher writes 'good job,' I feel encouraged. ChatGPT never says anything nice,it just lists mistakes."* This emotional absence was also commented on by teachers. T6 explained, *"Feedback isn't just about correction. It's about encouragement. That's what AI can't do."*

Hyland and Hyland (2006) and Winstone et al. (2016) emphasize that feedback plays an interpersonal role in learning. Encouragement and praise are known to foster motivation, especially in language learners. The inability of AI tools to deliver affectively rich feedback presents a major limitation in their pedagogical value.

4.4.4 Trust, authority, and pedagogical control

A central divergence between AI and human feedback lay in the trust and credibility attributed to each source. Students generally trusted teacher feedback more, perceiving it

as more accurate and context-sensitive. ST5 stated, *"I always believe what my teacher says because she has experience."* Others expressed confusion or skepticism when ChatGPT or Grammarly gave vague or tentative suggestions.

Teacher interviews reinforced this. T3 observed, *"Students follow AI suggestions blindly sometimes,even when they're wrong. I have to reteach things that were 'corrected' by ChatGPT."* T10 shared a concern about pedagogical authority: *"If students start relying more on AI than on us, we lose an important instructional channel."*

These concerns echo findings in Carless (2016), who notes that learners' acceptance of feedback depends heavily on their trust in its source. If feedback is doubted,or if it undermines the teacher's voice,it loses its formative power. The issue here is not merely technical accuracy, but educational authority: teachers felt that AI could inadvertently compete with or distort their role as knowledge-guides and mentors. This "tension of control" is increasingly common in hybrid classrooms (Winstone et al., 2016).

4.4.5 Toward an integrated feedback model

Despite their critiques, both students and teachers expressed openness to a blended feedback model, one where AI tools handle surface-level issues efficiently, while teachers focus on personalized and affective aspects of learning.

This hybrid approach is supported in the literature. Nicol and Macfarlane-Dick (2006) advocate for feedback systems that engage both cognitive and emotional processes. Ferris (2014) recommends that automated tools should be used strategically to reinforce (not replace) teacher insight. Çiftçi's (2024) study similarly emphasizes the value of complementing AI with human scaffolding, especially in EFL contexts where motivation and clarity are often fragile.

4.4.6 Summary of the cross-case discussion

The triangulated analysis reveals a complex but coherent picture:

1. AI feedback is appreciated for clarity, speed, and surface-level accuracy,
2. Teacher feedback is valued for depth, personalization, emotional support, and trust,

3. Students and teachers both see potential in combining the strengths of each, but warn against overreliance on AI tools, particularly for formative and interpersonal dimensions of feedback.

These insights address all three research questions, affirming that while AI-originated feedback plays a useful supporting role, it cannot replace the multidimensional, pedagogically rich feedback provided by experienced teachers. Rather, it should be positioned as a complementary scaffold in EFL writing education.

5. CONCLUSION AND SUGGESTIONS

This study investigated the comparative effectiveness of AI-generated feedback and teacher-written feedback in English as a Foreign Language (EFL) writing instruction. The research adopted a qualitative design to examine the linguistic focus and pedagogical quality of different feedback sources,ChatGPT, Grammarly, and FDT1, FDT2, and FDT3 (teachers), and to explore how students and teachers perceive, interpret, and respond to these feedback types. The study aimed not only to assess the instructional value of these feedback sources but also to uncover the factors that contribute to their perceived usefulness and limitations in a real-world EFL context.

Summary of Findings

The findings reveal that AI tools such as ChatGPT and Grammarly are particularly effective for addressing surface-level errors, including grammar, vocabulary, and mechanics. Students appreciated the speed and objectivity of AI-generated feedback, particularly for low-stakes editing tasks. However, AI feedback was often found lacking in terms of depth, contextual understanding, and emotional tone. In contrast, teacher feedback, especially from FDT3,was characterized by scaffolding, personalization, and a deeper engagement with students' ideas and organization. This made it more motivational and actionable, especially for higher-order concerns like coherence and argumentation.

A notable finding was the contrast between FDT1 and FDT3, both experienced teachers. While FDT1 provided brief, impersonal comments with little contextualization, FDT3's feedback was richer, clearer, and more targeted,largely due to her direct teaching relationship with the students. This supports the notion that teacher familiarity and feedback personalization significantly affect student reception and revision behaviour. Students repeatedly stated that they trusted and valued teacher feedback more when it felt connected to their learning history and individual needs.

The qualitative data also revealed that students often felt overwhelmed or unsure when interpreting AI feedback,especially when it was too generic or lacked explanation. On the

other hand, they described teacher feedback as more relatable, clear, and motivating. Teachers themselves acknowledged the potential of AI tools as time-saving and useful for basic revisions but cautioned against overreliance, noting that such tools cannot replace human insight, emotional awareness, and pedagogical judgment. These findings echo theoretical models such as Vygotsky's Zone of Proximal Development (ZPD), which emphasize the importance of socially mediated, guided learning.(1978)

Implications for ELT

This research provides several important insights for the field of English Language Teaching (ELT). First, it demonstrates the complementary roles that AI and human feedback can play in writing instruction. AI-generated feedback is best suited for immediate, mechanical corrections, while teacher-written feedback remains essential for addressing deeper, content-based aspects of writing.

This study also contributes to the development of feedback literacy in ELT. Students' ability to interpret and act on feedback depends on how, by whom, and in what context it is delivered. Teachers should help learners develop reflective strategies for using feedback from multiple sources.

Another critical contribution lies in highlighting the role of familiarity in teacher effectiveness. Feedback is not only about correctness,it is also about tone, trust, and pedagogical alignment. A teacher's knowledge of a student's background, goals, and prior work enhances the relevance and impact of their feedback. This reinforces the idea that effective feedback is a deeply human process shaped by relational dynamics.

Key Contributions

1. This study provides an empirical comparison between AI-generated and teacher-written feedback on the same student texts, offering a controlled perspective on their respective characteristics and impacts.
2. By triangulating feedback data with student and teacher perceptions, this research contributes to the growing field of AI-assisted language instruction and offers practical insights for developers, educators, and curriculum designers.

3. It identifies trust, personalization, and familiarity as central factors shaping students' preferences for different types of feedback, highlighting the emotional and relational dimensions of feedback reception.
4. The findings emphasize the importance of dialogic and scaffolded feedback, reinforcing the notion that effective feedback involves ongoing interaction and guided support.
5. The study validates that the credibility of feedback is strongly influenced by the quality of the student-teacher rapport, underscoring the interpersonal element in educational contexts.
6. It contributes to the Turkish EFL classroom literature by filling a gap in comparative research on AI and teacher feedback, an area with limited prior exploration.

Educational Implications

1. Institutions should adopt blended feedback models that integrate AI for surface-level corrections and teachers for higher-order development.
2. Teachers require training to critically assess AI-generated feedback and guide students in using it effectively.
3. Feedback literacy should be embedded in writing instruction to improve students' ability to reflect on and revise based on different feedback types.
4. The human element of teaching, particularly empathy, scaffolding, and contextual responsiveness, must be preserved and prioritized.
5. Teacher-student familiarity should be recognized as a pedagogical asset when assigning feedback responsibilities.

Suggestions for Further Research

To build on the limitations of this study, which focused on a single institution, one writing task, and a small sample of students and teachers, future research could expand and deepen the findings by:

1. Including public and private schools, and diverse institutional contexts,
2. Investigating other genres of writing and how different feedback types affect them,
3. Using longitudinal methods to study feedback impact on revision habits over time,

4. Incorporating student writing outcomes and performance data,
5. Exploring feedback from additional AI tools and platforms,
6. Analyzing stakeholder perspectives, including administrators and curriculum designers,
7. Investigating how teacher demographics (e.g., age, tech proficiency) affect feedback styles and AI integration.

Final Remarks

This study contributes to the evolving landscape of EFL writing instruction by demonstrating that while AI feedback offers efficiency and accessibility, it lacks the pedagogical depth, personalization, and emotional intelligence provided by human teachers. A balanced feedback ecosystem,where AI supports and human judgment guides,emerges as the most promising approach. Teacher feedback, particularly when grounded in familiarity and care, continues to play a vital role in helping students become reflective, confident writers. The responsible integration of AI into EFL classrooms must be shaped by pedagogical priorities, ethical standards, and a firm commitment to the irreplaceable value of the teacher’s voice.

REFERENCES

Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *Modern Language Journal, 78*(4), 465–483. https://doi.org/10.2307/328585

Alsowat, H. H. (2022). An investigation of Saudi EFL teachers' perceptions of learning-oriented language assessment. *European Journal of English Language and Literature Studies, 10*(3), 16–32. https://doi.org/10.37745/ejells.2013/vol10no3pp.16-32

Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education, 32*(2), 159–181. https://doi.org/10.1080/02602930600801928

Apriani, Eka & Cardoso, Luis Miguel & Obaid, Ahmed & , Muthmainnah & Wijayanti, Erfin & Esmianti, Farida & Supardan, Dadan. (2024). Impact of AI-Powered ChatBots on EFL Students' Writing Skills, Self-Efficacy, and Self-Regulation: A Mixed-Methods Study. *Global Educational Research Review. 1*. 57-72. https://doi.org/10.71380/GERR-08-2024-8

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57*, 100745. https://doi.org/10.1016/j.asw.2023.100745

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. New York, NY: Routledge. https://doi.org/10.4324/9780203832400

Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12*(3), 409–431. https://doi.org/10.1177/1362168808089924

Bitchener, J., & Knoch, U. (2010). The contribution of written corrective feedback to language development: A ten-month investigation. *Applied Linguistics, 31*(2), 193–214. https://doi.org/10.1093/applin/amp016

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

Brindley, G. (1989). *Assessing achievement in the learner-centred curriculum*. National Centre for English Language Teaching and Research, Sydney, Australia: Macquarie University.

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford, UK: Oxford University Press.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31*(2), 219–233. https://doi.org/10.1080/03075070600572132

Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*. Routledge. https://doi.org/10.4324/9781315740621

Carless, D. (2016). Feedback as dialogue. In M. A. Peters (Ed.), *Encyclopedia of educational philosophy and theory* (pp. 1–6). Singapore: Springer. https://doi.org/10.1007/978-981-287-532-7_389-1

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4), 395–407. https://doi.org/10.1080/03075071003642449

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing, 12*(3), 267–296. https://doi.org/10.1016/S1060-3743(03)00038-9

Chodorow, M., & Burstein, J. (2004). *Beyond essay length: Evaluating e-rater's performance on TOEFL essays* (TOEFL Research Report No. RR-73, ETS RR-04-04). Princeton, NJ: ETS.

Çiftçi, A. (2024). *AI-assisted teaching: Practices and perspectives of instructors on using AI tools in ELT* (Publication No. 880474) [Master's thesis, Maltepe University]. Retrieved June 18, 2025, from https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=usXiZIM9Lp0wk-YzRoaT--NCcSfExvLVloPAG2NUNhgKPHR3LR9ncFfu0fN2Ub1W

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Thousand Oaks, CA: SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge, UK: Cambridge University Press.

Dembsey, J. M. (2017). Closing the Grammarly gaps: A study of claims and feedback from an online grammar program. *The Writing Center Journal, 36*(1), Article 5. https://doi.org/10.7771/2832-9414.1815

Dikli, S. (2006). Automated Essay Scoring. *Turkish Online Journal of Distance Education, 7*(1), 49-62. Retrieved February 19, 2024, from https://dergipark.org.tr/en/pub/tojde/issue/16923/176620

Dodigovic, M., & Tovmasyan, A. (2021). Automated writing evaluation: The accuracy of Grammarly's feedback on form. *International Journal of TESOL Studies, 3*(6), 71–87. https://doi.org/10.46451/ijts.2021.06.06

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford, UK: Oxford University Press.

Ene, E., & Upton, T. A. (2014). Learner uptake of teacher electronic feedback in ESL composition. *System, 46*, 80–95. https://doi.org/10.1016/j.system.2014.07.011

Evans, N. W., Hartshorn, K. J., & Tuioti, E. A. (2010). Written corrective feedback: The practitioners' perspective. *International Journal of English Studies, 10*(2), 47–77. https://doi.org/10.6018/ijes/2010/2/119191

Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly, 31*(2), 315–339. https://doi.org/10.2307/3588049

Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing, 8*(1), 1–11. https://doi.org/10.1016/S1060-3743(99)80110-6

Ferris, D. R. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410607201

Ferris, D. R. (2010). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (2nd ed., pp. 81–104). Cambridge, UK: Cambridge University Press.

Ferris, D. R. (2011). *Treatment of error in second language student writing* (2nd ed.). Ann Arbor, MI: University of Michigan Press.

Ferris, D. R. (2014). Responding to student writing: Teachers' philosophies and practices. *Assessing Writing, 19*, 6–23. https://doi.org/10.1016/j.asw.2013.09.004

Fithriani, R. (2019). ZPD and the benefits of written feedback in L2 writing: Focusing on students' perceptions. *The Reading Matrix: An International Online Journal, 19*(1), 63–73. Retrieved February 19, 2025, from https://readingmatrix.com/files/20-c6t93b93.pdf

Gee, J. P. (2014). *An Introduction to Discourse Analysis: Theory and Method* (4th ed.). New York, NY: Routledge.

Han, T., & Sarı, E. (2022). An investigation on the use of automated feedback in Turkish EFL students' writing classes. *Computer Assisted Language Learning, 37*(1), 1–24. https://doi.org/10.1080/09588221.2022.2067179

Hattie, J., Crivelli, J., Gompel, K., West-Smith, P., & Wike, K. (2021). Feedback that leads to improvement in student essays: Testing the hypothesis that "where to next" feedback is most powerful. *Frontiers in Education, 6*, Article 645758. https://doi.org/10.3389/feduc.2021.645758

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hoang, T. L. G. (2022). Feedback precision and learners' responses: A study into ETS "Criterion" automated corrective feedback in EFL writing classrooms. *JALT CALL Journal, 18*(3), 444-467. https://doi.org/10.29140/jaltcall.v18n3.775

Huawei, S., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*. Advance online publication. https://doi.org/10.1017/S0958344023000265

Hwang, G.-J., & Chang, C.-Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2021.1952615

Hyland, K. (2003). *Second language writing*. Cambridge, UK: Cambridge University Press.

Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge, UK: Cambridge University Press.

Hyland, K., & Hyland, F. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing, 10*(3), 185–212. https://doi.org/10.1016/S1060-3743(01)00038-8

Hyland, K., & Hyland, F. (2006). *Feedback in second language writing: Contexts and issues*. Cambridge, UK: Cambridge University Press.

Kang, E. Y., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal, 99*(1), 1–18. https://doi.org/10.1111/modl.12189

Kumaravadivelu, B. (2006). *Understanding language teaching: From method to postmethod.* Mahwah, NJ: Routledge.

Merriam, S. B. (2009). *Qualitative Research: A Guide to Design and Implementation.* San Francisco, CA: Jossey-Bass.

Lee, I. (2008). Student reaction to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing, 17*(3), 144–164. https://doi.org/10.1016/j.jslw.2007.12.001.

Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Singapore: Springer. https://doi.org/10.1007/978-981-10-3924-9

Leki, I. (1991). The preferences of ESL students for error correction in college-level writing classes. *Foreign Language Annals, 24*(3), 203–218. https://doi.org/10.1111/j.1944-9720.1991.tb00464.x

Li, Z., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing, 27*, 1–18. https://doi.org/10.1016/j.jslw.2014.10.004

Lim, S. C., & Renandya, W. A. (2020). Efficacy of written corrective feedback in writing instruction: A meta-analysis. *TESL-EJ, 24*(3), 1–26. Retrieved February 27, 2024, from https://eric.ed.gov/?id=EJ1275821

Lin, S., & Crosthwaite, P. (2024). The grass is not always greener: Teacher vs. GPT-assisted written corrective feedback. *System, 127*, Article 103529. https://doi.org/10.1016/j.system.2024.103529

Llausas, S. M., Ruiz, E., Ayucan, S. M., & Evardo Jr., O. J. (2024). A systematic literature review on the use of Grammarly in improving the writing skills of ESL/EFL

students. *International Journal of Multidisciplinary: Applied Business and Education Research, 5*(9), 3507–3516. https://doi.org/10.11594/ijmaber.05.09.10

Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education, 10*(2), Article 2236469. https://doi.org/10.1080/2331186X.2023.2236469

Min, H.-T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing, 15*(2), 118–141. https://doi.org/10.1016/j.jslw.2006.01.003

Mohammadi, R., Borna, P., & Zoghi, R. (2023). The effect of employing Grammarly on Iranian EFL learners' writing performance. *International Journal of Research in English Education,* 8(5),187-197. Retrieved May 19, 2024, from http://ijreeonline.com/article-1-869-fa.html

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*(4), 375–401. https://doi.org/10.1007/s11251-008-9053-x

Nicol, D. J. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education, 35*(5), 501–517. https://doi.org/10.1080/02602931003786559

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1). https://doi.org/10.1177/1609406917733847

O'Neill, R., & Russell, A. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology, 35*(1), 42–56. https://doi.org/10.14742/ajet.3795

Paltridge, B. (2012). *Discourse Analysis: An Introduction* (2nd ed.). London, UK: Bloomsbury.

Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods* (3rd ed.). Thousand Oaks, CA: Sage.

Raihan, R., Febriani, R., & Irianti, L. (2023). EFL students' perception of using Grammarly in undergraduate paper writing. *International Journal of Innovation and Education Research, 2*(1), 26–36. https://doi.org/10.33369/ijier.v2i1.29089

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd ed.). Cambridge, UK: Cambridge University Press.

Roscoe, R. D., & McNamara, D. S. (2013). Writing pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology*, *105*(4), 1010–1025. https://doi.org/10.1037/a0032340

Roscoe, R. D., Wilson, J., Johnson, A., & Mayra, C. (2017). Presentation, expectations, and experience: Sources of student perceptions of automated writing evaluation. *Computers in Human Behavior, 70*, 207–221. https://doi.org/10.1016/j.chb.2016.12.076

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133–164). Amsterdam, Netherlands: John Benjamins.

Sarıcaoğlu, A., & Bilki, Z. (2021). Voluntary use of automated writing evaluation by content course students. *ReCALL: Journal of EuroCALL, 33*(3), 265–277. https://doi.org/10.1017/S0958344021000021

Shao, Z., Gong, Y., Shen, Y., Huang, M., Duan, N., & Chen, W. (2023). Enhancing retrieval-augmented large language models with iterative retrieval-generation synergy. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 9248–9274. https://doi.org/10.18653/v1/2023.findings-emnlp.620

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Steiss, J., Tate, T., Graham, S., Cruz, R., Hebert, M., Wang, J., Moon, Y., Tseng, A., & Warschauer, M. (2023). Comparing the quality of human and ChatGPT feedback on students' writing. *OSF Preprints*. https://doi.org/10.35542/osf.io/ty3em

Teng, F., & Zhang, L. J. (2016). A questionnaire-based validation of multidimensional models of self-regulated learning strategies. *The Modern Language Journal, 100*(3), 674–701. https://doi.org/10.1111/modl.12339

Teng, M. F. (2025). Metacognitive awareness and EFL learners' perceptions and experiences in utilising ChatGPT for writing feedback. *European Journal of Education, 60*(1), Article e12811. https://doi.org/10.1111/ejed.12811

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*(2), 327–369. https://doi.org/10.1111/j.1467-1770.1996.tb01238.x

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing, 16*(4), 255–272. https://doi.org/10.1016/j.jslw.2007.06.003

Villamil, O. S., & De Guerrero, M. C. M. (1996). Peer revision in the L2 classroom: Social-cognitive activities, mediating strategies, and aspects of social behavior. *Journal of Second Language Writing, 5*(1), 51–75. https://doi.org/10.1016/S1060-3743(96)90015-6

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. & Trans.). Cambridge, MA: Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Wang, X. (2025). Students' perceptions of automated writing evaluation in autonomous learning. *Education and Information Technologies*, 1–33. https://doi.org/10.1007/s10639-025-13378-y

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research, 10*(2), 157–180. https://doi.org/10.1191/1362168806lr190oa

Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education, 100*, 94–109. https://doi.org/10.1016/j.compedu.2016.05.004

Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research, 58*(1), 87–125. https://doi.org/10.1177/0735633119830764

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2016). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*, Article 3087. https://doi.org/10.3389/fpsyg.2019.03087

Wu, Y., & Schunn, C. D. (2020). The effects of providing and receiving peer feedback on writing performance and learning of secondary school students. *American Educational Research Journal, 58*(4), 795–830. https://doi.org/10.3102/0002831220945266

Yoon, S.-Y., Miszoglad, E., & Pierce, L. R. (2023, October 10). Evaluation of ChatGPT feedback on ELL writers' coherence and cohesion [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2310.06505

Zhang, J., & Zhang, L. J. (2022). The effect of feedback on metacognitive strategy use in EFL writing. *Computer Assisted Language Learning*, 1–26. https://doi.org/10.1080/09588221.2022.2069822

Zhang, V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing, 43*, 100439. https://doi.org/10.1016/j.asw.2019.100439

Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing, 36*, 90–102. https://doi.org/10.1016/j.asw.2018.02.004

Zhang, J., Bao, K., Zhang, Y., Wang, W., Feng, F., & He, X. (2023). Is ChatGPT fair for recommendation? Evaluating fairness in large language model recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems* (pp. 993–999). New York, NY: Association for Computing Machinery. https://doi.org/10.48550/arXiv.2309.07803

Zou, S., Guo, K., Wang, J., & Liu, Y. (2025). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: A comparison study. *Computer Assisted Language Learning*, 1–30. https://doi.org/10.1080/09588221.2024.2447279

APPENDICIES

Appendix:1: Semi Structered Interview Forms with the Consent Form

Informed Consent Form

The study titled "A Comparison of AI-Based Chatbot Feedback and Teacher-Written Feedback in EFL Writing Classes" will be conducted by Sema Yıldız Huseynov. The research is planned to compare the effectiveness and perceptions of AI-based chatbot feedback with teacher-written feedback. Participation in this research is voluntary. Even if you have agreed to participate in the study and have started the interview, you can decide to withdraw from the study if you wish. Answering all interview questions completely and consenting to audio recording implies your voluntary participation in the research.

All the information obtained from you will be used in a scientific research. The results of the research will not contain any names or signs that could reveal identities. All records related to you in this research will be kept confidential and will be completely destroyed after 5 years.

Therefore, it is very important that you answer all questions truthfully and completely. If you have any questions about the study, you can contact the responsible researcher Sema Yıldız Huseynov at the contact details below without hesitation.

Thank you for taking the time for our study and for your contribution to the scientific world...

Researcher Title, Name Surname: Lecturer Sema YILDIZ HUSEYNOV

Phone Number:

Email address:

Consent to Participate in the Study:

- I have read the information provided above that should be given to the participant/volunteer before the research and fully understood the scope and purpose of the study I am asked to participate in, as well as the responsibilities that fall on me voluntarily.
- I understand that my voice will be recorded during the interview within the scope of the study and will be used as data in the study.
- The written and oral explanation about the study was made by the researcher named above, I had the opportunity to ask questions and discuss, and I received satisfactory answers.
- The potential risks and benefits of the study were also explained to me orally.
- I understand that I can leave this study at any time and for any reason without having to provide a reason, and that I will not face any negative consequences if I do so.
- Under these conditions, I agree to participate in the aforementioned research of my own will, without any pressure or coercion.
- Below, the participant will sign in their own handwriting in the appropriate place.

Participant's signature:

Participant's name and surname:

Date:

(Optional to be filled in)

Semi- Structured Interview Form for Students (Feedback Perception):

Participant Details:

- **Name (Optional): _____**
- **Age: _____**
- **Years of EFL Study: _____**

Questions:

1. Which feedback helped you more with your writing errors, the one from AI or the teacher? Can you explain why?
    - _____
    - _____
2. Did you find the AI feedback clear and easy to understand?
    - _____
    - _____
3. Did you find the AI feedback clear and easy to understand?

4. Which feedback was more helpful for your writing, the one from AI or the teacher?
    - _____
    - _____
5. How helpful do you think the AI feedback will be for your future writing?
    - _____
    - _____
6. How helpful do you think the AI feedback will be for your future writing?

7. Were there any parts of the feedback from AI that were really helpful or confusing for you? Can you give examples?
    - _____
    - _____
8. Were there any parts of the feedback teacher that were really helpful or confusing for you? Can you give examples?

9. Do you have any other thoughts or suggestions about how you received feedback on your writing?
    - _____
    - _____

Semi- Structured Interview Form for Teachers (Feedback Evaluation):

Participant Details:

- **Name (Optional): _____**
- **Years of Teaching Experience: _____**

Questions:

1. After reviewing the feedback provided on the student writings, which feedback source (AI-generated or teacher-generated) do you believe was more effective in identifying and addressing errors? Why?
    - _____
    - _____
2. How would you rate the clarity and comprehensibility of the feedback provided on the student writings from AI?
    - _____
    - _____
3. How would you rate the clarity and comprehensibility of the feedback provided on the student writings from teachers?

4. In your opinion, which feedback source provided more relevant comments and corrections for the student writings?
    - _____
    - _____
5. How valuable do you believe AI-generated feedback is for students' future writing improvement?
    - _____
    - _____
6. How valuable do you believe teacher feedback is for students' future writing improvement?

7. Were there specific comments or corrections in the feedback that stood out as particularly effective? Please specify and mention the source (AI or teacher).

- _____
- _____

8. Were there specific comments or corrections in the feedback that stood out as particularly ineffective? Please specify and mention the source (AI or teacher).

9. Any additional insights or suggestions regarding the feedback process or the potential of AI in EFL writing feedback?

    - _____

Appendix:2: Maltepe University Ethical Clearance

Bu belge, Yükseköğretim Kurulu tarafından 19.01.2021 tarihli "Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge" ile bildirilen 6698 Sayılı Kişisel Verilerin Korunması Kanunu kapsamında gizlenmiştir.

Appendix:3: Tables Showing Frequencies of Feedback Focus For Each Student

Feedback Type Distribution for ST1

Feedback Type	**ChatGPT**	**FDT1**	**FDT2**	**FDT3**	**Grammarly**
Grammar & Mechanics	5	3	4	6	8
Vocabulary	1	0	1	1	2
Organization	2	1	1	3	0
Content/Ideas	1	2	1	3	0
Style & Tone	0	0	0	1	0
Total Feedback	9	6	7	14	10

Feedback Type Distribution for ST2

Feedback Type	**ChatGPT**	**FDT1**	**FDT2**	**FDT3**	**Grammarly**
Grammar & Mechanics	6	3	5	7	11
Vocabulary	1	0	1	2	1
Organization	1	1	1	3	0
Content/Ideas	2	1	2	4	0
Style & Tone	0	0	0	1	0
Total Feedback	10	5	9	17	12

Feedback Type Distribution for ST3

Feedback Type	**ChatGPT**	**FDT1**	**FDT2**	**FDT3**	**Grammarly**
Grammar & Mechanics	4	2	4	5	9
Vocabulary	1	1	1	2	1
Organization	2	1	2	3	0
Content/Ideas	2	0	2	4	0
Style & Tone	0	0	1	1	0
Total Feedback	9	4	10	15	10

Feedback Type Distribution for ST4

Feedback Type	**ChatGPT**	**FDT1**	**FDT2**	**FDT3**	**Grammarly**
Grammar & Mechanics	5	3	4	6	9
Vocabulary	1	1	1	2	1
Organization	2	0	2	3	0
Content/Ideas	2	1	2	4	0
Style & Tone	0	0	0	1	0
Total Feedback	10	5	9	16	10

Feedback Type Distribution for ST5

Feedback Type	**ChatGPT**	**FDT1**	**FDT2**	**FDT3**	**Grammarly**
Grammar & Mechanics	6	2	4	7	11
Vocabulary	1	0	1	1	2
Organization	2	1	2	3	0
Content/Ideas	2	1	2	4	0
Style & Tone	0	0	0	1	0
Total Feedback	11	4	9	16	13

Feedback Type Distribution for ST6

Feedback Type	**ChatGPT**	**FDT1**	**FDT2**	**FDT3**	**Grammarly**
Grammar & Mechanics	7	2	5	8	12
Vocabulary	1	0	1	1	2
Organization	2	1	2	3	0
Content/Ideas	1	1	1	4	0
Style & Tone	0	0	0	1	0
Total Feedback	11	4	9	17	14

Feedback Type Distribution for ST7

Feedback Type	**ChatGPT**	**FDT1**	**FDT2**	**FDT3**	**Grammarly**
Grammar & Mechanics	5	3	4	6	10
Vocabulary	1	1	0	1	1
Organization	1	1	1	2	0
Content/Ideas	1	0	1	3	0
Style & Tone	0	0	0	1	0
Total Feedback	8	5	6	13	11

Feedback Type Distribution for ST8

Feedback Type	**ChatGPT**	**FDT1**	**FDT2**	**FDT3**	**Grammarly**
Grammar & Mechanics	6	3	5	7	10
Vocabulary	1	0	1	1	1
Organization	2	1	1	2	0
Content/Ideas	1	1	2	4	0
Style & Tone	0	0	0	1	0
Total Feedback	10	5	9	15	11

Feedback Type Distribution for ST9

Feedback Type	ChatGPT	FDT1	FDT2	FDT3	Grammarly
Grammar & Mechanics	5	2	4	6	9
Vocabulary	1	0	1	1	1
Organization	2	1	1	3	0
Content/Ideas	2	1	1	3	0
Style & Tone	0	0	1	1	0
Total Feedback	10	4	8	14	10

Feedback Type Distribution for ST10

Feedback Type	ChatGPT	FDT1	FDT2	FDT3	Grammarly
Grammar & Mechanics	6	2	5	7	11
Vocabulary	1	0	1	1	1
Organization	1	1	2	2	0
Content/Ideas	1	0	1	3	0
Style & Tone	0	0	0	1	0
Total Feedback	9	3	9	14	12

Feedback Type Distribution for ST11

Feedback Type	**ChatGPT**	**FDT1**	**FDT2**	**FDT3**	**Grammarly**
Grammar & Mechanics	7	3	6	8	13
Vocabulary	1	0	1	1	2
Organization	2	1	2	3	0
Content/Ideas	2	1	1	4	0
Style & Tone	0	0	0	1	0
Total Feedback	12	5	10	17	15

Feedback Type Distribution for ST12

Feedback Type	**ChatGPT**	**FDT1**	**FDT2**	**FDT3**	**Grammarly**
Grammar & Mechanics	6	2	5	8	12
Vocabulary	1	1	1	2	2
Organization	2	0	2	3	0
Content/Ideas	2	0	1	3	0
Style & Tone	0	0	0	1	0
Total Feedback	11	3	9	17	14

Feedback Type Distribution for ST13

Feedback Type	ChatGPT	FDT1	FDT2	FDT3	Grammarly
Grammar & Mechanics	7	2	6	9	13
Vocabulary	1	0	1	1	2
Organization	2	1	2	2	0
Content/Ideas	1	0	1	3	0
Style & Tone	0	0	0	1	0
Total Feedback	11	3	10	16	15

Feedback Type Distribution for ST14

Feedback Type	ChatGPT	FDT1	FDT2	FDT3	Grammarly
Grammar & Mechanics	5	2	5	7	12
Vocabulary	1	0	1	1	2
Organization	1	1	1	2	0
Content/Ideas	2	0	1	3	0
Style & Tone	0	0	0	1	0
Total Feedback	9	3	8	14	14

Feedback Type Distribution for ST15

Feedback Type	**ChatGPT**	**FDT1**	**FDT2**	**FDT3**	**Grammarly**
Grammar & Mechanics	6	3	4	8	11
Vocabulary	1	1	1	1	1
Organization	1	1	2	2	0
Content/Ideas	1	0	1	3	0
Style & Tone	1	0	0	1	0
Total Feedback	10	5	8	15	12

CURRICULUM VITAE

Sema YILDIZ HUSEYNOV

Education

Degree	*Year*	*University, Institute, Department*
MA	2025	Maltepe University, Graduate School, Department of Foreign Languages, English Language Teaching
BA	2012	Middle East Technical University, Educational Sciences, Department of Foreign Language Education
High School	2008	Kırıkkale Anatolian Teacher Training High School

Work Experience

Year	Job
2012 -2013	English Teacher, Ankara Aziziye Anatolian High School
2017-Current	English Instructor, Piri Reis University

Publications

Huseynov, S. Y. (2020). Yükseköğretimde İngilizce derslerinde teknoloji kullanımına ilişkin öğretmen görüşleri [Teacher opinions on the use of technology in English classes in higher education]. *Ege Eğitim Teknolojileri Dergisi*, *4*(1), 1–13. https://dergipark.org.tr/tr/pub/eetd/issue/69573/845377

Huseynov, S. Y. (2019). Blog kullanımının EFL öğrencilerinin yazma becerisine etkisi ve öğrencilerin blog kullanarak yazmaya ilişkin düşüncelerinin incelenmesi [The effect of using blogs on EFL students' writing skills and an investigation of students' opinions on writing by using blogs]. *Ege Eğitim Teknolojileri Dergisi*, *3*(2), 28–41. https://dergipark.org.tr/tr/pub/eetd/issue/50897/655885

Huseynov, F., Huseynov, S. Y., & Özkan, S. (2016). The influence of knowledge-based e-commerce product recommender agents on online consumer decision-making. *Information Development, 32*(1), 81–90. https://doi.org/10.1177/0266666914537827

Conferences

Demirkan, Ö., & Huseynov, S. Y. (2021, June 1–2). *How English teachers integrate teaching technologies in their courses* [Paper presentation]. International Asian Congress on Contemporary Sciences-V, Nakhchivan, Azerbaijan. https://www.asyakongresi.org/

Workshops

Huseynov, S. Y. (2025, May 2). *Creativity and problem-solving techniques for English instructors*. ELT CONNECT: Sharing Voice, Shaping Practice, Piri Reis University, Turkey.