

REPUBLIC OF TURKEY
SULEYMAN DEMIREL UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES
DEPARTMENT OF FOREIGN LANGUAGES EDUCATION

**RASCH ANALYSIS OF TEACHER, ARTIFICIAL INTELLIGENCE AND
HYBRID FEEDBACK ON EFL STUDENTS' WRITING**

Ali Dinçay AYAN

Advisor: Assoc. Prof. Dr. Nihan ERDEMİR

Co-Advisor: Asst. Prof. Dr. Sabahattin YEŞİLÇINAR

MASTER'S THESIS

ISPARTA, 2025

© 2025 by Ali Dinçay AYAN. All rights reserved.

TABLE OF CONTENTS

TABLE OF CONTENTS	i
ABSTRACT	iv
ÖZET	vi
ACKNOWLEDGEMENTS	viii
LIST OF TABLES.....	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xi
OPERATIONAL DEFINITIONS	xii
1. INTRODUCTION	1
1.1. Statement of the Problem.....	1
1.2. Purpose of the Study	3
1.3. Significance of the Study	4
1.4. Assumptions.....	5
1.5. Limitations.....	6
2. LITERATURE REVIEW.....	7
2.1. Writing Skills in English Language Teaching	7
2.2. Writing Evaluation.....	8
2.2.1. Automated writing evaluation.....	9
2.3. Writing Feedback.....	12
2.3.1. Higher-order and lower-order concerns.....	13
2.3.2. Teacher feedback	15
2.3.3. AWE feedback	16
2.3.4. AWE-supported teacher feedback	17
2.4. Artificial Intelligence	18
2.4.1. The use of AI in language learning and teaching.....	18
2.5. Chatbots	19
2.5.1. Teachers' and students' perceptions regarding chatbots	20
2.5.2. Concerns regarding chatbot integration in language learning and teaching.....	21
2.6. ChatGPT	22
2.6.1. The use of ChatGPT in education.....	24
2.6.2. ChatGPT in language learning and teaching	26
2.6.3. ChatGPT and language feedback.....	28

2.6.4. Ethical considerations and limitations of ChatGPT.....	30
2.7. Relevant International Studies	31
2.8. Relevant National Studies.....	32
3. METHODOLOGY	35
3.1. Research Design	35
3.2. Participants	36
3.3. Data Collection Tools	38
3.3.1. Students' essays	38
3.3.2. Semi-structured interviews	39
3.4. Data Collection Procedure	39
3.5. Data Analysis	44
3.5.1. Quantitative data analysis	44
3.5.2. Qualitative data analysis	45
4. PILOT STUDY	48
4.1. Research Design of the Pilot Study	48
4.2. Participants of the Pilot Study	48
4.3. Data Collection Tools and Procedure of the Pilot Study	49
4.5. Data Analysis of the Pilot Study	49
4.6 Findings of the Pilot Study	50
5. RESULTS	53
5.1. Quantitative Research Results	53
5.1.1. To what extent do different feedback types address higher-order concerns and lower-order concerns in EFL students' writing?.....	53
5.1.1.1. Do the essays differ in main feedback categories (HOCs and LOCs) across different feedback types (teacher, ChatGPT, and hybrid)?.....	53
5.1.1.2. Do the raters differ in feedback categories and patterns across different feedback types (teacher, ChatGPT, and hybrid) based on single raters and combined rater types?	55
5.1.1.3. Are the feedback categories in the rubric distinctive across subcategorized benchmark facets, rater facets, rater type facets, and the main categorized benchmark facets?	57
5.1.1.4. How does the interaction between feedback types and given feedback frequency vary across main-feedback categories (HOCs and LOCs) and sub-feedback categories (content, organization, vocabulary, grammar, and mechanics)?.....	63

5.2. Qualitative Research Results	65
5.2.1. What are EFL teachers' perceptions of the strengths, challenges, and further suggestions for different feedback types on EFL students' writing?	66
5.2.1.1. What strengths and challenges do teachers perceive in each feedback type's ability to address main-feedback categories (HOCs and LOCs) and sub-feedback categories (content, organization, vocabulary, grammar, and mechanics)?.....	66
5.2.1.2. What suggestions do teachers have for improving each feedback type to better support EFL students' writing feedback categories, specifically in relation to HOCs and LOCs?	78
6. DISCUSSION	84
6.1. Interaction of Teacher, AI, and Hybrid Feedback across Feedback Categories	84
6.1.1. Variations in feedback categories across feedback types.....	84
6.1.2. Variations in feedback categories and patterns	86
6.1.3. Distinctiveness of feedback categories in the rubric	89
6.1.4. Variations in the interaction between feedback types and frequency across feedback categories	91
6.2. EFL Teachers' Perceptions.....	93
6.2.1. Strengths and challenges in feedback categories across feedback types	93
6.2.2. Suggestions for enhancing teacher, ChatGPT, and hybrid feedback	97
7. CONCLUSION.....	100
8. SUGGESTIONS FOR FURTHER STUDIES	102
REFERENCES	103
Appendix A. Ethics Committee Approval	130
Appendix B. Writing Tasks.....	132
Appendix C. Rubric for Main- and Sub-feedback Categories.....	133
Appendix D. Semi-structured Interview Questions.....	134
CURRICULUM VITAE	Hata! Yer işareti tanımlanmamış.

ABSTRACT

RASCH ANALYSIS OF TEACHER, ARTIFICIAL INTELLIGENCE AND HYBRID FEEDBACK ON EFL STUDENTS' WRITING

Ali Dinçay AYAN

**Master's Thesis, Süleyman Demirel University, Graduate School of Educational
Sciences, Department of Foreign Languages Education**

Advisor: Assoc. Prof. Dr. Nihan ERDEMİR

Co-Advisor: Asst. Prof. Dr. Sabahattin YEŞİLÇINAR

2025, 137 pages

Recently there has been a remarkable rise in interest regarding artificial intelligence (AI) tools in educational settings, particularly within language teaching and learning environments. As these technologies evolve, their potential to enhance traditional feedback mechanisms, particularly in writing instruction, is being explored extensively. This study investigates the effectiveness of types of writing feedback—teacher, AI and hybrid feedback—on tertiary-level English as a foreign language (EFL) students' writing. Specifically, the research focuses on the feedback categories including higher- and lower-order concerns in EFL instruction. The study employed an explanatory mixed-methods approach, integrating quantitative analysis of feedback types and qualitative insights from semi-structured interviews with EFL instructors. 80 essays written by tertiary-level EFL students were analyzed, with feedback from four experienced EFL teachers and an AI-based tool. A rubric adapted from existing literature categorized and assessed the feedback units. The results were statistically analyzed, incorporating Rasch analysis, to identify the interaction among feedback types and feedback categories. The results suggest that AI effectively provides feedback on lower-order concerns (LOCs) while less effective when addressing higher-order concerns (HOCs). On the other hand, the teachers effectively provide feedback on HOCs at the discourse level and they are less effective when addressing LOCs due to the workload of grammar, vocabulary and mechanics. Therefore, the hybrid feedback type, which involves EFL teachers supplementing AI's

feedback, demonstrated potential for improving the HOCs of students' writing, as well. The qualitative phase of the study was conducted by semi-structured interviews with EFL instructors to discover their perspectives on using AI-generated feedback following the completion of the quantitative analysis. This study demonstrated AI's capability to help reduce the workload of EFL teachers when giving feedback on students' written work at the sentence level. Thus, EFL teachers can concentrate more on the HOCs of EFL students' writing. This study concluded that AI-driven feedback integration with traditional teacher feedback could result in more effective writing instruction in EFL contexts. Contributing to the growing interest in AI in education, the research provides valuable insights for EFL teachers on how to incorporate AI technologies into their writing lessons.

Keywords: teacher feedback, artificial intelligence feedback, hybrid feedback, higher order concerns, lower order concerns

ÖZET

YABANCI DİL OLARAK İNGİLİZCE ÖĞRENEN ÖĞRENCİLERİN YAZILARINA VERİLEN ÖĞRETMEN, YAPAY ZEKA VE HİBRİT GERİ BİLDİRİM TÜRLERİNİN RASCH ANALİZİ

Ali Dinçay AYAN

Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi, Eğitim Bilimleri Enstitüsü,

Yabancı Diller Eğitimi Anabilim Dalı

Danışman: Doç. Dr. Nihan ERDEMİR

2. Danışman: Dr. Öğr. Üyesi Sabahattin YEŞİLÇINAR

2025, 137 sayfa

Son zamanlarda eğitim ortamlarında, özellikle de dil öğretimi ve öğrenme ortamlarında yapay zekâ (YZ) araçlarına yönelik ilgide kayda değer bir artış olmuştur. Bu teknolojiler geliştikçe, özellikle yazma öğretiminde geleneksel geri bildirim mekanizmalarını geliştirme potansiyelleri kapsamlı bir şekilde araştırılmaktadır. Bu çalışma, yazma geribildirim türlerinin (öğretmen, yapay zekâ ve hibrit geribildirim) İngilizce dil öğrenimi bağlamında hazırlık seviyesinde yabancı dil olarak İngilizce (EFL) öğrenen öğrencilerinin yazıları üzerindeki etkinliğini araştırmaktadır. Araştırma özellikle, EFL öğretiminde geribildirim kategorileri olan üst- ve alt-düzey kategorilere odaklanmaktadır. Çalışma, geri bildirim türlerinin nicel analizini ve EFL öğretmenleriyle yapılan yarı yapılandırılmış görüşmelerden elde edilen nitel verileri birleştiren açıklayıcı bir karma yöntem yaklaşımı kullanmıştır. Hazırlık seviyesindeki EFL öğrencileri tarafından yazılan 80 kompozisyon, dört deneyimli EFL öğretmeninden ve yapay zekâ tabanlı bir araç olan ChatGPT ile alınan geri bildirimlerle analiz edilmiştir. Mevcut literatürden uyarlanan geribildirim kategorilerini içeren rubrik, geribildirimleri kategorize etmiş ve değerlendirmiştir. Sonuçlar, geri bildirim türleri ve geribildirim kategorileri arasındaki etkileşimi belirlemek için Rasch analizi kullanılarak istatistiksel olarak analiz edilmiştir. Sonuçlar, YZ'nin alt düzey kategoriler (ADK) hakkında etkili bir şekilde geribildirim sağladığını, ancak üst düzey kategoriler (ÜDK) ele alırken daha az etkili olduğunu

göstermektedir. Öte yandan, öğretmenler söylem düzeyinde ÜDK'lara etkili bir şekilde geribildirim sağlarken, dilbilgisi, kelime bilgisi ve mekanik gibi kategorilerde iş yükü nedeniyle ADK'ları ele alırken daha az etkilidir. Bu nedenle, EFL öğretmenlerinin YZ'nin geribildirimiyle desteklendiği hibrit geribildirim türü, öğrencilerin yazılarında ÜDK'ların da iyileştirilmesi potansiyelini göstermiştir. Çalışmanın nitel aşaması, nicel analizin tamamlanmasının ardından, yazma derslerinde YZ kullanımı ve YZ tarafından oluşturulan geribildirim hakkındaki bakış açılarını keşfetmek için EFL öğretmenleri ile yarı yapılandırılmış görüşmeler yapılarak gerçekleştirilmiştir. Bu çalışma, YZ'nin, öğrencilerin yazılarına geribildirim verirken EFL öğretmenlerinin iş yükünü azaltmaya yardımcı olma kapasitesini göstermiştir. Böylece, EFL öğretmenleri öğrenci yazılarında ÜDK'lara daha fazla konsantre olabilmıştır. Bu çalışma, geleneksel öğretmen geribildirimine YZ destekli geribildirim entegrasyonunun, dil öğrenme bağlamlarında yazma derslerinde daha etkili geribildirim verme ile sonuçlanabileceği sonucuna varmıştır. Eğitimde YZ'ye artan ilgiye katkıda bulunan bu araştırma, aynı zamanda EFL öğretmenlerine YZ teknolojilerini yazma derslerine nasıl dahil edecekleri konusunda değerli bilgiler sunmaktadır.

Anahtar Kelimeler: öğretmen geribildirimi, yapay zeka geribildirimi, hibrit geribildirim, üst düzey kategoriler, alt düzey kategoriler

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to my advisor, Assoc. Prof. Dr. Nihan ERDEMİR, for her invaluable guidance and unwavering support throughout the preparation of my thesis. I am equally grateful to my co-advisor, Asst. Prof. Dr. Sabahattin YEŞİLÇINAR, for his insightful feedback and dedicated assistance, greatly contributed to completing this work.

I would also like to sincerely thank the esteemed members of my thesis committee: Prof. Dr. Nazlı BAYKAL, Assoc. Prof. Dr. Kağan BÜYÜKKARCI and Assoc. Prof. Dr. Galip KARTAL for their valuable comments, constructive feedback, and support during my thesis defense.

I would like to express my sincere gratitude to Assoc. Prof. Dr. Mehmet ŞATA for his invaluable support and guidance in conducting the Rasch analysis for this study. His expertise, patience, and insightful contributions have been instrumental in ensuring the accuracy and rigor of the analysis. I am also grateful to the participating students and teachers for their willingness to contribute to this study. Their support and cooperation are deeply appreciated.

I am also deeply thankful to my beloved family, whose steadfast encouragement and unconditional support have been pivotal to my academic journey.

LIST OF TABLES

Table 1. Historical development of ChatGPT	22
Table 2. Research design by research questions	36
Table 3. Demographic information of the EFL students	37
Table 4. Demographic information of the participant teachers	37
Table 5. The rubric used to categorize feedback	44
Table 6. Comparison of teacher and ChatGPT feedback	50
Table 7. ChatGPT versus hybrid feedback	51
Table 8. Measurement report of the essay facets	54
Table 9. Measurement report for the rater facets	55
Table 10. Measurement report on the facet of the feedback types	56
Table 11. Measurement report for the subcategorized benchmark facets	57
Table 12. Measurement report for the rater facets	60
Table 13. Measurement report for feedback type facets	61
Table 14. Measurement report for the main categorized benchmark facets	61
Table 15. Measurement report on the interaction between rater types and subcategories	64
Table 16. Measurement report on the interaction between rater types and main categories	65

LIST OF FIGURES

Figure 1. Roles of generative AI in education	25
Figure 2. Demonstration of data collection and analysis	35
Figure 3. Training structure for participating teachers	40
Figure 4. An example of teacher feedback process	41
Figure 5. The final prompt used for ChatGPT feedback	42
Figure 6. An example of ChatGPT feedback	42
Figure 7. Data collection and feedback process timeline	43
Figure 8. Data analysis in MaxQDA	46
Figure 9. Variable map for the tested model	59
Figure 10. Variable map for the tested model (second model)	63
Figure 11. Strengths of teacher, AI and hybrid feedback in MaxQDA	66
Figure 12. Challenges of teacher, AI and hybrid feedback in MaxQDA	72
Figure 13. Suggestions for teacher, AI and hybrid feedback in MaxQDA	78

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AWE	Automated Writing Evaluation
CAF	Complexity, Accuracy, Fluency
CEFR	Common European Framework of Reference
EFL	English as a Foreign Language
ELT	English Language Teaching
GPT	Generative Pre-trained Transformer
HOCs	Higher Order Concerns
LOCs	Lower Order Concerns
L1	First Language
L2	Second Language
SPSS	Statistical Package for the Social Sciences

OPERATIONAL DEFINITIONS

Artificial Intelligence: AI replicates the cognitive functions of human intelligence in computer systems. AI technologies can carry out activities that typically necessitate human cognitive skills, like comprehending spoken words, identifying associations, addressing challenges and creating solutions. AI tools are being increasingly utilized in various fields, including education, where growing applications have emerged as they improve learning experiences and offer creative solutions for teaching methods.

Automated Writing Evaluation: AWE utilizes computer-based systems for evaluation and feedback on written texts. This software examines different aspects of writing, such as grammar, punctuation, syntax and vocabulary. AWE tools help writers by providing immediate feedback, typically using algorithms and natural language processing methods to assess writing quality and recommend enhancements.

Chatbot: A chatbot is a computer program that engages with users by mimicking a conversation through text or voice. It can perform functions, respond to inquiries, and offer data using preset reactions or artificial intelligence algorithms.

ChatGPT: ChatGPT is an AI model made by OpenAI. It is designed to produce text resembling human writing using the input given. It can talk, respond to queries, clarify, and help with different tasks.

Higher-Order Concerns: Higher-order concerns are aspects of writing related to content, organization and coherence/cohesion, including clarity of thesis, coherence of argument, and depth of analysis at the discourse level.

Lower-Order Concerns: Lower-order concerns refer to surface-level writing errors such as grammar, punctuation, mechanics and spelling at the sentence level.

1. INTRODUCTION

The study aims to analyze teacher, artificial intelligence (AI) and hybrid feedback in terms of feedback categories including higher-order concerns (HOCs) and lower-order concerns (LOCs) in English as a foreign language (EFL) students' writing at the tertiary level. Therefore, this chapter presents the study's problem statement, purpose, significance, assumptions, and limitations.

1.1. Statement of the Problem

In today's world, every teacher needs to improve their technological knowledge to keep up with the latest advancements in their field. Nowadays, every field of study has seen some improvements or changes, including English language teaching (ELT). The field of language learning is rapidly evolving, particularly with the incorporation of new technologies in language education (González-Lloret, 2016). Language proficiency, assessment and evaluation skills, classroom management, instructional strategies, relationships with education stakeholders, and student achievement and involvement are all expected of English language teachers. Online teaching becomes obligatory one day because of a pandemic; hybrid learning becomes popular the next day, and teachers are always expected to be familiar with every new model. However, there are many time-consuming obligations that teachers need to pay attention to, such as evaluating students' work, preparing exams or quizzes, creating materials for the class, and providing feedback on students' writing. Due to these obligations, English language teachers sometimes lack the time to follow the latest developments. However, teachers struggle extensively to physically and pedagogically implement technology into every aspect of education, particularly in their classrooms. This struggle can often lead to feelings of overwhelm and frustration (Agyei & Voogt, 2011; Andrei, 2016; Kafyulilo et al., 2015).

According to modern educators, gaining writing ability in another language is considered one of the most challenging skills because it involves multiple stages like brainstorming, combining ideas, and revising (Oshima & Hogue, 2007). Studies emphasize the importance of understanding the cognitive processes in the development of writing skills to enhance instructional practices and highlight the roles of feedback, training, and motivation in improving writing skills (Agyei & Voogt, 2011; Andrei, 2016; Dewi et al.,

2019; Kafyulilo et al., 2015). Guiding students and providing regular feedback in writing classes are crucial for improving students' writing ability. This interaction helps students improve their writing abilities and become more proficient in expressing themselves in L2 (Graham et al., 2013; Reid, 1993; Vo, 2022). This process also raises students' awareness and ability to write more confidently and masterfully (Dung, 2020).

For this reason, teachers' written feedback is vital in guiding students to improve their potential as learners regarding their essay-writing skills (Hamid & Romly, 2021). However, providing detailed feedback on every written work can be challenging and takes a long time. Ferris (2003) stated that effective feedback demands a well-thought-out review and analysis of the student's writing, which can be time-consuming. EFL teachers require longer time to write detailed comments for each student on writing assignments due to large class sizes and additional administrative duties.

It is also likely that the EFL teachers may get overwhelmed while providing detailed feedback. Hattie and Timperley (2007) suggested that feedback given should have three characteristics: It should be specific, it should be acted upon, and it should be provided in a reasonable time frame. In this case, EFL teachers must provide students with quick, detailed, personalized feedback. At the same time, they must find ways to correct and improve students' writing abilities. Maintaining this load of work when providing feedback to many students might be challenging for EFL teachers.

EFL instructors frequently need help providing personalized feedback to students with diverse needs. Research conducted by Hyland and Hyland (2006) highlighted that every student has distinct language abilities, learning styles, and objectives. Therefore, EFL teachers must know these when giving feedback on students' written work. If they have limited resources and time, delivering individual feedback that meets each learner's needs can be challenging for educators.

EFL educators are required to complete a wide range of responsibilities. These include time management, assessing students, and giving timely feedback on students' written work. Nevertheless, providing immediate feedback requires significant time for EFL teachers. Automated Writing Evaluation (AWE) tools help make the feedback process on students' written assignments easier and have many benefits, such as boosting learner

autonomy and providing easy access to educational resources (Chen & Cheng, 2008). Therefore, these systems are currently known by many educational institutions and the adoption of these systems is on the rise in EFL writing instruction worldwide.

Aligned with the AWE systems, chatbots have also been investigated in educational studies (Okonkwo Ade-Ibijola, 2020). Chatbots offer immediate responses to learner inquiries, and it is stated that they foster learners' academic performance (Hiremath et al., 2018; Kim, 2019). They are accepted as practical, supportive tools that provide personalized assistance to learners (Pham et al., 2018). Furthermore, with the help of chatbots, students can access instructional resources rapidly, such as dictionaries and student portfolios, making it easier to assess their written pieces (Durall & Kapros, 2020).

A chatbot called ChatGPT has been very popular in the news lately. It can give feedback on students' written work like an EFL teacher. This study evaluates teacher, AI (ChatGPT-based), and hybrid feedback types to see if AI can effectively replace or complement the feedback given by EFL teachers in student writings. Comparing the feedback frequency and categories provided by AI to those provided by EFL teachers might clarify the possible advantages and drawbacks of including chatbots in writing instruction and feedback processes.

1.2. Purpose of the Study

Teacher feedback is essential to improve students' writing abilities. Ferris (1995) stated that students appreciate and benefit from teacher feedback, especially during early drafts of written work. Students perceive the feedback from their teachers as crucial for their growth as writers (Furqoni, 2019). Feedback, including praise, is valued by students and helps them become better writers. Nonetheless, obtaining feedback from AI represents an innovative strategy in the field of ELT and there exists a limited number of studies addressing this topic (Kohnke et al., 2023; Nugroho et al., 2023; Xiao & Zhi, 2023). This study seeks to address this gap by investigating teacher, AI, and hybrid feedback types focusing on feedback categories, specifically higher-order concerns (HOCs) and lower-order concerns (LOCs), within tertiary EFL classrooms. The aim is to assess whether the AI tool, a chatbot named ChatGPT, can be a valuable resource for EFL educators in delivering feedback on written work.

In this study, the following research questions are sought:

1- To what extent do different feedback types address higher-order concerns and lower-order concerns in EFL students' writing?

- a) Do the essays differ in main-feedback categories (HOCs and LOCs) across different feedback types (teacher, ChatGPT, and hybrid)?
- b) Do the raters differ in feedback categories and patterns across different feedback types (teacher, ChatGPT, and hybrid) based on single raters and combined rater types?
- c) Are the feedback categories in the rubric distinctive across subcategorized benchmark facets, rater facets, rater type facets, and the main categorized benchmark facets?
- d) How does the interaction between feedback types and given feedback frequency vary across main-feedback categories (HOCs and LOCs) and sub-feedback categories (content, organization, vocabulary, grammar, and mechanics)?

2- What are EFL teachers' perceptions of the strengths, challenges, and further suggestions for different feedback types on EFL students' writing?

- a) What strengths and challenges do teachers perceive in each feedback type's ability to address main-feedback categories (HOCs and LOCs) and sub-feedback categories (content, organization, vocabulary, grammar, and mechanics)?
- b) What suggestions do teachers have for improving each feedback type to better support EFL students' writing feedback categories, specifically in relation to HOCs and LOCs?

1.3. Significance of the Study

Offering feedback on students' written work is one of the most demanding responsibilities for EFL teachers due to its time-consuming nature and attention to detail. Constructive feedback is widely acknowledged as crucial for facilitating learning and development in various contexts, including educational and professional settings (Jiang & Wang, 2023; Li, 2021; Link et al., 2020; Wilson & Czik, 2016). According to Cho and MacArthur (2010), student writing feedback must be comprehensive, detailed, and personalized to meet each learner's needs. However, providing this level of individualized feedback often requires significant time and effort from EFL teachers, leading to a substantial workload.

EFL teachers' challenges are further compounded by large class sizes, making it difficult to provide immediate and timely feedback. Research has shown that delays in providing feedback can diminish its effectiveness, underscoring the importance of timely responses in improving student outcomes (Hattie & Timperley, 2007). In response to these challenges, many EFL teachers have started utilizing technological tools such as online platforms and automated writing evaluation (AWE) systems. Recent advancements in AWE tools have the potential to save time, maintain feedback quality, and enhance students' learning processes. Nonetheless, AWE systems can sometimes be unreliable and insufficient in offering detailed, context-specific feedback (Gikandi et al., 2011). For this reason, EFL teachers are advised to use AWE tools to complement rather than replace their personalized feedback. Adopting such an approach can improve student achievement while fostering a more engaging learning environment.

Despite the increasing use of AWE systems and AI tools in educational settings, the lack of national and international studies specifically comparing AI-generated feedback, teacher-provided feedback, and hybrid feedback in EFL contexts indicates a significant gap in the literature. The current study aims to bridge this gap by analyzing AI feedback, teacher feedback, and hybrid feedback, focusing on HOCs and LOCs in students' written work. By examining whether AI tools, teacher feedback, or a combination of both can serve as valuable resources for EFL teachers when providing feedback, this research seeks to uncover the potential advantages and limitations of these approaches. Filling this gap will contribute to understanding how AI technologies, in collaboration with teacher expertise, can be integrated into feedback practices and offer insights into their practical applications in EFL writing instruction.

1.4. Assumptions

It is assumed that the participant EFL teachers in the study gave truthful and precise answers during the data collection process. The feedback from EFL teachers and the students' completion of writing assignments play a crucial role in the study's results. Participants, especially the EFL teachers, declared their familiarity with ChatGPT, which may improve the reliability of the results and help their generalizability. Therefore, the results of this research, which focused on a group of Turkish EFL teachers and university students, can be applied to academic environments. The findings and results seen in this

research study indicate larger patterns in incorporating AI technology in language learning and comparing the efficacy of feedback generated by AI versus feedback generated by teachers.

1.5. Limitations

Although this study has made significant contributions, several limitations must be addressed, which may affect the generalizability and interpretation of the findings. First, the study included four EFL teachers and 40 students from one state university in Türkiye. This small and specific sample might not accurately reflect the larger population of EFL teachers and students, which could restrict the findings from being applied to different contexts, institutions, or regions.

Furthermore, the research took place in a particular academic environment, highlighting students enrolled at a state university in Türkiye. The study's results may not apply to other educational systems or universities in different nations due to differences in organizational culture, available resources, or student demographics.

Thirdly, the research heavily depended on the GPT-4 version of ChatGPT for producing AI-generated feedback. Although ChatGPT is an advanced AI tool, its feedback is restricted by the limitations of current natural language processing technologies. This addresses possible issues such as the inability to understand context, details, or students' specific needs, which could affect the quality and relevance of feedback compared to that provided by human educators.

Another limitation is the possibility of bias in self-reported data provided by participant teachers, particularly during the qualitative phase of the study. During interviews, the responses provided by participant teachers might have been influenced by social norms or fulfill anticipated standards. Therefore, although this study concentrated on the integration of AI with teacher feedback, it did not thoroughly investigate additional factors, including student motivation, engagement and the significance of peer feedback on the effectiveness of feedback. These factors might improve the understanding of the dynamics involved in delivering feedback on writing within educational settings and should be considered in future research initiatives.

2. LITERATURE REVIEW

2.1. Writing Skills in English Language Teaching

Writing is a complex skill because it involves various cognitive processes such as brainstorming, higher-order thinking abilities, producing and assessing, and lower-order thinking skills like remembering and processing (Bloom, 1974; Hyland, 2019; Nation, 2008). That's why, writers need to acquire specific linguistic knowledge and skills. This involves several aspects, including syntax, mechanics, semantic knowledge, and organizational skills. Acquiring information using listening and reading activities is essential to this process (Harmer, 2006). Therefore, teaching and assessing this competence pose significant challenges. It is stated by EFL teachers that mastering writing skills is difficult to acquire comparing other language skills (Hyland, 2019) since it demands a combination of numerous cognitive processes (Ahangari et al., 2014; Bai et al., 2020; Senel, 2018; Ying, 2018). The challenging nature of writing often leads students to make various errors, as they mainly transfer their native language patterns and try to use them in the target language.

Furthermore, mistakes in the same language arise from a lack of adequate understanding of the target language (Richards, 2015). Yağız et al. (2009) indicated that L2 writers encounter more significant challenges compared to L1 writers. The reason behind that is the influence of L1 on L2 writing. When learning a second language, learners often convert their first language into the one they acquire. However, when L2 writers do not fully acquire the target language, they tend to overgeneralize rules, use words and grammatical structures in the same way, misuse grammar rules, make spelling and punctuation mistakes, have trouble delivering meaning and commit other errors that stem from not fully understanding the target language (Kaweera, 2013; Phuket & Othman, 2015).

It has been stated that effectively transmitting meaning through writing is essential for academic accomplishment (Buitrigo & Diaz, 2017). However, the integration of writing into language learning syllabuses has been undervalued in the context of EFL. That's why, it is demanding for EFL teachers to have a more comprehensive approach to teaching writing within the context of language education. EFL teachers can better

support students in developing their proficiency in written expression to improve communication and academic outcomes.

Supporters of the input hypothesis claim that language acquisition occurs when learners have exposure to input. However, some other researchers believe obtaining input is insufficient to enhance foreign language competency. Similarly, Swain (1985) asserted that input alone does not ensure that students achieve the required level of accuracy and L2 competency. The output hypothesis proposed by Swain (1985) supported stated that producing output motivates students to concentrate on their language use, ultimately identifying discrepancies between the actual outputs and the desired outputs. Cumming (1990) established a correlation between the act of noticing and educational achievements in writing. He proposed that cognitive activities, such as selecting suitable vocabulary, identifying gaps, and evaluating language, contribute to the development of language skills. Manchón and De Larios (2007) suggested that writing is a comprehensive and cognitively demanding activity, necessitating that learners establish continuous goals until they complete the writing process. Therefore, producing a text is regarded as a form of problem-solving, and this accomplishment empowers students to increase their proficiency in the target language and gain a deeper understanding of L2.

Writing exercises in language instruction can also improve students' critical thinking skills, creativity, and overall communication abilities. By completing writing activities, students can learn to convey their thoughts and ideas in an organized and comprehensible way. This also improves their language proficiency, which is crucial for future academic and professional environments where strong writing skills are needed. As a result of recent developments and studies, researchers have noticed the importance of writing and its contributions to language acquisition (Cumming, 1990; Manchón & De Larios, 2007; Swain & Lapkin, 1995). Since adopting writing in an EFL context has encouraging results in language learning, it has recently become an essential part of language instruction worldwide.

2.2. Writing Evaluation

Essentially, evaluation collects evidence to assess whether an educational process is effective; it can inform necessary adjustments to improve the performance of both

learners and teachers. Studies describe evaluation as gathering information to determine to what extent teaching and learning goals are achieved as their intended objectives (Buttram & Covert, 1978; Ioannou-Georgiou & Pavlou, 2007). Evaluation is considered an act of judgment to determine quality or value (Scriven, 1991) and is often associated with assigning grades. The feedback obtained through evaluation provides students and teachers with insights into their progress, enabling them to identify areas that require further attention or revision to improve learning outcomes.

Information for evaluation can be collected using different kinds of tools like tests, questionnaires, assignments or observations. Even though evaluation and assessment are frequently used interchangeably in the literature, they are not synonymous. One key distinction between the two is that evaluation typically involves making judgments. For example, scoring a student's homework or assigning a grade are kinds of evaluations (Hibbert et al., 2012). On the other hand, assessment is more diagnostic, which means there are specific areas that need improvement, while evaluation tends to focus on overall success regarding whether a certain goal is achieved. Furthermore, Baehr (2013) claimed that assessment gives feedback to improve future performance and learning outcomes, whereas evaluation assists decision-making by assessing the level of achievement. Assessment does not necessarily reflect a quality judgment (Cen & Ross, 2009). Therefore, evaluation is considered a more comprehensive term encompassing assessment (Grainger et al., 2008). Despite these differences, evaluation and assessment share certain similarities. Both can be defined as formative, summative, process-oriented, or product-oriented based on their purpose, time, substance, and scope (Scriven, 1996). To summarize, while evaluation and assessment are closely related concepts that are frequently used interchangeably, evaluation is typically connected with judgment and has a greater scope than assessment. Additionally, both are essential components of EFL writing.

2.2.1. Automated writing evaluation

Automated writing evaluation (AWE) refers to assessing and evaluating written texts using computer-based tools and algorithms. It can provide feedback on writing, including grammar, syntax, style, coherence and overall quality. The AWE system is commonly

used in educational settings to improve students' writing skills, especially in language learning and composition courses. Several research findings indicate that after using AWE programs, second language learners' writing improves in grammatical and mechanical accuracy (Chen & Cheng, 2008; Grimes & Warschauer, 2010; Rock & Service, 2007).

Most research on AWE tools examines how valid and reliable their assessment scoring systems are, usually focusing on students' opinions and comparing the tool's performance and teacher feedback (Chen & Cheng, 2008; Grimes & Warschauer, 2010; Link et al., 2020). There is a debate about whether students favor AWE tools. In some studies, students thought AWE tools were beneficial since they helped them improve grammatical accuracy (Dikli & Bleyle, 2014). However, the grammar corrections students received in some studies were generally dissatisfactory (Chen & Cheng, 2008). Despite the mixed feedback, many students still see the value in using AWE tools to improve their writing skills. Li (2021) highlights that students appreciate the instant feedback and assistance AWE tools provide, even if it is imperfect. Similarly, Ranalli (2018) observed that students perceive the accessibility of AWE tools as a significant advantage despite their concerns regarding accuracy. Overall, while there may be some dissatisfaction with the grammar corrections provided by AWE tools, students generally believe that these tools can be helpful in their writing process.

Examining how teachers use and view automated feedback helps one to understand how AWE affects instruction. For example, Dikli (2010) and Dikli and Bleyle (2014) examined two AWE tools: *MyAccess* and *Criterion*. Though there were notable differences between the two forms of feedback, in both studies, feedback from teachers was better. Teacher feedback was more concise, focused, and individualized than its lengthy and generic automated counterpart. Instructors provided a greater amount of higher-quality, accurate, and formative feedback.

On the other hand, Wilson and Czik (2016) compared the amount, type, and level of teacher feedback on the L1 English writings of middle school students in two conditions: teacher feedback alone, and combined teacher and automated feedback. The study examined feedback types, including directive, query, informative, and praise, along with feedback levels addressing specific writing skills. The feedback levels were categorized

into lower-level and higher-level subskills. Despite equal feedback in both conditions, the AWE condition showed significantly more feedback on higher-level writing subskills than the teacher-feedback-only condition. Teachers could focus on fostering critical thinking, organization, and argument development by providing students with automated feedback on lower-level revisions. This study highlights the potential of AWE to enhance writing instruction in the classroom.

Only some studies have explored AWE's potential to encourage students to revise or rewrite their work. For example, after receiving feedback from *Criterion*, Chapelle et al. (2015) examined student revisions. The research classified revisions into six categories based on their presence and manner: no change, removal, addition, deletion, modification, and transposition. While 50% of feedback resulted in no change, the remaining 50% of revision acts were successful, indicating the utility of *Criterion's* feedback in guiding students' revision decisions. However, the study highlighted a lack of knowledge regarding factors influencing these decisions beyond the accuracy of automated feedback.

AWE feedback can promote writing development; however, the results could be more conclusive, making it insufficient to establish its validity in L2 writing classrooms. For instance, Li et al. (2015) explored how AWE feedback influenced the overall writing accuracy of ESL university students, utilizing the AWE tool *Criterion* error reports. They observed enhanced linguistic accuracy from initial to final drafts. A comparable study (Li et al., 2017) noted a short-term positive impact on accuracy, specifically in addressing run-on sentences, though it was only sometimes reflected in long-term gains across all error categories. Similarly, Wang et al. (2013) conducted a quasi-experimental study showing that, following assistance from the AWE tool *CorrectEnglish*, the experimental group demonstrated superior performance in grammar, word usage, and spelling compared to the control group. Findings suggest that utilizing AWE tools can improve specific aspects of writing accuracy, such as grammar, word usage, and spelling. However, long-term gains in overall linguistic accuracy may only sometimes be consistent across all error categories. Educators must consider the potential benefits and limitations of incorporating such tools into the writing instruction curriculum. Further research is needed to explore the most effective ways to integrate these tools for sustained improvement in writing skills.

2.3. Writing Feedback

Error correction and feedback in L2 writing are significant concerns for many researchers and educators. L2 instructors often report that correcting errors in student writing is one of the most challenging and exhausting aspects of their teaching practices (Ferris, 1999). According to Truscott (1996), nearly all language teachers engage in some form of grammar correction. However, this process can be problematic for both teachers and students: Students invest substantial effort for grammatical accuracy, while teachers face uncertainty regarding the long-term effectiveness of such corrections (Ferris, 1999). Truscott (1996) also noted that teachers frequently resist the idea that grammar feedback might be ineffective, and research findings alone may not be sufficient to change their perceptions of this issue.

The impact of grammar feedback on writing frequently suffers due to teachers' challenges in consistently providing feedback and the heavy workload they encounter, which includes evaluating a significant number of papers (Truscott, 1996). This burden can negatively impact the quality of the feedback provided. Moreover, inconsistent teacher feedback might hinder students' comprehension and utilization of corrections, a challenge raised by the time constraints teachers encounter while simultaneously addressing several mistakes. Even if a teacher could identify all errors and maintain consistency in grammar feedback, this focus on grammatical accuracy would likely detract from attention to the content and organization of ideas, making the process time-intensive and impractical.

On the other hand, teachers' comments on student writing were generally accurate and understandable to learners (Ferris, 2006). Research suggests that teachers must develop strategies and plan feedback sessions in the absence of a definitive guide for providing effective feedback (Ferris, 1999, 2004). This careful preparation could lead to more accurate feedback. However, Truscott (1999) critiques this approach, pointing out that such detailed preparation and learner-specific grammar correction are likely time-consuming and may divert attention from more critical aspects of writing. The substantial time commitment required for teachers and students to focus on grammar correction could diminish students' interest in writing, highlighting potential disadvantages associated with this feedback approach.

Strategy training for self-editing can significantly enhance learners' grammatical accuracy (Ferris, 1995). Ferris et al. (1997) noted that continuous feedback over a semester enhanced students' writing and fostered their self-editing abilities. According to the study's participants, receiving comments helped them improve as writers. Ferris (1999) suggested that L2 teachers should concentrate on improving the efficacy of their feedback rather than completely giving up on grammar correction. Ferris and Roberts (2001) conducted an experimental study revealing that students who got feedback enhanced their writing accuracy, whereas those who received less explicit feedback with error-type codes were more inclined to participate in self-editing. When learners are trained in self-editing and teachers provide focused and selective feedback, the outcomes can be very promising for improving students' writing abilities (Ferris, 1999, 2004, 2006). Ferris et al. (2013) also noted that learners had positive perceptions of written corrective feedback, especially when it was supplemented with techniques like revisions and interviews, which improved their writing accuracy.

2.3.1. Higher-order and lower-order concerns

Writing concerns can be categorized into two main categories: Higher-Order Concerns (HOCs) and Lower-Order Concerns (LOCs). HOCs focus on the overall structure, organization, argumentation, coherence, and clarity of ideas within a piece of writing at the discourse level. In contrast, LOCs deal with surface-level issues such as grammar, punctuation, spelling, and word choice, which are rule-based and more mechanical at the sentence level.

The complementary nature of automated feedback demonstrates a system's competent capacity to deliver input on LOCs (Ranalli et al., 2016). Because AWE can handle LOCs, it has been suggested that it could free up instructors' time to concentrate more on HOCs (Warschauer & Grimes, 2008; Wilson & Czik, 2016). Little empirical evidence, however, supports this statement. The limited number of studies that have examined the influence of AWE on teacher feedback indicates conflicting results, thus necessitating further investigation in this area (Jiang et al., 2020; Link et al., 2020; Wilson & Czik, 2016). Thus far, a limited number of studies have specifically examined the influence of AWE on teacher feedback.

In a quasi-experimental study by Wilson and Czik (2016), two eighth-grade English Language Arts (ELA) classes were divided into two distinct conditions: one group received feedback via the *Project Essay Grade (PEG) Writing* system combined with teacher input, while the other relied solely on teacher feedback. Teachers in both groups provided feedback according to their standard practices and the nature of this feedback was subsequently analyzed to evaluate how *PEG Writing* influenced the emphasis on HOCs compared to LOCs. The findings revealed that, despite using *PEG Writing* to supplement their feedback, teachers in the *PEG Writing* condition focused more heavily on LOCs. Conversely, teachers in the teacher-only feedback condition tended to prioritize HOCs. The study's authors concluded that while AWE systems like *PEG Writing* may assist in addressing LOCs, their impact on redirecting teacher focus toward HOCs was limited, with only minor differences observed between the two conditions.

In their study, Link et al. (2020) built on the work of Wilson and Czik (2016) by investigating the influence of the Educational Testing Service's AWE tool, *Criterion*, on teacher feedback, focusing specifically on EFL instructors in Iran. The research involved two groups: One received a combination of *Criterion*-generated feedback and teacher input, while the other relied solely on teacher feedback. Results indicated that teachers in the *Criterion* + teacher feedback condition provided less feedback overall compared to those in the teacher-only condition. Interestingly, *Criterion's* use did not result in an increased focus on HOCs, diverging from the findings of Wilson and Czik (2016). Instead, teachers in the *Criterion* + teacher feedback condition concentrated exclusively on HOCs, delegating LOCs to *Criterion*. In contrast, teachers in the teacher-only condition addressed both HOCs and LOCs, complicating direct comparisons between the two groups.

Unlike two comparative studies, Jiang et al. (2020) conducted a longitudinal qualitative study inside classrooms to investigate how *Pigai*, an automated feedback tool, affected Chinese EFL teachers' feedback practices. Two of the 11 EFL teachers did not trust *Pigai's* feedback and maintained traditional methods. Three teachers entirely relied on *Pigai*, reducing both the time spent and the amount of feedback provided. The remaining six teachers used *Pigai* alongside traditional methods, which saved their time and enabled them to focus more on HOCs. In line with the findings by Wilson and Czik (2016), the study revealed that educators continued to provide feedback on LOCs.

2.3.2. Teacher feedback

Many scholars and writing teachers are concerned about error correction and feedback in L2 writing. EFL teachers claim that correcting writing mistakes is a challenging and tiring aspect of their language instruction (Ferris, 1999). According to Truscott (1996), most EFL teachers have different methods to detect and correct grammar mistakes in their students' written work. Providing grammar feedback to students is vital to motivate and support them when learning a language (Ferris et al., 1997, 2013). However, there is an ongoing debate about the effectiveness of grammar correction in improving students' writing ability. It is suggested that direct teacher correction of errors may lead students to feel demotivated and discouraged (Truscott, 1996). Accordingly, demotivated and discouraged students in the classrooms generally have a fear of making mistakes which hinders their active participation actively in language classes.

Researchers believe that providing feedback on errors is crucial for language development and language accuracy. In other words, error correction helps students spot and correct their mistakes while fostering their language proficiency (Ellis et al., 2009). However, while correcting students' errors, teachers' maintaining students' motivation and engagement is also crucial (Ellis et al., 2009). Therefore, EFL teachers should create a supportive environment where students feel motivated to improve their writing skills while also feeling free to take risks and make mistakes while expressing themselves.

Regarding the role of teachers in providing feedback, Truscott (1996) provided evidence that teachers' poor ability to spot errors and their excessive workload, which includes grading a sizable volume of papers, compromises the efficacy of grammar feedback on writing. As a result, these factors might hurt the standard of teachers' feedback. Another issue with feedback problems associated with teachers is the need for more consistency in their comments, which hinders learners' comprehension of the feedback. This issue develops because of teachers' demanding schedules and the simultaneous need to address several errors. Nevertheless, even if the instructor could identify every mistake and maintain consistency in providing grammar feedback, it would hinder their ability to concentrate on content and concept organization, resulting in a significant amount of time consumed, which is impossible to implement. Alternatively, Ferris (2006) observed that the feedback provided by teachers regarding students' writing was determined to be both

precise and easily understandable for the learners. This suggests that there is a balance between providing detailed grammar feedback and focusing on the overall content and clarity of the writing. By finding a middle ground, instructors can still help students improve their language skills while also allowing them to develop their composition and critical thinking abilities. In this way, students can receive comprehensive feedback addressing their grammar errors and overall writing quality, which leads to more effective learning outcomes.

Even if providing feedback is essential to improve learners' writing ability, EFL teachers believe it is time-consuming and challenging, and sometimes self-editing and peer correction can also work in language classrooms (Ranalli, 2018). Therefore, with the use of technological tools, teachers can address the grammatical errors of students rapidly and reduce their workload (Folt et al., 2020). Thus, it can be said that AI tools have a dual role: While EFL teachers are saving their time, EFL students can also receive detailed immediate feedback on their pieces.

2.3.3. AWE feedback

With the help of using AWE tools, teachers may reduce their workload, especially while they are dealing with minor grammatical mistakes, and can focus more on critical aspects of writing, such as organization and content. AWE tools' ability to provide immediate responses to errors makes them faster than EFL teachers in providing feedback (Dikli, 2010; Dikli & Bleyle, 2014). Similarly, Chukharev-Hudilainen and Saricaoglu (2014, as cited in Ranalli et al., 2016) stated that AWE systems have notably improved feedback correction on LOCs and mechanical mistakes in students' writing. This might be because AWE systems are designed for accuracy and related to their lack of ability to deliver feedback related to HOCs.

Liao (2016) highlighted the benefits of AWE feedback. He indicated that AWE feedback reduces EFL teachers' workload when giving feedback by minimizing the time spent on repetitive, surface-level errors. This also allows them to focus more on the content and organization of the paper while promoting learner responsibility and fostering autonomy. Scholars such as Harvey-Scholes (2017) and Wang et al. (2013) recognized AWE systems as tools that enhance learner autonomy in language learning. Chen and Cheng

(2008) attributed the improvements in learner autonomy to using AWE systems within a process-based writing approach, which is advantageous because it enables learners to revise their writing. Additionally, Cotos (2011) emphasized the potential of AWE systems to support the process-based writing approach, which involves multiple drafting stages, as learners can write several drafts based on the feedback provided by the AWE system.

2.3.4. AWE-supported teacher feedback

Numerous studies (e.g., Chen & Cheng, 2008; Choi, 2010; Link et al., 2020) have favored combining AWE tools with teacher feedback rather than relying solely on the AWE tools. However, in Chen and Cheng's (2008) study, one teacher exclusively used AWE feedback, leading to student dissatisfaction by the end of the semester as they felt the need for teacher guidance. Typically, AWE tools were trusted to provide feedback on lower-order skills, while teachers focused on higher-order skills (Link et al., 2020).

Wilson and Cziki (2016) noted that students who received feedback from their teachers and an AWE tool demonstrated improvement in writing. However, there was no significant improvement in the final quality of their drafts. In another study, students who received teacher feedback improved 8 out of 9 CAF (complexity, accuracy, fluency) measures, while those using only an AWE tool improved solely in accuracy. It's worth mentioning that the AWE group continued to enhance their accuracy until the delayed post-test, a trend not observed in the teacher feedback group (Link et al., 2020). Shi et al. (2020) also found that using an AWE tool improved writing fluency and complexity.

Contrary to standard recommendations, Wali and Huijser (2018) relied only on feedback from the AWE tool. Over half of the students reported that the AWE tool helped improve their grammar, content, and vocabulary; however, they still felt the need for teacher feedback. In a related study, Choi (2010) discovered that students who used *Criterion* in combination with teacher feedback outperformed those who either did not use *Criterion* or used it optionally, particularly regarding overall writing quality and grammatical accuracy.

2.4. Artificial Intelligence

Artificial intelligence (AI) is a branch of science focused on enabling computer systems to perform mental functions as humans do. Research in AI centers on mimicking and enhancing the capabilities of computers to think, learn, solve problems, and make decisions. Nowadays, AI applications play an increasingly significant role through various tools and technologies in our daily lives. The field of AI encompasses multiple sub-disciplines, including machine learning, natural language processing, knowledge representation, problem-solving, and logic, which allows AI to operate in a complex and multifaceted domain (Russell & Norvig, 2016).

Transferring human thought processes to machines dates back to the ancient Greek philosopher Aristotle. However, despite its philosophical origins, practical applications did not materialize until the mid-20th century, with the advent of modern computer technology. The field of modern AI research gained momentum during this period. Early AI work emerged alongside the development of computers and electronic devices. A significant milestone in AI was the Turing Test, introduced by Alan Turing in the 1950s, which asked whether a machine could demonstrate thinking abilities (Turing, 2009). The formal beginning of AI as a field was marked by the Dartmouth Conference in 1956, where key figures such as John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon laid the foundation for AI research (McCarthy et al., 1955). Despite significant advancements, the period between 1974 and 1980 saw a downturn in AI progress, as negative critiques led governments to withdraw their support, which caused a period known as the AI winter. When the UK started supporting AI research due to competition with Japan, this pause came to an end (Öztürk & Şahin, 2018).

2.4.1. The use of AI in language learning and teaching

Artificial intelligence is now a major part of our lives and many studies in different areas examine AI tools from various perspectives. New AI technologies and tools emerge almost daily in today's world, and this makes it difficult to keep up with all the latest developments. AI offers opportunities for improved learning experiences, evidence-based decision-making, and modern judicial processes in various sectors (Mijwil et al., 2022; Perbawa, 2021). Organizations can improve their effectiveness, productivity, decision-

making processes, and overall performance by using AI systems. This ultimately leads to improved outcomes and advancements in diverse fields.

Several AI tools related to language learning and teaching have also emerged recently. AI has significantly impacted the EFL context by revolutionizing traditional methods. Integrating AI technology in language education has opened new possibilities for improving learning outcomes. AI can provide innovative teaching materials and platforms primarily through technologies like deep learning algorithms and natural language processing mechanisms (Menekşe, 2023). Advancements like intelligent English teaching systems have optimized language teaching models with the aim of improving teaching quality and learning outcomes (Meng-Yue et al., 2020).

The use of AI in language learning extends to various aspects, including teaching spoken English and evaluating mobile English learning platforms (Cui & Li, 2022; Li, 2022). AI technologies can enhance the ecological perspective of teaching, as stated in research that explored AI systems in flipped classrooms (Hong & Zhang, 2023). Furthermore, the application of AI in language education has paved the way for the emergence of chatbots by offering personalized and interactive learning experiences (Haristiani, 2019).

The impact of AI on language education is not limited to English. It extends to languages like Chinese, where innovative teaching methods are being developed with a focus on AI integration (Sumo & Bah, 2021). The potential of AI to transform teaching and learning practices is inevitable in higher education settings because AI has already begun to reshape the structure of learning processes and communication between educators and students (Popenici & Kerr, 2017). AI offers personalized, efficient, and engaging learning experiences. Thus, its role in language learning and teaching is becoming increasingly significant (Roll & Wylie, 2016).

2.5. Chatbots

The term chatbot originates from the word chatterbot. Michael Mauldin first introduced it in 1997 to describe robots that can communicate with humans (Smestad & Volden, 2019). Chatbots have recently evolved remarkably. Modern chatbots operate using complex algorithms and they can respond faster to user inputs than human beings

(Okonkwo & Ade-Ibijola, 2020; Shi et al., 2020). Chatbots use algorithms and natural language processing (NLP) and by utilizing these, they can interact with users through written text or audio (Brennan, 2006; Hwang & Chang, 2021). In a study, chatbots are described as AI software that has the ability to engage with users and provide human-like interactions in various contexts (Shawar & Atwell, 2007). Chatbots can carry out multiple tasks from offering travel plans to preparing exercises and answers for teachers. Thus, their potential impact on people's lives is undeniable (Copulsky, 2019). While ChatGPT is the most known AI chatbot, there are other examples of chatbot models like *Microsoft Copilot*, *Amazon Comprehend* and *Gemini* (Alier et al., 2024).

Today, these chatbots are increasingly recognized as valuable tools in various sectors. This includes educational settings since chatbots are believed to enhance students' learning experiences (Clarizia et al., 2018). In an educational setting, chatbots can accomplish various tasks. For instance, studies have shown that chatbots can significantly improve students' learning performance by providing real-time responses to queries, which creates a more personalized learning experience for students (Lin & Lin, 2023). With ongoing developments and a focus on AI, chatbots are expected to continue to evolve and take a crucial role in education.

2.5.1. Teachers' and students' perceptions regarding chatbots

EFL teachers and students are the primary users of chatbots when using them in the EFL context (Chuah & Kabilan, 2021). This made researchers focus on how EFL teachers and students view the use of chatbots in the classroom in numerous studies. For example, Bii et al. (2018) stated in their research that EFL teachers generally feel supportive when integrating chatbots into their classes. Because EFL teachers believe that chatbots improve students' learning by providing an interactive environment that increases their curiosity and comprehension (Hu & Wang, 2023). Similarly, faculty members also view chatbots as a valuable tool in language learning and teaching. Because they believe that chatbots have the ability to provide personalized feedback to students when improving their writing skills (Kohnke, 2023; Quasem et al., 2023). Hew et al. (2022) also found similar results in their study. Faculty members mostly valued chatbots' accessibility and ease of use when providing immediate feedback. This also might lessen their burden and save time.

While it is believed that chatbots can potentially improve language learning, their effectiveness in educational settings is skeptical. One of the main concerns is the limited ability of chatbots to provide meaningful feedback and engage in complex conversations. For instance, Yang et al. (2022) highlighted that while AI chatbots can serve as virtual agents who can talk with users, they often lack the process of natural language and provide contextually appropriate responses. This limitation can lead to a superficial learning experience without completing the required interaction when mastering a second language (Annamalai et al., 2023). Similarly, faculty members were concerned in terms of the accuracy of the language generated by chatbots and their ability to understand the details of human language (Sun & Hoelscher, 2023; Tlili et al., 2023).

2.5.2. Concerns regarding chatbot integration in language learning and teaching

Chatbots can improve student motivation and learning outcomes. They accomplish this by giving students easy access to various educational resources, such as dictionaries and exercises. Besides, chatbots can answer students' questions quickly and foster an interesting learning atmosphere. However, EFL teachers must be aware of several issues before integrating chatbots into language instruction.

Chatbots function by processing and analyzing extensive datasets, a capability that raises critical ethical concerns regarding the protection of students' privacy and identity (Ruane et al., 2019). For example, these systems may collect sensitive personal information, such as students' names, email addresses, and academic records (Liden & Nilros, 2020). Such data collection practices can significantly impact the willingness of both teachers and students to engage with chatbots in educational contexts due to privacy and security concerns (Shumanov & Johnson, 2021). To address these issues, chatbot systems must follow strict privacy and security standards, and educational institutions, along with educators, must ensure that all collected data is securely managed in accordance with applicable data protection regulations (Liden & Nilros, 2020).

Dinçer (2018) stated in his study that integrating technology effectively into teaching and learning practices is a significant challenge for teachers. Liden and Nilros (2020) emphasized that teachers need to get training or support from their institutions or professionals to learn how to incorporate chatbots into their teaching styles. This training

should demonstrate the effective use of chatbots in educational settings, such as integrating chatbots into lesson plans and student engagement with these tools.

The financial part is also critical when it comes to implementing and maintaining chatbots in an educational context. To keep chatbots updated with fast-evolving trends, ongoing programming and maintenance are required. Liden and Nilros (2020) support this view by highlighting the importance of regular maintenance and updates to ensure the chatbot remains practical and relevant to the changing needs of students.

2.6. ChatGPT

ChatGPT is a sophisticated language model developed by OpenAI¹ based on the Generative Pre-trained Transformer (GPT) architecture. The term GPT signifies its dual characteristics: It is generative, meaning it can produce new content, and pre-trained, reflecting that it has been initially trained on extensive textual data before undergoing a specific task. The GPT model is a type of AI that uses deep learning techniques to understand and generate human-like text based on the input it receives.

ChatGPT was released on November 30, 2022, and reached 100 million users within two months. This uptrend interest in ChatGPT has made it the fastest-growing application (Reuters, 2023). While it was first available only in English, ChatGPT now supports more than 50 languages (OpenAI, 2024). The latest version of ChatGPT (ChatGPT-4o) was released in May 2024. Table 1 below shows the key developments related to ChatGPT.

Table 1. Historical development of ChatGPT (OpenAI, 2024)

Date	Development
June 2018	Release of GPT (Generative Pre-trained Transformer) by OpenAI.
February 2019	Release of GPT-2, an improved version with 1.5 billion parameters, showing significant progress in language understanding and generation.

¹ It represents the organization's goal of developing AI technologies in a way that benefits humanity.

Table 1. (continued)

June 2020	Release of GPT-3 with 175 billion parameters.
January 2021	Announcement of ChatGPT, a special version of GPT-3 designed by OpenAI.
November 2022	Introduction of ChatGPT-3.5, an advanced version with enhanced understanding and response capabilities.
January 2023	Integration of ChatGPT with Microsoft Bing search engine, starting to provide AI-supported responses to search queries.
March 2023	Release of GPT-4 by OpenAI, a more advanced version capable of understanding and generating text and images.
May 13, 2024	Release of GPT-4o, a newest model that provides GPT-4-level intelligence that is much faster and improves on its capabilities across text, voice, and vision.
July 25, 2024	OpenAI introduced SearchGPT, a prototype AI search tool created to provide users with direct answers to their queries.
September 12, 2024	OpenAI revealed the GPT o1 model, stating that it "possesses human-like reasoning abilities."

Customer service, education, healthcare consulting and technical support are some example applications provided by ChatGPT. In their research, Memarian and Doleck (2023) have outlined several features of ChatGPT:

- **Text Generation and Editing:** ChatGPT can generate and edit various text types, including articles, stories, blog posts, and code snippets.
- **Question-Answering and Guidance:** It can respond to user inquiries and provide information and guidance.
- **Educational and Learning Support:** In education, ChatGPT can be utilized to create educational materials, support learning processes and assist with assignments.

- **Language Translation:** ChatGPT can translate between different languages, making it useful for multilingual communication and content creation.
- **Creative Writing and Script Development:** It can be used in creative writing projects to generate ideas and develop scripts.
- **Customer Service and Support:** As a chatbot, it can handle customer inquiries and provide support.
- **Research and Data Analysis:** ChatGPT can assist research projects by analyzing data, summarizing findings, and generating reports.

It is clear from ChatGPT's many features that it is much more than solely a text creator or translation tool. It can be used for various tasks, from creative writing to customer service. Thus, studying ChatGPT's potential is crucial in education because it can raise the progress of teaching and learning to a whole new level.

2.6.1. The use of ChatGPT in education

ChatGPT and other language models have the potential to transform education by offering students and educators innovative and effective learning tools. These modern technologies can improve educational processes by providing individual learning experiences and easy access to material and new concepts previously only available in textbooks. This represents a significant turning point in the growth of educational processes.

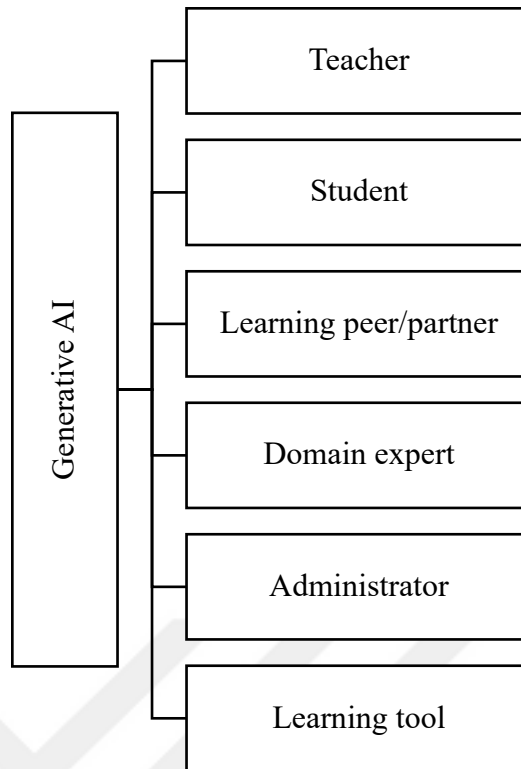


Figure 1. Roles of generative AI in education (Hwang & Chen, 2023)

According to Hwang and Chen (2023), the roles of generative AI applications in education are categorized in Figure 1:

- **Teacher/Tutor:** Supports students by summarizing learning content, providing examples, solving problems, or creating artwork.
- **Student/Tutee:** Learns from users, mainly when students take on the role of a teacher and assess the quality of tasks based on the information provided by the students.
- **Learning Peer/Partner:** Acts as a teammate in collaborative learning activities, working alongside students to complete specific tasks.
- **Domain Expert:** Serves as a specialist by offering advice on specific topics and providing solutions to problems.
- **Administrator:** Aids the decision-making process by conducting data analysis and offering summary reports.
- **Learning Tool:** Assists students in collecting and analyzing data, allowing them to focus on creative thinking.

Hwang and Chen's categories (2023) highlighted AI's diverse significance in education. Specifically, ChatGPT can contribute to the educational process by learning from users' input. It can assist learners in collaborative activities and provide expert advice on specific topics. In addition, it can support administrative work such as data analysis and reporting and assist students with data collection and analysis as a learning tool.

There are various perspectives on integrating ChatGPT into education. Some studies emphasize its positive aspects, such as its ability to enhance learning and teaching experiences through personalized learning and support for different educational approaches like flipped learning (Zhu et al., 2023) and its potential to increase student performance and motivation (Crawford et al., 2023; Lee, 2023). On the contrary, concerns have been raised about ChatGPT's potential issues, such as the spread of misinformation and inaccurate references (Fergus et al., 2023; Choi et al., 2023), academic integrity (Khan et al., 2023; Su & Yang, 2023) and a decline in students' creative and critical thinking skills (Su & Yang, 2023). Future research findings will shed light on how and through which methods ChatGPT can be effectively integrated into education.

2.6.2. ChatGPT in language learning and teaching

Recent advancements in deep machine learning technology have led to the development of tools capable of generating text and mimicking human writing. The number of AI-powered content writing tools, such as *RytrAI*, *JasperAI*, and *CopyAI*, continues to grow. Among these, ChatGPT stands out as the most influential tool which is described as a language model that interacts with users conversationally in the study of Deng and Lin (2022). Its user-friendly interface has contributed to its rapid adoption by many users (Gleason, 2022). Students' rapid adoption of ChatGPT has triggered a dramatic evolution in AI language learning, resulting in essential changes in foreign language instruction (Tlili et al., 2023).

In language education, the use of ChatGPT in language learning offers several advantages. It can be utilized for tasks, such as text generation, content editing, problem-solving, summarization, grammar correction, and answering questions (Kim et al., 2023). Kasneci et al. (2023) suggested that ChatGPT can generate human-like texts and natural dialogues for each individual and assist students in improving their writing skills and

vocabulary (Kasneci et al., 2023; Tlili et al., 2023). Furthermore, its extensive pre-training and intensive learning capabilities enable ChatGPT to produce contextually and semantically rich texts, making it a powerful tool for language learning (Raman et al., 2023).

The positive impacts of ChatGPT in the ELT context are highlighted in the study (Ali et al., 2023). Gilson et al. (2023) found that ChatGPT combines supervised and effective learning techniques to generate human-like texts and maintain user conversational exchanges. Su and Yang (2023) stated that ChatGPT provides students with a more personalized and practical learning experience while allowing teachers to give quicker and more accessible feedback. Qadir (2023) favored this feature in his study because it is particularly engaging for students. Similarly, Rahman and Watanobe (2023) noted that ChatGPT's deep understanding of language and writing skills could help students improve their language abilities. ChatGPT's advanced language abilities can facilitate the development of linguistic intelligence (Wu et al., 2023).

Altogether, some of the advantages of ChatGPT in language learning include:

- Having natural language processing capabilities
- Providing personalized and efficient learning experiences
- Engaging students
- Supporting the development of language skills
- Facilitating the enhancement of linguistic intelligence

ChatGPT has several positive effects on language learning. However, there are also potential negative aspects of ChatGPT related to issues such as cheating and assessment (Kohnke et al., 2023). One concern is that texts produced by ChatGPT may seem to originate from other sources without appropriate citation which could result in plagiarism. Second, ChatGPT can provide different and inconsistent answers from repeated entries, as pointed out in the study of Kohnke et al. (2023), which answers the possibility of inconsistencies in the text generated by ChatGPT. Third, ChatGPT's lack of human-like understanding and empathy could affect the students' knowledge and the quality of personalized feedback, which could hinder their language-learning process (Koubaa et al., 2023).

2.6.3. ChatGPT and language feedback

ChatGPT can serve multiple roles in the context of language learning and teaching. It can provide instant and personalized feedback to students on their written assignments, help them practice conversational skills and explain various language concepts. Studies have shown that AI-powered tools like ChatGPT can enhance students' learning experiences by providing timely and consistent feedback, thus improving their engagement and motivation (Al-Abdullatif & Alsubaie, 2022; Holmes & Tuomi, 2022).

GPT-4 is the latest version of ChatGPT, and it provides large language models for natural language processing tasks, including writing, summarizing, translating, and answering questions (Qu et al., 2023). ChatGPT performs well in these tasks through its two-stage comprehensive training on about 45 terabytes of web data (Dwivedi et al., 2023). In addition to previous training, ChatGPT acquires knowledge from everyday users who can submit input or upvote/downvote to help the chatbot respond better (AlAfnan et al., 2023).

Feedback provided by ChatGPT has been investigated in terms of teaching and learning as a possible educational aid. Although ChatGPT's output may not always be informative or original, ChatGPT can provide teachers with helpful feedback for improving classroom instruction (Chang et al., 2024). ChatGPT's feedback is typically more thorough and consistent for students, especially regarding student reports (Dai et al., 2023). Students' task performance in other areas, such as writing argumentative essays (Su et al., 2023) may be improved by ChatGPT-based feedback.

Theoretical studies about ChatGPT-based feedback have identified some advantages and benefits in this context of EFL learners. Firstly, ChatGPT provides instant and personalized feedback, enabling learners to make real-time improvements (Hong, 2023). Secondly, the feedback generated by ChatGPT is unlimited, offering students numerous opportunities for practice and refinement (Kim et al., 2023). Thirdly, interacting with ChatGPT exposes ESL/EFL learners to authentic language usage, thereby unconsciously enhancing their proficiency (Liu & Ma, 2023).

Regarding instant and personalized feedback to language learners, ChatGPT can evaluate students' tasks, correct grammar and vocabulary errors, and suggest alternative phrasing and expressions in real-time (Allehyani & Algamdi, 2023; Athanassopoulos et al., 2023; Nugroho et al., 2023; Shaikh et al., 2023). This can help learners identify and address their weaknesses, enhancing their language skills (Allehyani, 2023; Nugroho, 2023; Shaikh, 2023). In addition, ChatGPT's interactive and conversational nature can make language learning more engaging and motivating for students (Athanassopoulos et al., 2023; Zhang & Tur, 2023). It can act as a virtual language assistant, providing practice opportunities and boosting learners' confidence (Athanassopoulos et al., 2023; Nugroho, 2023). Furthermore, ChatGPT's ability to generate personalized content and adapt to individual needs can lead to more effective and tailored language instruction (Athanassopoulos et al., 2023; Zhang & Tur, 2023).

Studies have also explored the potential of integrating ChatGPT with teacher feedback, suggesting that the two can complement each other effectively (Wang & Han, 2022). While ChatGPT can address LOCs and provide initial feedback, teachers can then focus on HOCs, such as content and organization, and offer more personalized guidance (Wang & Han, 2022). Integrating both feedback sources, ChatGPT can support EFL teachers by reducing the workload of grading and offering feedback. As ChatGPT manages routine and repetitive tasks, educators can focus more on advanced educational priorities and activities, including curriculum development and personalized student interactions. Nonetheless, the incorporation of ChatGPT into educational methodologies requires careful consideration, such as data privacy, the reliability of feedback, and the risk of excessive dependence on AI resources. (Ruane et al., 2019; Shumanov & Johnson, 2021).

While ChatGPT offers numerous benefits, there are also some drawbacks and limitations to consider. Learners need to develop specific skills to use ChatGPT effectively for language practice, and over-reliance on the tool could hinder their ability to engage in authentic language use and understand cultural nuances (Javier & Moorhouse, 2023; Widianingtyas et al., 2023). Additionally, the language and cultural biases present in ChatGPT's training data may lead to inaccuracies or inappropriate responses, particularly in multilingual and multicultural contexts (Li et al., 2023). Careful integration of ChatGPT with other language teaching approaches and materials is necessary to provide a balanced and diverse learning experience (Li et al., 2023).

However, it is important to note that the effectiveness of ChatGPT's feedback, particularly in addressing HOCs, may depend on factors such as the learner's language proficiency, writing experience, and engagement with the feedback (Shen et al., 2023). Some studies have found that higher-proficiency-level learners may benefit more from ChatGPT's feedback on HOCs, while lower-proficiency-level learners may require more targeted support on LOCs (Geckin, 2020; Shen et al., 2023).

Currently, limited empirical research has tested the actual effectiveness of ChatGPT-based feedback in the EFL context. For instance, Mohamed (2023) and Nguyen (2023) interviewed teachers who viewed ChatGPT as an affordable and convenient tool for providing feedback. Furthermore, Schmidt-Fajlik (2023) reported positive student experiences with ChatGPT's feedback through questionnaire surveys. However, the study relied on self-reported outcomes and lacked statistical evidence. Therefore, further research is necessary to comprehensively understand the effectiveness of ChatGPT for EFL learners.

2.6.4. Ethical considerations and limitations of ChatGPT

ChatGPT faces several challenges and ethical considerations despite its capabilities. A major challenge of ChatGPT lies in ensuring the accuracy and reliability of the information it delivers. It may sometimes provide inaccurate and misleading answers to users' inquiries. That's why educators have expressed significant concerns about bias and discrimination in ChatGPT. For example, Kasneci et al. (2023) highlighted that large language models, like those used in ChatGPT, can inherit social biases from their training data. Such biases can lead to inaccurate and unfair responses, which can impact the learning negatively (Lund & Wang, 2023). Therefore, it is essential to ensure diverse training databases and continually monitor and test the model's performance across different demographic groups (Lund & Wang, 2023).

ChatGPT's ability to provide fast responses to users' inputs and superior performance compared to other AI technologies have led to its widespread popularity. However, Kasneci et al. (2023) indicate that the fast solutions provided by ChatGPT might hinder students' critical thinking and problem-solving abilities. Similarly, teachers who rely heavily on ChatGPT when preparing lesson plans or activities may experience a decline

in their teaching quality (Lund & Wang, 2023). However, if users understand the value of using ChatGPT as a supplementary tool rather than a replacement, these issues can be minimized (Pavlik, 2023).

2.7. Relevant International Studies

There are numerous studies that explored AI's educational applications in terms of benefits and drawbacks for teachers and students. One of the primary advantages of AI in education is its ability to personalize learning experiences. For instance, Patiño et al. (2024) emphasized that AI can adapt oral English teaching to suit each student's proficiency level, which optimizes the learning process through personalized feedback and targeted exercises. These tools can significantly promote learners' motivation and engagement by providing personalized learning patterns and immediate feedback (Holmes & Tuomi, 2022). Additionally, Ayala-Pazmiño (2023) and Holmes et al. (2019) examined the potential of AI technology in education, and their studies indicate that AI can improve the personal learning experience, facilitate the procedures of student performance evaluation and assessment as well as it can offer more efficient assistance to educators.

AI applications can ease teachers' burdens and workloads. AI tools can reduce teachers' workload and facilitate the tracking of student development, highlighting both current uses and future possibilities in education (Hashem et al., 2023). Additionally, research suggests that AI can adapt learning materials, monitor student performance, and provide teachers with practical guidance while addressing concerns about ethics, privacy, and data security (Crompton & Song, 2021).

Numerous studies over the world have explored the use of AI applications in language education. For example, using AI in teaching other languages, such as Arabic, has also effectively improved students' grammar and vocabulary skills by integrating new concepts into language instruction (Al-Abdullatif & Alsubaie, 2022). As educators adapt to developing teaching methods, AI provides creative solutions that promote student engagement and support effective learning experiences in educational contexts (Idham et al., 2024).

Furthermore, the higher education curriculum has begun to integrate extensive AI courses, which is crucial for preparing students for future professions. The Ministry of Education in several nations has launched action plans to promote AI innovation in higher education, highlighting the importance of multidisciplinary approaches to AI education (Zhang et al., 2022). The establishment of domain-specific AI courses is becoming increasingly important, as it allows students to apply AI concepts in practical contexts relevant to their fields of study (Schleiss, 2023). These advancements highlight the increasing significance of AI in the field of education (Nabiyev & Erumit, 2020). Consequently, these advancements led to the founding of the first AI university in Abu Dhabi and offering programs in machine learning, language models, and related fields.

2.8. Relevant National Studies

The First National Congress on AI Applications in Education was held in 2021, which marked a significant milestone in exploring and evaluating the application of AI in the educational sector in Türkiye. The status of AI technology in the context of education was examined through a literature review by Bozkurt et al. (2021). Their research indicated that AI is still a new concept and in the developmental stage in the educational system of Türkiye. They claimed that AI may offer many opportunities in educational settings. The literature review also revealed that numerous studies have been conducted on the usage of AI applications in education in Türkiye. These studies mainly focus on the contributions of AI applications to learning processes, their impact on students' learning experiences, and teachers' attitudes, awareness, and interest in AI technologies. For instance, Aktay et al. (2023) explored students' use of AI for studying different subjects. The students in the study believed that using ChatGPT improved their academic success. Students also stated that ChatGPT provides easy access to information and quick answers to their questions. However, accessibility is the biggest challenge for students when using ChatGPT in Türkiye's educational settings. Karabıyık's (2024) study showed that AI-supported instruction in subjects like mathematics can increase students' academic success and motivation toward learning as well as foster positive attitudes. Since AI offers personalized learning materials according to students' needs and learning styles, it is highlighted that the use of AI in elementary education improves learning outcomes (Çayır, 2023). This aligns with findings from Alkan (2024), who notes that AI

facilitates classroom management and automates assessment processes, allowing for quicker evaluation of assignments and improved feedback mechanisms.

There has also been an increase in research examining teachers' attitudes, awareness, and interest in AI in Türkiye. For example, Yetişensoy (2024) explored prospective teachers' attitudes and knowledge about AI in education. He found that teachers' technological knowledge affects their attitudes. While knowledgeable teachers generally have a positive attitude toward AI, some teachers may lack the knowledge to use the technology effectively. In a similar study, Güneyli et al. (2024) examined teachers' awareness and use of AI in educational technologies. He concluded that although teachers mostly approach AI-based educational technologies positively, they may not be fully prepared to use them effectively. Further research by Göçen and Aydemir (2020) investigated Turkish teachers' perceptions of AI, their usage levels, and their views on the role of AI in education. Although the participants generally seem to have positive perceptions of AI, there are also several negative aspects regarding the future of education.

EFL teachers' perceptions of AI technologies in education vary. Many recognize their potential benefits; conversely, others show low levels of awareness. While some research indicates that EFL novice teachers see AI as a valuable tool for classroom activities, assisting teachers and providing students' individual needs (Çam et al., 2021), some other studies indicate that EFL teachers' awareness of AI technologies remains low (Küçükkara et al., 2024). Despite the lower awareness among EFL teachers, younger generations, including those with significant academic qualifications, have practical knowledge of AI. Besides, they are willing to adopt these technologies in their classes (Çam et al., 2021; Uygun et al., 2024). In contrast, others also express their concerns that too much reliance on AI might block students' development, such as social skills, as interactions with AI systems cannot replicate interactions between humans (Akgün & Greenhow, 2021).

This study addresses a critical gap in the existing literature by focusing on the comparative effectiveness of teacher, AI, and hybrid feedback (combining teacher and AI feedback) types on feedback categories, HOCs and LOCs, in EFL students' writing. While numerous studies have explored the applications of AI, including ChatGPT, in education, few have rigorously examined how AI-generated feedback type compares to teacher feedback, particularly in the context of tertiary-level EFL writing instruction. This

research is characterized by its dual emphasis: First, on quantitatively and qualitatively assessing the efficacy of ChatGPT in providing meaningful feedback alongside human instructors, and second, on capturing the perceptions of EFL instructors regarding the integration of AI in their pedagogical practices. In this sense, this research addresses a significant gap in the literature. It aims to provide practical implications for integrating AI in language education, potentially leading to more effective and personalized feedback mechanisms in EFL writing classes.



3. METHODOLOGY

This chapter presents the main components of this study's methodology. Detailed information about the research design, participants, data collection tools and procedures, and data analysis was provided.

3.1. Research Design

This study aims to reveal the interaction among teacher, AI, and hybrid feedback in relation to HOCs and LOCs. It also seeks to identify EFL teachers' perceptions of using an AI tool for providing feedback on students' written work. Explanatory mixed-method research integrates quantitative and qualitative approaches within a single study to investigate and clarify a phenomenon (Creswell, 2014; Creswell & Clark, 2017). This study employed a sequential explanatory design because researchers can first collect and analyze quantitative data, followed by qualitative data to explain or elaborate on the quantitative findings. This approach allows for a richer interpretation of results, as it combines the strengths of both methods, enhancing the validity and depth of the research outcomes (Darmaji, 2023; Sweetman et al., 2010). The demonstration of data collection and analysis can be seen in Figure 2.

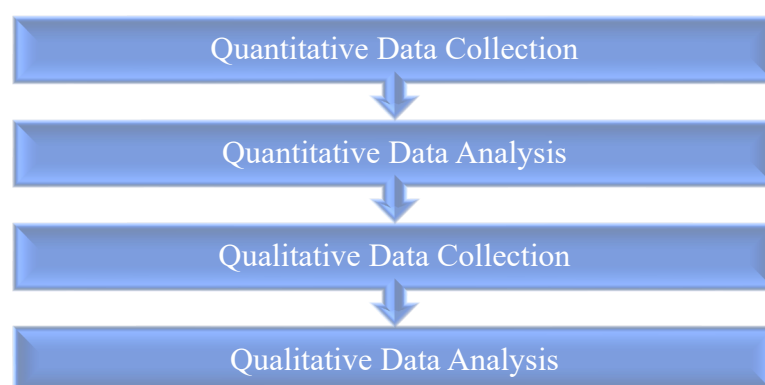


Figure 2. Demonstration of data collection and analysis

The quantitative part includes collecting, analyzing, interpreting, and reporting numerical data in a planned way (Williams, 2011). According to Dörnyei (2007), the outcomes of quantitative studies are empirical, reliable, and logical. This study used quantitative design to identify the differences between teacher, AI, and hybrid feedback types. It aims to identify the feedback given to students' written work in terms of HOCs (e.g., content,

organization, and coherence/cohesion) at the discourse level, and LOCs (e.g., vocabulary, grammar, and mechanics) at the sentence level.

For the qualitative part of the study, the researcher utilized semi-structured interviews which allowed EFL teachers to express their perspectives on the use of AI when providing feedback and facilitated a deep and complex understanding of the topic (Silverman, 2016).

Creswell (2003) highlighted the combination of qualitative and quantitative methods as a beneficial strategy for comprehending the research problem from the perspective of the researcher. Therefore, quantitative and qualitative methodologies were employed to answer the study's research questions. Mixed-method research design was developed by integrating a quantitative analysis of students' writing with semi-structured interviews. The overall research design is presented below in Table 2.

Table 2. Research design by research questions and data analysis

Quantitative Data	Qualitative Data
Research Question 1	Research Question 2
Rasch Analysis	Semi-structured interviews

3.2. Participants

The participants in the quantitative phase were tertiary-level English as a Foreign Language (EFL) students enrolled at a state university in Türkiye. Two classes, with 19 and 21 students, respectively, participated in the study, totaling 40 students. Students' ages were between 18-20. The criteria for selection depended on their proficiency level, which was decided by their university's placement exam at the beginning of the first semester of 2024-2025, and their level of English was pre-intermediate, according to the Common European Framework of Reference (CEFR). Gender, age, and previous writing exam grades of the students were not considered variables in this study. Their demographic information is provided below in Table 3, ensuring a clear overview of the participants' backgrounds for a better understanding of the study's context.

Table 3. Demographic information of the EFL students

Participants	Number of Students	Age Mean	Proficiency Level
Class 1	19	20	Pre-intermediate
Class 2	21	20	Pre-intermediate

The participants of the qualitative phase of the study were four EFL teachers who work at a state university in Türkiye with the participant students. The convenient sampling method was chosen to include participants who were readily accessible and willing to participate in the study, ensuring the feasibility of the research process. According to Dörnyei and Griffee (2010), convenient sampling is commonly used when the researcher selects participants based on ease of access and availability, especially in studies with time or resource constraints. Thus, the participant teachers who had sufficient technological knowledge, had experience using AI for their writing courses or feedback purposes, and had at least three years of experience teaching writing, were included in this study.

Participant teachers had an average of 6 years of teaching experience, ranging from 3 to 12 years. They had experience teaching L2 writing for an average of 6 years, ranging from 3 to 12 years. They were familiar with ChatGPT, with an average of 2 years as a tool for providing feedback on writing assignments. Their demographic information is provided below in Table 4.

Table 4. Demographic information of the participant teachers

Participants	Educational Background	Experience of English Language Teaching	Experience with Teaching Writing	Experience with L2 <i>ChatGPT</i>
T1	MA degree	4 years	3 years	2 years
T2	MA degree	12 years	12 years	2 years
T3	PhD student	4 years	4 years	2 years
T4	PhD student	3 years	3 years	2 years

All the participants provided written informed consent before the data collection procedures. The researcher also ensured that all participants were aware of their rights and that confidentiality measures were in place to protect their data. Moreover,

participants were allowed to ask any questions about the study before giving their consent. This level of transparency and communication helped establish trust between the researcher and participants, leading to a successful data collection process. This study's thorough selection criteria and ethical considerations laid a solid foundation for the research findings and analysis.

3.3. Data Collection Tools

3.3.1. Students' essays

In the quantitative phase of this study, the primary data collection tools were the 80 essays written by the students and the rubric used by both EFL teachers and ChatGPT, which includes feedback categories. The students were asked to write essays on two topics, which were later analyzed to identify HOCs, such as content, organization/coherence/cohesion, and LOCs, like vocabulary, grammar, and mechanics. The essays were collected in the first semester of the 2024-2025 academic year. The students were required to write two essays on “*daily life hobbies*,” totaling 40 essays, and “*daily life relations*,” totaling 40 essays with 250-300 words each, as an in-class activity. The students were given a supplementary document, which included two written tasks and instructions (see Appendix B).

The reason why students were tasked with writing two essays on closely related topics was to avoid teachers' familiarity with EFL students' writing while providing feedback. For the first essay, feedback was provided separately by ChatGPT and the teacher. For the second essay, hybrid feedback, which combined the feedback from first ChatGPT and then the teacher, was delivered. In this way, this design choice allowed us to avoid redundancy by ensuring that ChatGPT and teachers were not providing hybrid feedback on the same essay. To achieve this, two essay topics with comparable levels of difficulty and scope were created based on expert opinions, ensuring consistency and minimizing the potential for repetition during the evaluation process.

3.3.2. Semi-structured interviews

In the qualitative phase, semi-structured interviews were conducted using a set of pre-prepared questions designed to gather insights from the participating teachers regarding their perceptions of ChatGPT-generated feedback and its potential to support traditional teacher feedback. Interview questions were prepared by the researcher and advisor in line with the research questions. A draft was prepared for interview questions, expert opinion was obtained from another researcher, and a pilot study with an EFL teacher was conducted. The pilot teacher was interviewed in advance to check the interview questions' validity and reliability and modify them if needed. After the interview, possible modifications were discussed with two ELT professors, and the final version of the interview questions was created (see Appendix D). The feedback rubric of the study was also used during the interviews as a supplementary tool (see Appendix C).

3.4. Data Collection Procedure

Teacher Training: Before beginning the study, the researcher organized a training session for the teachers via Zoom in September 2024. The training lasted approximately two hours and covered several key areas to ensure the teachers were well-prepared to participate. First, an overview of the study was presented, detailing the objectives and various types of feedback (teacher, AI, and hybrid feedback). In this study, hybrid feedback is conducted as AI feedback supported by teacher feedback. First, the teacher gets feedback from AI; then, the teacher provides feedback on AI's feedback. Next, the rubric for evaluating students' writing was introduced, and the researcher demonstrated how to apply it by providing feedback on two sample essays. This was followed by a live demonstration of how ChatGPT would generate feedback, showing the step-by-step process using the same essays. The session concluded with a Q&A, during which teachers' questions were addressed to ensure they were confident in their roles. The timeline of the training session can be seen in Figure 3.

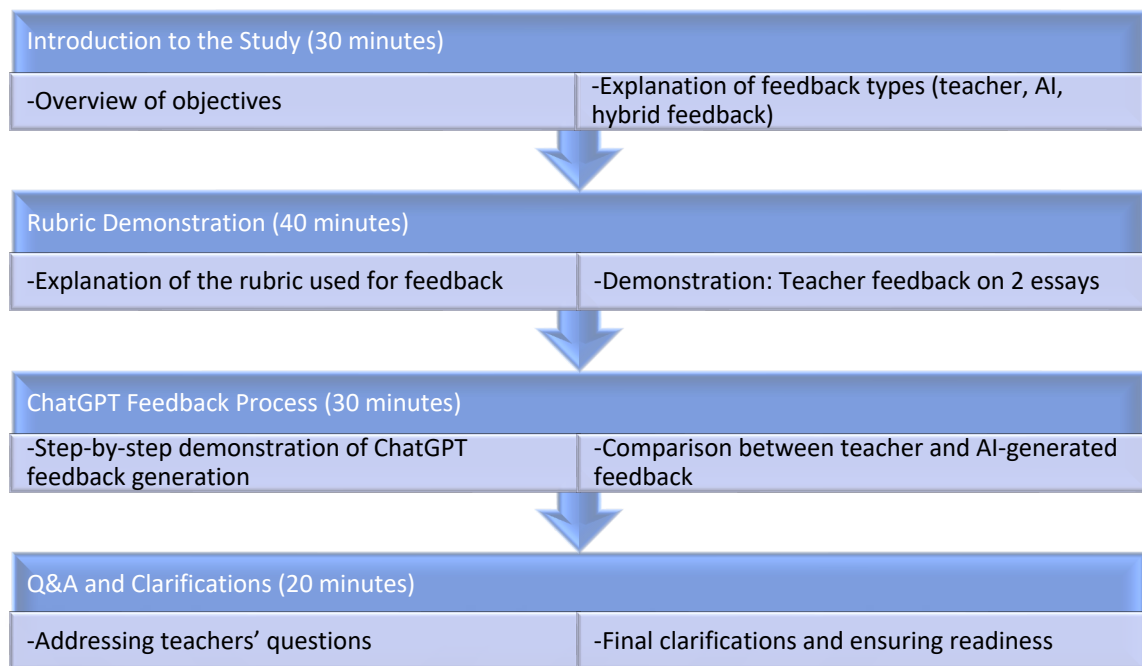


Figure 3. Training structure for participating teachers (September 2024)

After an overview of the study, a teacher feedback process was demonstrated with the participant teachers. The rubric of the study was introduced, and a demo feedback process was completed. An example of a teacher feedback process can be seen in Figure 4. If teachers have any additional feedback or comments, they were told that they can state it as a note under the essay. LOCs were shown in orange color and HOCs were shown in red color in the essay.

Jack's Daily Routine

Jack usually wakes up at 9 am. Jack loves sleeping a lot. Jack sometimes gets up at 9:30 am and then **have (grammar)** a nice breakfast. He always drinks coffee and eats donuts for breakfast. Jack sits at his desk at **10 o'clock in the afternoon (coherence/cohesion)**. He doesn't like to work too hard before noon and usually does simple things. He goes to have lunch at nearly 11:40 am. He likes to eat hamburgers and fries for lunch. By midday, he starts getting ready to go to work. He takes a clean shower and then **wears clothes (vocabulary)**. He always leaves home for work at 1 o'clock and **than (grammar)** goes **(grammar)** to work. Jack works as a ship engineer (in a shipyard). He starts work at half past one. He's very busy at work but he loves his job because he's interested in ships. He usually leaves work at half past five. Jack arrives home and prepares for dinner. He always eats dinner at half past six. He loves to eat chicken with his girlfriend for dinner. Jack usually doesn't go out much in the evenings because he is tired from work. He sometimes reads **(grammar) book** at 8 pm. He usually reads historical books. Lastly, he often goes **sleeping (grammar)** at half past ten.:

*Improve paragraph order.

*Include a topic sentence.

Figure 4. An example of teacher feedback process

When deciding on the prompt for ChatGPT, different versions were tried with a colleague of the researcher. They decided on the most effective prompt by producing different ones and acquiring different feedback types. Lastly, the researcher decided on a prompt that had some steps before getting feedback.

First, the researcher provided information about the students' background, age, and level of English. He also provided information about the university. Then, he provided the rubric used for teacher feedback and an example of an essay including teacher feedback. He wanted ChatGPT to provide feedback as an EFL teacher. By asking questions about the expected process, the researcher made sure that ChatGPT understood the details and important points. A final prompt can be seen in Figure 5.

What can I help with?

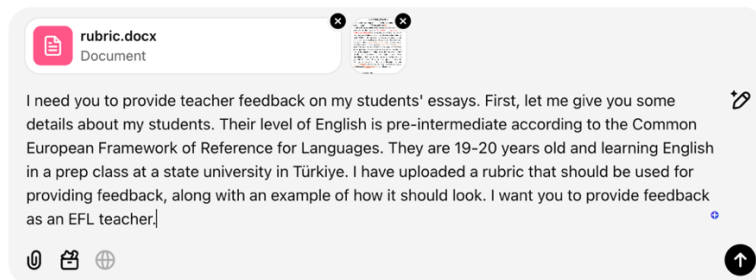


Figure 5. The final prompt used for ChatGPT feedback

Next, a demonstration of ChatGPT feedback was completed with the participant teachers on an essay. The prompt was introduced and practiced on ChatGPT. An example of the ChatGPT feedback process can be seen in Figure 6.

Jack usually wakes up at 9 am. Jack loves sleeping a lot. Jack sometimes gets up at 9:30 am and then **have** (grammar: should be "has") a nice breakfast. He always drinks coffee and eats donuts for breakfast. Jack sits at his desk at **10 o'clock in the afternoon** (coherence: "in the afternoon" is inaccurate; use "10 am"). He doesn't like to work too hard before noon and usually does simple things. He goes to have lunch at nearly 11:40 am. He likes to eat hamburgers and fries for lunch. **He always drinks coke with hamburger** (vocabulary: replace with "soda with his hamburger"). By midday, he starts getting ready to go to work. He takes a clean shower and then **wears clothes** (vocabulary: "gets dressed" sounds more natural). He always leaves home for work at 1 o'clock and **than** (grammar: should be "then") **goes work** (grammar: should be "goes to work").

Jack works as a ship engineer in a shipyard. He starts work at half past one. **He's very busy at work but he loves his job because he's interested in ships** (coherence: elaborate on why he's interested in ships for more depth). He usually leaves work at half past five. Jack arrives home and prepares for dinner. He always eats dinner at half past six. He loves to eat chicken with his girlfriend for dinner. Jack usually doesn't go out much in the evenings because he is tired from work. **He sometimes reads book** (grammar: should be "a book") at 8 pm. He usually reads historical books. **Lastly, he often goes slee** (grammar: should be "goes to sleep") at half past ten.

Figure 6. An example of ChatGPT feedback

Lastly, a demonstration of the hybrid feedback process was completed with the participant teachers. In hybrid feedback, firstly, teachers were required to get ChatGPT feedback. Then, they were required to provide teacher feedback on ChatGPT feedback. A demo hybrid feedback process was completed with participant teachers on an essay. Feedback from ChatGPT was obtained and then the researcher provided teacher feedback on ChatGPT feedback.

Teacher Feedback: In the study's first phase, each participating teacher was assigned to provide feedback on 10 of the 40 essays on the first topic according to the rubric.

AI Feedback: Meanwhile, the researcher utilized ChatGPT to generate feedback on the same essays using a predetermined prompt developed with the assistance of two ELT experts to ensure the feedback's relevance and effectiveness.

Hybrid Feedback: In the second phase, the teachers were required to obtain feedback from ChatGPT for the essays on the second topic using the predetermined prompt. Each participant teacher was required to get ChatGPT feedback for 10 essays on the second topic. Then, the teachers were asked to provide their feedback in addition to ChatGPT's feedback and send the hybrid feedback (a combination of ChatGPT and teacher feedback) to the researcher.

Qualitative data were collected through semi-structured teacher interviews, followed by quantitative data analysis. These interviews aimed to explore the teachers' perspectives on the quality of ChatGPT's feedback and its potential as a replacement or complementary tool for providing feedback on students' writing. After the quantitative part of the study, semi-structured interviews with the participant teachers were conducted. The interviews were recorded, transcribed, and analyzed using the qualitative software program MaxQDA 24. The entire data collection process can be seen below in Figure 7.



Figure 7. Data collection and feedback process timeline

3.5. Data Analysis

3.5.1. Quantitative data analysis

To address research question 1, each feedback message produced by ChatGPT and the participating teachers was initially divided into feedback units (i.e., idea units) following the procedures outlined by Cho et al. (2006) and Link et al. (2020). A feedback unit is an independent message that addresses a specific problem or targets a text feature. To ensure accuracy, two researcher colleagues participated in the feedback process. The researcher and colleagues divided all feedback messages into idea units and noted the feedback category and the number of feedback units in Excel. The researcher applied the same process to the study's second phase with two colleagues. The second group of 40 essays, which were provided with hybrid feedback, was noted in Excel as the number of feedback units and category of feedback.

The rubric of the study was adapted from Ene and Upton (2014) and Ferris (2006). Both ChatGPT and teacher feedback were provided depending on this rubric. Regarding the rubric, there are two main categories: HOCs and LOCs. Under HOCs, there are two subcategories: *content and organization/coherence/cohesion*. Under LOCs, there are three subcategories: *vocabulary, grammar, and mechanics*, as can be seen in Table 5. A comprehensive version of the rubric, including the classification of error types, is provided in Appendix C.

Table 5. The rubric used to categorize feedback

HOCs (Higher-order concerns)	LOCs (Lower-order concerns)
Discourse level	Sentence level
	-Vocabulary
-Content	
	-Grammar/Syntax/Morphology
-Organization/Coherence/Cohesion	
	-Mechanics

In the data analysis process, descriptive statistics and multifaceted Rasch analysis were performed for the socio-demographic information of the participants.

The main reason for using multifaceted Rasch analysis is that the common interaction between variables is examined rather than the variables within the scope of the research (Kassim, 2007). In addition, more detailed results are obtained because it produces both group and individual-level statistics at the same time (Myford & Wolfe, 2003). In the multifaceted Rasch measurement model, many sources of variability (such as rater, item, task, individual, and time) can be placed on a single equal interval scale (Bond & Fox, 2015; Linacre, 1993). These features provide the opportunity to examine the effects of rater biases, beliefs, or personal characteristics on scoring behavior using a multifaceted Rasch measurement model approach.

Since multifaceted Rasch analysis is a member of item response theory, it has some assumptions that need to be met (Farrokhi, et al., 2012; Farrokhi, et al., 2011). These assumptions are one-dimensionality, local independence and model-data fit. Considering that the rubric used as a measurement tool was used to assess writing skills, which is a single construct, it was accepted that the measurement tool represented a unidimensional construct. Since one-dimensionality was ensured, it was accepted that the local independence assumption was also met. Finally, standardized residual values were examined for model-data fit. It is stated that the number of standardized residual values outside the ± 2 range should not be more than 5% of the total number of observations to ensure model-data fit (Linacre, 2017). When the standardized residual values were examined, it was found that there were 37 (4.63%) in the ± 2 range for the first model and 9 (2.81%) for the second model, and it was concluded that the model data fit was good (total number of observations $80 \times 2 \times 5 = 800$ for the first model and $80 \times 2 \times 2 = 320$ for the second model). The data were analyzed utilizing SPSS 25 for statistical computations and FACETS 3.70.1 for multifaceted Rasch analysis.

3.5.2. Qualitative data analysis

To address research question 2, semi-structured interviews were conducted to reveal the perceptions of participant EFL teachers. Interviews were conducted via an online meeting platform, *Zoom*. The interviews were transcribed via an online tool, *Trint*². Thematic analysis was used to analyze the data, and it is a valuable method for collecting insights

² <https://trint.com>.

into individuals' beliefs, values, or worldviews derived from qualitative data (Creswell, 2012; McCombes, 2023). As shown in Figure 8, the qualitative software program MaxQDA 24 was utilized for thematic analysis.

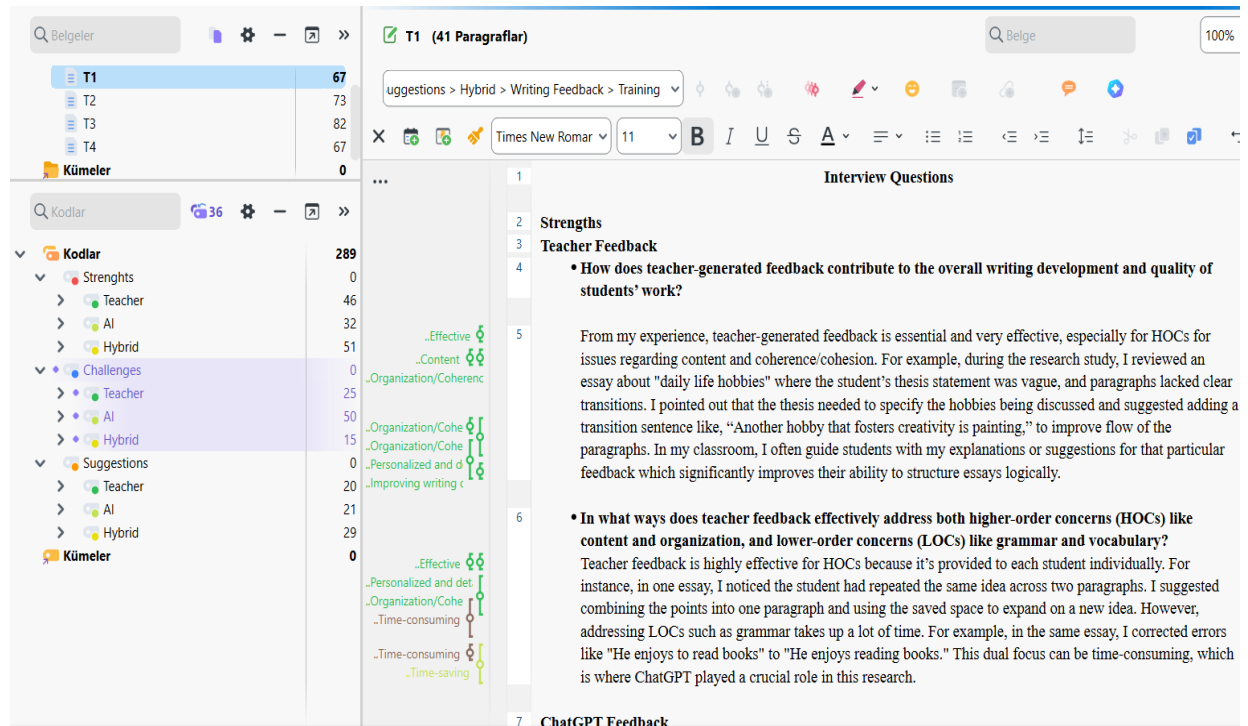


Figure 8. Data analysis in MaxQDA

The researcher generated main themes and sub-themes in accordance with research questions. Participant teachers' answers were coded under three main themes: Strengths, challenges, and suggestions in terms of teacher, AI, and hybrid feedback. Under each feedback source, three sub-themes were created: LOCs, HOCs, and additional comments. According to the answers of EFL teachers, codes were created under these sub-themes. Coding was completed through MaxQDA, and an example extract of how it was done is below:

*ChatGPT feedback is great for providing **quick, immediate responses**. I've noticed it's especially helpful with lower-order concerns, like **grammatical errors** and **spelling**. When I first use ChatGPT to provide feedback, I appreciated how it could **instantly identify and correct errors**, which allowed me see errors in seconds. It's also useful for suggesting improvements for **repetitive sentences** or **correcting word forms**. However, its impact on higher*

concerns such as *developing ideas* or ensuring *accuracy of information* is more limited (T3).

This extract was coded under the strengths of AI feedback as *quick feedback* and *grammar*. On the other hand, it was coded under the challenges of AI feedback as *organization/coherence/cohesion* and *content*.

An EFL teacher as a second rater coded in accordance with the themes and codes to test the interrater reliability of coding. The outcomes were compared using the formula (the number of agreements/the number of agreements (x) the number of disagreements multiplied by 100) of Tawney and Gast (1984). The interrater reliability level between the raters was found .86 and indicated a high degree of reliability.

4. PILOT STUDY

4.1. Research Design of the Pilot Study

The start of the pilot study aimed to evaluate the research design, data collection instruments, and methodologies to confirm their suitability and efficacy for the main research study. 10 student essays were collected and assessed, including feedback from both the participant teacher and ChatGPT. The feedback was categorized into HOCs and LOCs to compare the effectiveness and focus of the feedback provided by each source. Next, the researcher assigned another topic at the same level to the same students. After collecting all the essays, the researcher sent them to the participating teacher and asked him to obtain feedback from ChatGPT using the predetermined prompt. After completing the ChatGPT feedback process, the researcher asked the teacher to provide feedback on ChatGPT's feedback according to the rubric adapted from Ene and Upton (2014) and Ferris (2006). Then, the participant teacher sent the researcher both AI-only feedback and hybrid feedback (combining ChatGPT and teacher feedback). Next, the researcher conducted descriptive, frequency, and t-test analyses of the AI feedback and hybrid feedback in terms of HOCs and LOCs using the quantitative data analysis program SPSS 2024.

For the qualitative part, a semi-structured interview was conducted on Zoom with the participant teacher. Interview questions were tested, and future modifications were discussed according to this interview. Then, the final version of the interview questions was decided with the help of an ELT expert.

4.2. Participants of the Pilot Study

The quantitative part of the pilot study included one experienced EFL teacher and ten university-level students. Every student's essay received a traditional written evaluation from the teacher. In the meantime, the researcher got feedback from ChatGPT using predetermined prompts to generate feedback.

The qualitative part of the pilot study included 1 participating EFL teacher to be able to test the semi-structured interview questions.

4.3. Data Collection Tools and Procedure of the Pilot Study

The researcher gathered ten essays from one of his classes and collaborated with an ELT expert during the first stage of the pilot study. The participant teacher provided traditional teacher feedback on the essays according to the rubric adapted for the study. In the meantime, the researcher got feedback from ChatGPT for the same essays. The teacher's and ChatGPT feedback was documented independently and analyzed through statistical methods. The feedback was categorized and analyzed using the rubric adapted from existing literature, focusing on HOCs (e.g., content, organization, coherence/cohesion) and LOCs (e.g., vocabulary, grammar, mechanics). The results were compared to identify any discrepancies or alignments between the two sources of feedback. The purpose was to evaluate whether ChatGPT could effectively complement the teacher's feedback.

For the qualitative part, a semi-structured interview was conducted via Zoom. Interview questions were asked, and the answers were recorded for future analysis.

4.5. Data Analysis of the Pilot Study

The researcher used the quantitative data analysis program SPSS 2024 to conduct descriptive, frequency, and t-test analyses of the teacher and AI feedback on 10 essays. For the qualitative part, the data of semi-structured interview was analyzed through content analysis by the researcher.

After the pilot study, the researcher gathered 80 essays for the main study. However, the researcher and his advisor had a meeting before running a statistical analysis. They noticed that participant teachers hesitated to give sufficient feedback on essays and had problems using ChatGPT while providing feedback on students' essays. Thus, the outcome was not enough to run a statistical analysis.

As a result, the researcher planned to hold a training session with participant teachers to provide them with a demo of teacher feedback using a predetermined rubric and a preview of using ChatGPT while giving feedback on students' essays. After the training session, the researcher collected new data for the main study and completed it with the participant teachers.

4.6 Findings of the Pilot Study

In the first part of the pilot study, evaluating essays from 10 students involved a comparison using feedback from both a teacher and ChatGPT for each student. The feedback was categorized according to specific rubric categories, and the numerical data from the teacher's and ChatGPT's assessments were recorded for each category. Relevant categories included content, organization, coherence/cohesion, vocabulary, grammar, and mechanics. This data allowed us to compare the teacher's and ChatGPT's feedback strategies in different areas when assessing each student's writing.

Table 6. Comparison of teacher and ChatGPT feedback

Feedback categories		Teacher		ChatGPT		Mann-Whitney U	Z	p
		Median	Mean	Median	Mean			
Higher-order concerns (HOCs)	Content	1.00	1.00	0.50	0.8	45.00	-0.403	0.687
	Organization	3.00	2.60	0.50	0.5	2.5	-3.702	0.00
Lower-level concerns (LOCs)	Vocabulary	1.50	2.00	1.00	1	38.00	-0.947	0.344
	Grammar	7.50	10.00	10.50	10.6	37	-0.987	0.324
	Mechanics	8.50	10.10	5.50	5.5	25.5	-1.867	0.062

* $p < 0.05$

Table 6 shows a comparison of the feedback provided by the teacher and ChatGPT. The feedback was analyzed in two main categories: higher-order and lower-level concerns. For higher-order concerns, the teacher's median and mean values for content and organization were 1.00 and 2.60, respectively, while ChatGPT's values for these issues were 0.50 and 0.8, respectively (Mann-Whitney U = 45.00, Z = -0.403, p = 0.687). For lower-level concerns, on topics such as vocabulary and grammar, the teacher's median and mean values were 1.50 and 2.00, respectively, whereas ChatGPT's values were 1.00

and 1.00, respectively (Mann-Whitney U = 38.00, Z = -0.947, p = 0.344). In addition, while the median and mean values of the teacher on mechanics were 8.50 and 10.10, respectively, the ChatGPT values were 5.50 and 5.50, respectively (Mann-Whitney U = 25.5, Z = -1.867, p = 0.062).

In the second phase of the pilot study, the ten newly collected student essays were presented to the participating teacher, who was asked to get feedback for each student using ChatGPT. The teacher was then asked to provide his feedback on the feedback given by ChatGPT and to make relevant corrections or additions. In this process, for each feedback category (content, organization, vocabulary, grammar, and mechanics), the amount of feedback provided by the teacher and the amount of feedback provided by ChatGPT were recorded in Excel.

Table 7. ChatGPT versus hybrid feedback

Feedback categories		Hybrid		ChatGPT		Mann-Whitney U	Z	p
		Median	Mean	Median	Mean			
Higher-order concerns (HOCs)	Content	1.00	0.70	0.00	0.1	24.50	-2.304	0.021
	Organization	1.00	0.90	0.00	0.2	28.00	-1.909	0.056
Lower-level concerns (LOCs)	Vocabulary	0.00	0.60	0.00	0.6	50.00	0.00	1.00
	Grammar	5.50	6.80	5.50	6.9	49.50	-0.038	0.97
	Mechanics	5.00	8.30	5.00	8.9	46.50	-0.272	0.786

* $p < 0.05$

Table 7 shows the comparison between hybrid feedback and ChatGPT feedback for HOCs and LOCs. For HOCs, hybrid feedback demonstrated higher scores for content, with a median of 1.00 and a mean of 0.70, compared to ChatGPT's median of 0.00 and mean of 0.10. This difference was statistically significant (Mann-Whitney U = 24.50, Z = -2.304, p = 0.021). Similarly, for organization, hybrid feedback scored a median of 1.00 and a mean of 0.90, while ChatGPT scored 0.00 and 0.20, respectively. However, this difference was not statistically significant (Mann-Whitney U = 28.00, Z = -1.909, p = 0.056).

For LOCs, both feedback sources showed comparable results. In vocabulary, the median and mean values for both hybrid feedback and ChatGPT were 0.00 and 0.60, respectively, with no significant difference (Mann-Whitney U = 50.00, Z = 0.00, p = 1.00). Regarding

grammar, both scored a median of 5.50, with hybrid feedback having a mean of 6.80 and ChatGPT a mean of 6.90, indicating no statistically significant difference (Mann-Whitney $U = 49.50$, $Z = -0.038$, $p = 0.97$). Lastly, for mechanics, hybrid feedback had a median of 5.00 and a mean of 8.30, while ChatGPT had a median of 5.00 and a mean of 8.90, again showing no statistically significant difference (Mann-Whitney $U = 46.50$, $Z = -0.272$, $p = 0.786$).

This analysis highlights that while hybrid feedback significantly outperformed ChatGPT in addressing content, both feedback sources demonstrated similar performance in LOCs such as vocabulary, grammar, and mechanics.



5. RESULTS

This section provides the results of the analysis of numerical data, including Rasch analysis related to teacher, AI, and hybrid feedback, and thematic analysis of semi-structured interview questions regarding EFL teachers' perspectives. In the quantitative part, statistical analysis of teacher, AI, and hybrid feedback is presented by tables and figures. In the qualitative part, findings included themes, patterns, and direct citations from participants, offering a deeper comprehension of participants' perspectives and experiences concerning the research topic. The results are presented as quantitative and qualitative analyses, respectively, in line with research questions. This study's results offer insights regarding the interaction among teacher, AI, and hybrid feedback types in accordance with the feedback categories and the perspectives of EFL teachers regarding the use of AI in their writing instruction and their experience with AI.

5.1. Quantitative Research Results

5.1.1. To what extent do different feedback types address higher-order concerns and lower-order concerns in EFL students' writing?

In the quantitative part of the study, teacher, AI, and hybrid feedback were analyzed in terms of HOCs and LOCs. Each participant teacher was given ten essays for the first part which is ChatGPT and teacher feedback in addressing higher-order and lower-order concerns. Next, another 40 essays were given to EFL teachers and asked to get feedback from ChatGPT. Then, they were asked to provide teacher feedback on ChatGPT feedback to be able to analyze the essays in terms of hybrid feedback (combining ChatGPT and teacher feedback).

5.1.1.1. Do the essays differ in main feedback categories (HOCs and LOCs) across different feedback types (teacher, ChatGPT, and hybrid)?

Within the scope of the research, firstly, the measurement reports of each facet for the first measurement model were analyzed. The logit values, standard error of the logit value (SE), infit and outfit values and group statistics obtained for each facet are given respectively. First, the measurement report for the Essay facets is given in Table 8.

Table 8. Measurement report of the essay facets

Essay ID	Observed Mean	Adjusted Mean	Logit	SE	In-Fit	Out-Fit
E01	1.85	1.85	1.54	0.40	1.55	1.23
E02	1.85	1.85	1.54	0.40	0.94	0.96
E03	1.85	1.85	1.54	0.40	1.09	1.00
E04	2.05	2.10	0.93	0.38	0.88	0.83
E05	1.75	1.73	1.87	0.42	1.97	1.47
E06	1.70	1.66	2.05	0.43	0.76	0.71
E07	1.85	1.85	1.54	0.40	0.91	0.86
E08	2.10	2.16	0.78	0.38	1.51	1.37
E09	1.90	1.92	1.38	0.40	0.79	0.78
E10	1.75	1.73	1.87	0.42	1.54	1.43
E11	1.40	1.29	3.29	0.51	0.99	1.49
E12	1.50	1.41	2.80	0.48	0.97	1.01
E13	1.85	1.88	1.47	0.40	0.87	0.72
E14	1.85	1.88	1.47	0.40	0.59	0.54
E15	1.70	1.68	1.99	0.43	1.22	0.86
E16	1.70	1.68	1.99	0.43	0.76	0.57
E17	1.35	1.23	3.56	0.53	1.45	3.52
E18	1.90	1.94	1.31	0.40	1.00	1.04
E19	2.10	2.19	0.72	0.38	0.90	0.88
E20	1.70	1.68	1.99	0.43	1.53	1.45
E21	2.00	2.00	1.17	0.39	0.79	0.76
E22	1.85	1.81	1.64	0.40	1.19	1.12
E23	1.55	1.43	2.75	0.46	1.13	1.01
E24	1.55	1.43	2.75	0.46	1.15	1.44
E25	1.90	1.88	1.48	0.40	0.79	0.72
E26	1.60	1.49	2.54	0.45	2.62	1.72
E27	1.55	1.43	2.75	0.46	0.66	0.89
E28	2.30	2.35	0.33	0.37	0.93	0.91
E29	2.15	2.17	0.74	0.37	1.28	1.25
E30	1.85	1.82	1.63	0.40	0.94	0.91
E31	2.00	1.97	1.24	0.39	0.45	0.47
E32	2.10	2.09	0.95	0.38	0.68	0.68
E33	2.15	2.15	0.81	0.38	1.10	1.02
E34	1.65	1.53	2.42	0.44	0.77	0.69
E35	1.80	1.72	1.88	0.41	0.62	0.59
E36	2.05	2.03	1.10	0.38	0.68	0.70
E37	1.80	1.72	1.88	0.41	0.67	0.57
E38	1.85	1.79	1.71	0.40	0.91	0.83
E39	2.25	2.26	0.53	0.37	0.56	0.57
E40	1.95	1.91	1.40	0.39	0.87	0.76
Mean	1.84	1.81	1.68	0.41	1.03	1.01
S (Population)	0.22	0.27	0.74	0.04	0.41	0.51
S (Sample)	0.22	0.27	0.74	0.04	0.41	0.51
Model. (Population): RMSE = 0.42 Adjusted S = 0.61 Separation = 1.46 Strata = 2.28 Reliability = 0.68						
Model. (Sample): RMSE = 0.42 Adjusted S = 0.62 Separation = 1.49 Strata = 2.32 Reliability = 0.69						
Model. Chi-Square (Fixed): 114.10 sd = 39 p = .00						
Model. Chi-Square (Normal): 30.00 sd = 38 p = .82						

When Table 8 is analyzed, it is seen that the observed mean and the adjusted mean values are quite close to each other, which indicates that the values estimated by the model do not have a large deviation. The mean of the logit values was 1.68 and the standard deviation was 0.74. This shows that the number of feedback given for the essays is at a moderate level. When the logit values were analyzed, it was determined that the essay

with the least feedback was 17 and 28. Four essays (5, 17, 26 and 31) were found to be outliers compared to the other essays and deviated from the model. This discrepancy may result from differences in how the raters assessed the essays or unique characteristics inherent in the essays themselves. Based on the analysis, it can be concluded that the feedback provided for the essays is generally consistent, as indicated by the close observed and adjusted mean values and the moderate logit mean. However, the presence of outlier essays (5, 17, 26, and 31) suggests variability in rater evaluations or inherent differences in the essays themselves.

It is seen that the separation, strata, and reliability calculated for the sample are low. The low discrimination rate and discrimination index mean that the feedback given for the evaluated essays is similar to each other (homogeneous). It was determined that the reliability value calculated for the discrimination index was .68. The main reason for this was that the number of feedback given to the essays was similar. Whether the feedback given to the essays differed statistically was tested with chi-square statistics and it was observed that the fixed chi-square value was statistically significant ($\chi^2(38) = 114.10, p < .05$). This finding indicates that there is a statistically significant difference in the number of feedback given to the essays across different feedback types.

5.1.1.2. Do the raters differ in feedback categories and patterns across different feedback types (teacher, ChatGPT, and hybrid) based on single raters and combined rater types?

After the measurement report for the essay facets is presented, the measurement report for the rater facet is given. The measurement report estimated for the rater facet is given in Table 9.

Table 9. Measurement report for the rater facets

Rater	Observed Mean	Adjusted Mean	Logit	SE	In-Fit	Out-Fit
Teacher 1	1.90	1.74	-0.15	0.25	1.26	1.17
Teacher 2	1.68	1.64	-0.42	0.27	1.58	2.57
Teacher 3	1.93	1.87	0.19	0.27	1.14	1.05
Teacher 4	2.04	1.86	0.16	0.23	0.83	0.88
ChatGPT	1.72	1.66	-0.36	0.10	0.94	0.75
ChatGPT+Teacher 1	2.04	1.82	0.05	0.24	1.03	1.06

Table 9. (continued)

ChatGPT+Teacher 2	1.88	1.82	0.05	0.26	0.90	1.19
ChatGPT+Teacher 3	1.98	1.84	0.10	0.25	1.32	1.54
ChatGPT+Teacher 4	2.20	1.95	0.38	0.24	0.66	0.71
Mean	1.93	1.80	0.00	0.23	1.07	1.21
S (Population)	0.15	0.09	0.25	0.05	0.27	0.54
S (Sample)	0.16	0.10	0.26	0.05	0.28	0.57
Model. (Population): RMSE = 0.24 Adjusted S = 0.06 Separation = 0.26 Strata = 0.68 Reliability = 0.06						
Model. (Sample): RMSE = 0.24 Adjusted S = 0.11 Separation = 0.45 Strata = 0.93 Reliability = 0.17						
Model. Chi-Square (Fixed): 16.40 sd = 8 p = .04						
Model. Chi-Square (Normal): 4.90 sd = 7 p = .67						

When Table 9 is examined, it is seen that teachers have higher averages than ChatGPT according to the observed and adjusted averages of the raters who evaluated the essays. It was found that there was a statistically significant difference in the raters' level of giving feedback to the essays ($\chi^2(8) = 16.40$, $p < .05$). When the logit values of the raters were analyzed, the most feedback was given by Teacher 4 and ChatGPT+Teacher 1, while the least feedback was given by Teacher 2.

It is seen that the separation, strata and reliability calculated for the sample are quite low. Although separation and strata were low, raters' level of giving feedback was found to be statistically significant. It was determined that the reliability value calculated for the discrimination index was .06, and the main reason for this situation was that the number of feedback given to the essays was similar. As a result, although there is a significant difference in the number of feedback categories, we can reliably say that their feedback patterns are similar.

After the measurement report for each rater was given, the raters were divided into three feedback types as teacher, AI and hybrid and included in the analysis as a new variable. The measurement report for rater type is given in Table 10.

Table 10. Measurement report on the facet of the feedback types

Feedback Type	Observed Mean	Adjusted Mean	Logit	SE	In-Fit	Out-Fit
ChatGPT	1.72	1.70	-0.25	0.10	0.94	0.75
Teacher	1.89	1.80	0.01	0.13	1.17	1.41
Hybrid	2.02	1.89	0.23	0.12	0.97	1.13
Mean	1.88	1.80	0.00	0.11	1.03	1.10

Table 10. (continued)

S (Population)	0.12	0.07	0.20	0.01	0.10	0.27
S (Sample)	0.15	0.09	0.24	0.02	0.13	0.33
Model. (Population): RMSE = 0.12 Adjusted S = 0.16 Separation = 1.37 Strata = 2.15 Reliability = 0.65						
Model. (Sample): RMSE = 0.12 Adjusted S = 0.21 Separation = 1.82 Strata = 2.76 Reliability = 0.77						
Model. Chi-Square (Fixed): 9.80	sd = 2	$p = .01$				
Model. Chi-Square (Normal): 1.70	sd = 1	$p = .20$				

When Table 10 is examined, it is seen that hybrid is the one that gives the most feedback in essay evaluation and ChatGPT is the one that gives the least. It was determined that there was a statistically significant difference in the status of giving feedback to essays ($\chi^2(2) = 9.80, p < .05$).

The separation, strata, and reliability are low; however, it is statistically significant. It was found that the number of feedback given by the raters to the essays had a statistically significant difference. However, the separation and strata were low. It was determined that the reliability value calculated for the separation was .77, and the main reason for this situation was that the number of feedback given to the essays was similar. This value was found to be quite high according to the rater facet. This finding indicates that the feedback given in essay evaluation is better separated from each other according to the rater type. As a result, although the number of feedback items provided by raters is similar overall, AI feedback provided the lowest mean feedback, while hybrid feedback offered the highest mean feedback.

5.1.1.3. Are the feedback categories in the rubric distinctive across subcategorized benchmark facets, rater facets, rater type facets, and the main categorized benchmark facets?

The measurement report on the criteria of the rubric used in essay evaluation is given in Table 11.

Table 11. Measurement report for the subcategorized benchmark facets

Feedback Categories	Observed Mean	Adjusted Mean	Logit	SE	In-Fit	Out-Fit
Content	1.23	1.23	-1.90	0.19	0.82	0.96

Table 11. (continued)

Organization	1.42	1.44	-1.04	0.15	0.92	0.93
Vocabulary	1.33	1.33	-1.43	0.17	0.85	0.94
Grammar	2.83	2.93	2.70	0.12	1.00	0.99
Mechanics	2.39	2.48	1.66	0.12	1.23	1.22
Mean	1.84	1.88	0.00	0.15	0.97	1.01
S (Population)	0.65	0.69	1.83	0.03	0.15	0.11
S (Sample)	0.72	0.77	2.05	0.03	0.16	0.12
Model. (Population): RMSE = 0.15 Adjusted S = 1.82 Separation = 11.90						
Strata = 16.20 Reliability = 0.99						
Model. (Sample): RMSE = 0.15 Adjusted S = 2.04 Separation = 13.31						
Strata = 18.093 Reliability = 0.99						
Model. Chi-Square (Fixed): 801.50 sd = 4 p = .00						
Model. Chi-Square (Normal): 4.00 sd = 3 p = .26						

When Table 11 is analyzed, it is seen that grammar is the feedback category that is given the most feedback and content is the category that is given the least. It was found that there was a statistically significant difference in the feedback given to the category ($\chi^2(4) = 801.50$, $p < .05$). Accordingly, it was determined that the generally provided grammar feedback and content was the category that they provided the least feedback. The categories of the rubric can effectively distinguish different feedback conditions with high reliability, as demonstrated through the analysis of subcategories.

It is seen that the separation, strata, and reliability are high. The high separation and strata indicate that the feedback given to the categories in the rubric is quite different. The reliability value calculated for the discrimination index is .99, which indicates that the feedback given to the criteria in the rubric is highly reliable. The variable map for the tested multifaceted Rasch measurement model is given in Figure 9.

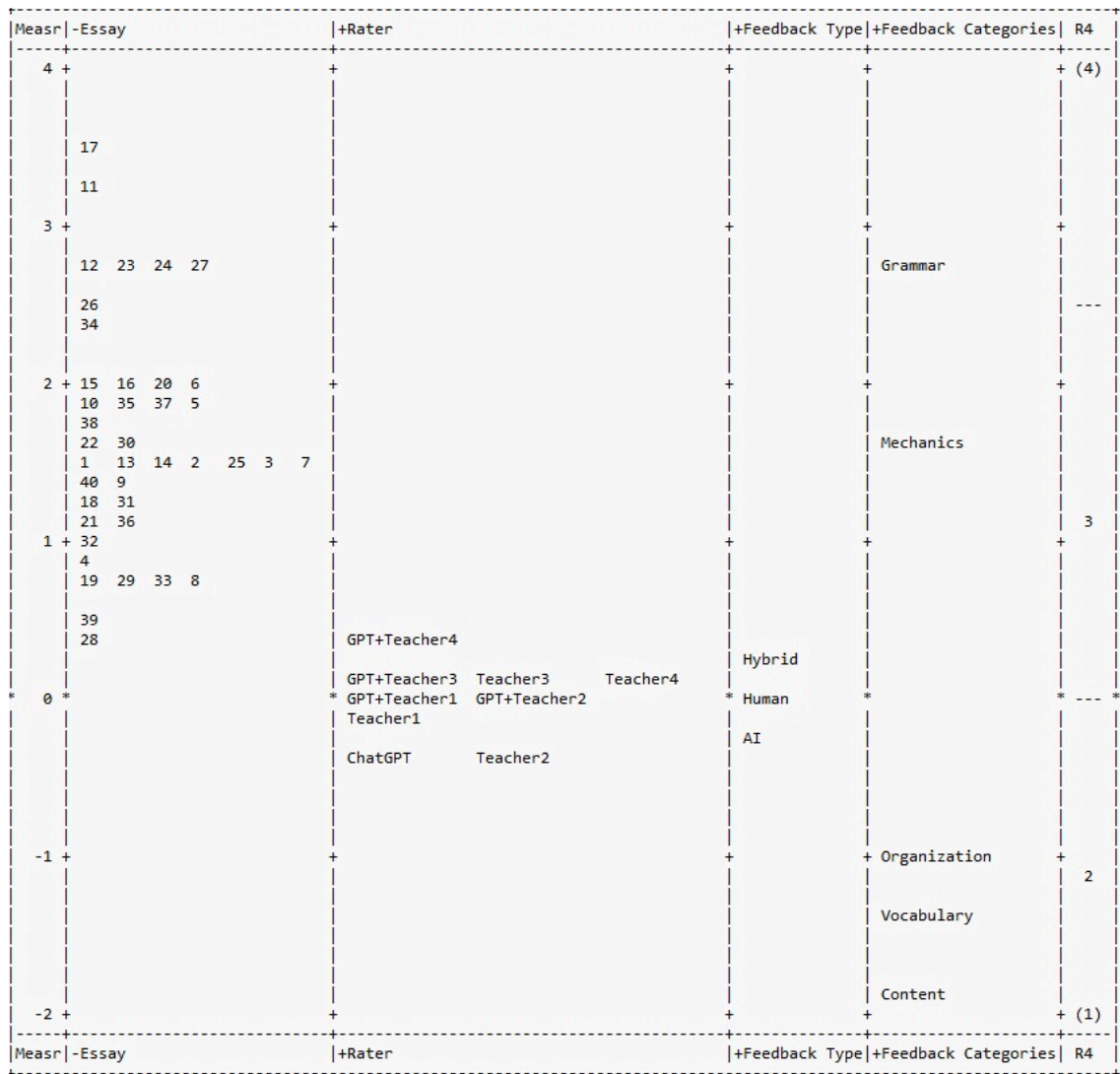


Figure 9. Variable map for the tested model

When Figure 9 is analyzed, it is seen that essay ID 17 received the least feedback, while essays ID 28 and 39 received the most feedback. It is seen that the rater giving the most feedback is ChatGPT+Teacher 4 and the feedback type is hybrid, while the least feedback is given by the raters, ChatGPT and Teacher 2 and the feedback type is AI. Finally, the category with the least feedback was content and the most feedback was grammar.

After the first model was tested, the categories in the rubric were divided into two as higher-order concerns and lower-order concerns and a second model was tested. Regarding the second tested model, firstly, the measurement report for the rater is given in Table 12.

Table 12. Measurement report for the rater facets

Rater	Observed Mean	Adjusted Mean	Logit	SE	In-Fit	Out-Fit
Teacher 1	3.40	4.00	1.34	0.73	0.26	0.12
Teacher 2	3.40	3.99	0.87	0.68	1.33	9.00
Teacher 3	3.56	3.98	0.09	0.78	1.36	0.62
Teacher 4	3.45	3.97	-0.60	0.68	0.79	0.33
ChatGPT	2.62	3.44	-4.20	0.31	0.89	0.74
GPT+Teacher 1	3.65	4.00	3.36	0.72	0.30	0.14
GPT+Teacher 2	3.50	3.99	1.11	0.70	0.74	0.32
GPT+Teacher 3	3.55	3.96	-0.79	0.73	0.90	0.43
GPT+Teacher 4	3.45	3.94	-1.17	0.72	0.88	0.41
Mean	3.40	3.92	0.00	0.67	0.83	1.34
S (Population)	0.29	0.17	1.97	0.13	0.36	2.71
S (Sample)	0.30	0.18	2.09	0.14	0.38	2.88
Model. (Population): RMSE = 0.68 Adjusted S = 1.85 Separation = 2.71 Strata = 3.94 Reliability = 0.88						
Model. (Sample): RMSE = 0.68 Adjusted S = 1.98 Separation = 2.89 Strata = 4.19 Reliability = 0.89						
Model. Chi-Square (Fixed): 171.00 sd = 8 p = .00						
Model. Chi-Square (Normal): 7.90 sd = 7 p = .34						

When Table 12 is examined, it is seen that teachers provided higher numbers of feedback than ChatGPT according to the observed and adjusted averages of the raters who evaluated the essays. It was found that there was a statistically significant difference in the frequency levels of feedback categories given by the raters ($\chi^2(8) = 171.00$, $p < .05$). When the logit values of the raters were examined, the most feedback was given by Teacher 3 and ChatGPT + Teacher 1, while ChatGPT gave the least feedback.

It is seen that the separation, strata, and reliability are high. The high separation and strata indicate that there is a statistically significant difference in the number of raters giving feedback to essays. The reliability value calculated for the discrimination index was .88 and it was determined that the differentiation of the raters' giving feedback to the essays had high reliability.

After the measurement report for each rater was given, the raters were divided into three types as teacher, AI and hybrid and included in the analysis as a new variable. The measurement report for the feedback type is given in Table 13.

Table 13. Measurement report for feedback type facets

Feedback Type	Observed Mean	Adjusted Mean	Logit	SE	In-Fit	Out-Fit
ChatGPT	2.62	3.61	-3.51	0.36	0.89	0.74
Teacher	3.45	4.00	1.42	0.36	0.93	9.00
Hybrid	3.54	4.00	2.10	0.31	0.70	0.32
Mean	3.20	3.87	0.00	0.34	0.84	3.36
S (Population)	0.41	0.18	2.50	0.02	0.10	3.99
S (Sample)	0.51	0.22	3.06	0.03	0.12	4.89
Model. (Population): RMSE = 0.34 Adjusted S = 2.47 Separation = 7.25 Strata = 10.00 Reliability = 0.98						
Model. (Sample): RMSE = 0.34 Adjusted S = 3.04 Separation = 8.90 Strata = 12.20 Reliability = 0.99						
Model. Chi-Square (Fixed): 177.10 sd = 2 p = .00						
Model. Chi-Square (Normal): 2.00 sd = 1 p = .16						

When Table 13 is examined, it is seen that hybrid is the one who gives the most feedback in essay evaluation and AI is the one who gives the least. It was determined that there was a statistically significant difference in the status of giving feedback to essays ($\chi^2(2) = 177.10, p < .05$).

It is seen that the separation, strata, and reliability are high. The high separation and strata indicate that there is a statistically significant difference in the number of feedback types giving feedback to the essays. The reliability value calculated for the discrimination index was .98, and it was determined that the differentiation of the feedback status of the essays according to the feedback type had high reliability. Based on the overall rater types, the rubric demonstrates reliable discriminating features.

The measurement report on the categories of the rubric used in essay evaluation is given in Table 14.

Table 14. Measurement report for the main categorized benchmark facets

Feedback Category	Observed Mean	Adjusted Mean	Logit	SE	In-Fit	Out-Fit
Lower	3.98	4.00	-8.85	0.62	1.20	9.00
Higher	2.13	2.57	8.85	0.21	0.81	0.80
Mean	3.06	3.29	0.00	0.41	1.00	4.90
S (Population)	0.92	0.71	8.85	0.21	0.20	4.10
S (Sample)	2.32	1.01	12.52	0.29	0.28	5.80
Model. (Population): RMSE = 0.46 Adjusted S = 8.84 Separation = 19.21 Strata = 25.94 Reliability = 1.00						

Table 14. (continued)

Model. (Sample): RMSE = 0.46	Adjusted S = 12.51	Separation = 27.18
	Strata = 36.58	Reliability = 1.00
Model. Chi-Square (Fixed): 739.90	sd = 1	p = .00

When Table 14 is examined, it is seen that feedback is given the most for lower and the least for higher. It was found that there was a statistically significant difference in the feedback given for the main -feedback categories (HOCs and LOCs) ($\chi^2(1) = 739.90$, $p < .05$). Accordingly, it was determined that the feedback types generally provided lower-order feedback category.

It is seen that the separation, strata, and reliability are high. The high separation and strata indicate that the feedback given according to the feedback categories in the rubric is quite different. The reliability value calculated for the strata was 1.00, which indicates that the difference in the feedback given to the categories in the rubric indicates high reliability. The variable map for the tested multi-facet Rasch measurement model is given in Figure 10. In this context, the rubric also demonstrates strong discriminative capacity when evaluated in terms of the main feedback categories.

Measr	+Essay	+Rater	+Feedback Type	+Feedback Categories	R4
15 +	28 29	+	+	+	(4)
14 +	18	+	+	+	
13 +	39	+	+	+	
12 +	22	+	+	+	
11 +	21 30	+	+	+	
10 +	19 20	+	+	+	
9 +	25 32 35 36 40	+	+	+	
8 +	27 31 37 4 6 9	+	+	Lower	
7 +	12	+	+	+	
6 +	33 34 38	+	+	+	
5 +	1 10 2 23 24 3 7	+	+	+	
4 +	11 13 14 15 17	+	+	+	
3 +	26	+	+	+	
2 +	5 8	+	+	+	---
1 +	16	+	+	+	
0 +		GPT+Teacher1	+	+	
-1 +		+	+	+	
-2 +		Teacher1	Hybrid	+	3
-3 +		GPT+Teacher2 Teacher2	Human	+	
-4 +		Teacher3	+	+	
-5 +		Teacher4	+	+	
-6 +		GPT+Teacher3 GPT+Teacher4	+	+	
-7 +		+	+	+	
-8 +		+	+	+	2
-9 +		ChatGPT	AI	+	
		+	+	+	
		+	+	+	---
		+	+	+	
		+	+	+	
		+	+	Higher	(1)
Measr	+Essay	+Rater	+Feedback Type	+Feedback Categories	R4

Figure 10. Variable map for the tested model (second model)

When Figure 10 is analyzed, it is seen that while essay ID 16 received the least feedback, essay ID 28 and 39 received the most feedback. It is seen that the rater who provided the most feedback was ChatGPT + Teacher 1 and the feedback type was hybrid, while the least feedback was given by ChatGPT, and the feedback type was AI.

5.1.1.4. How does the interaction between feedback types and given feedback frequency vary across main-feedback categories (HOCs and LOCs) and sub-feedback categories (content, organization, vocabulary, grammar, and mechanics)?

The interactions between feedback types and feedback categories in the rubric were examined and the findings are given in Tables 15 and 16.

Table 15. Measurement report on the interaction between feedback types and sub-feedback categories

Observed Value	Expected Value	Bias Magnitude	Model SE	t Value	sd	p	In-Fit	Out-Fit	Feedback type	Categories
54.00	50.32	0.43	0.32	1.32	39	0.194	0.90	1.30	Teacher	Content
80.00	92.86	-3.77	1.68	-2.25	79	0.027*	0.00	0.00	AI	Content
63.00	53.83	0.82	0.28	2.93	39	0.006*	0.80	0.90	Hybrid	Content
70.00	58.46	0.90	0.26	3.39	39	0.002*	1.00	1.00	Teacher	Organization
82.00	105.07	-2.88	0.72	-4.01	79	0.000*	0.90	0.40	AI	Organization
75.00	63.48	0.82	0.26	3.18	39	0.003*	0.60	0.70	Hybrid	Organization
55.00	54.42	0.06	0.32	0.19	39	0.852	0.90	1.20	Teacher	Vocabulary
97.00	98.81	-0.13	0.27	-0.47	79	0.637	0.80	0.80	AI	Vocabulary
60.00	58.78	0.10	0.29	0.36	39	0.721	0.90	1.00	Hybrid	Vocabulary
109.00	116.16	-0.43	0.24	-1.76	39	0.086	1.10	1.10	Teacher	Grammar
228.00	213.77	0.42	0.17	2.44	79	0.017*	1.00	0.90	AI	Grammar
116.00	123.07	-0.44	0.25	-1.78	39	0.084	0.60	0.60	Hybrid	Grammar
90.00	98.65	-0.52	0.25	-2.10	39	0.042*	1.00	1.00	Teacher	Mechanics
202.00	178.51	0.70	0.17	4.06	79	0.000*	1.10	1.10	AI	Mechanics
91.00	105.85	-0.88	0.25	-3.59	39	0.001*	0.80	0.80	Hybrid	Mechanics

Fixed (all = 0) chi-square: 99.50 df: 15 p = .000

When Table 15 is analyzed, it is seen that the interaction between feedback type and feedback categories has a statistically significant difference ($\chi^2(15) = 99.50$, $p < .05$). Accordingly, it was found that the feedback types provided different numbers of feedback to the categories while evaluating the student essays. Individual-level statistics were analyzed to determine which feedback type caused the difference at the group level. Considering that each feedback type had five interactions in total, it was found that four of the five interactions of AI were significant, while the teacher had two and the hybrid had 3. When the statistically significant interactions of AI were examined, it was found that AI provided less feedback than it should have provided in content, while it provided more feedback than it should have provided in organization, grammar, and mechanics. When teachers were examined, it was found that they provided more feedback for organization and less feedback for mechanics. It was found that teachers and hybrid feedback provided similar scores, while AI provided different scores.

In conclusion, AI showed limited effectiveness in providing feedback on content and organization while demonstrating significant success in grammar and mechanics. Teachers delivered more feedback on organization; however, their feedback on mechanics was notably insufficient. Hybrid feedback, on the other hand, significantly provided a higher amount of feedback across all domains, including content, organization, grammar, and mechanics.

Table 16. Measurement report on the interaction between rater types and main categories

Observed Value	Expected Value	Bias Magnitude	Model SE	t Value	sd	p	In-Fit	Out-Fit	Feedback type	Categories
117	116	0.13	0.36	0.36	39	0.724	0.80	0.70	Teacher	Higher
123	123	0.00	0.36	0.00	39	0.999	0.70	0.60	Hybrid	Higher
101	102	-0.13	0.36	-0.35	79	0.724	0.90	1.00	AI	Higher
159	160	-9.37	1.04	-8.99	39	0.000*	1.10	0.50	Teacher	Lower
318	317	0.45	0.74	0.61	79	0.543	1.10	0.90	AI	Lower
160	160	0.00	128.00	0.00	39	1.000	0.00	0.00	Hybrid	Lower

Fixed (all = 0) chi-square: 81.40 df: 6 p = .000

When Table 16 is examined, the interaction between the feedback type and the feedback categories of the rubric is statistically significant * ($\chi^2(6) = 81.40$, $p < .05$). When the number of higher and lower feedback categories by the feedback types is examined, it is found that the numbers of teacher and hybrid feedback are close to each other; however, the number of higher and lower feedback by the AI feedback has a significant difference. While 318 (75.89%) of 419 feedback given by AI were lower, 159 (57.61%) of 276 feedback given by teachers and 160 (56.54%) of 283 feedback given by hybrid were lower.

In conclusion, when analyzed in the main feedback category, AI is more effective in providing feedback for LOCs, while teacher and hybrid feedback demonstrate comparable performance. However, teachers show significantly lower performance in lower feedback categories.

5.2. Qualitative Research Results

The qualitative data was collected through semi-structured interviews with four EFL teachers regarding their experience and perceptions of the use of ChatGPT. Eighteen questions were asked about strengths, challenges, and further suggestions regarding

teacher, AI, and hybrid feedback. Recordings of the interviews were transcribed by the researcher and content analysis was employed to interpret the data.

5.2.1. What are EFL teachers' perceptions of the strengths, challenges, and further suggestions for different feedback types on EFL students' writing?

In the qualitative part of the study, semi-structured interviews were completed with the participant teachers via the online meeting platform Zoom. The transcribed data was coded by content analysis and divided into three main themes, strengths, challenges, and further suggestions in terms of teacher, ChatGPT, and hybrid feedback.

5.2.1.1. What strengths and challenges do teachers perceive in each feedback type's ability to address main-feedback categories (HOCs and LOCs) and sub-feedback categories (content, organization, vocabulary, grammar, and mechanics)?

In this part, the strengths of feedback sources were discussed respectively as teacher, AI, and hybrid feedback in line with research questions. Three sub-themes were created under teacher, AI, and hybrid feedback to be able to categorize the codes meaningfully. *LOCs*, *HOCs*, and *additional comments* were the sub-themes under each feedback source.

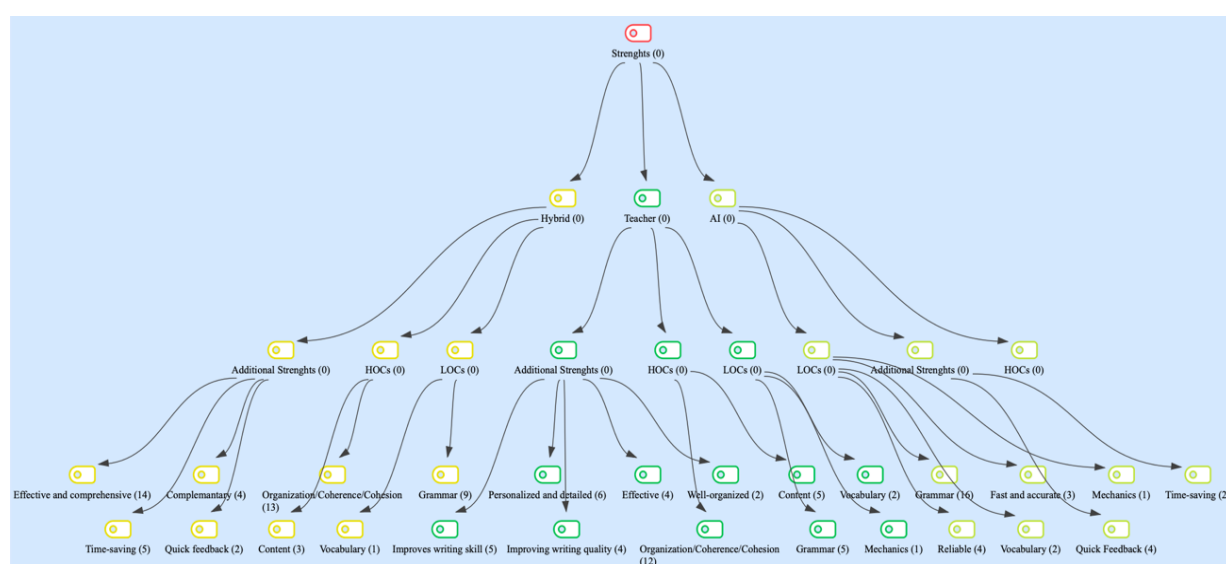


Figure 11. Strengths of teacher, AI and hybrid feedback in MaxQDA

Strengths of teacher feedback:

The analysis of interview data showed that teacher feedback has strengths for LOCs (n=8), HOCs (n=17) and additional strengths (n=21). Thus, the analysis showed that EFL teachers believe teacher feedback has additional strengths, HOCs and LOCs, respectively.

In terms of LOCs, teacher feedback was favored for the correction of ‘grammar’ (n=5), ‘vocabulary’ (n=2) and ‘mechanics’ (n=1) by the participant teachers. Extracts below show teachers’ favor for LOCs:

*Teacher feedback is capable of addressing LOCs, such as **correcting verb forms or articles** (T2).*

*Teacher feedback can effectively address LOCs like **grammar mistakes, punctuation mistakes and misspellings** (T4).*

‘Organization/coherence/cohesion’ (n=12) was the most favored part of the HOCs, and ‘content’ mistakes (n=5) were also mentioned by the participant teachers in terms of strengths of teacher feedback. The data showed that teacher feedback has more strengths in terms of HOCs than LOCs. Extracts below indicate teachers’ favor for HOCs:

*Teacher-generated feedback is **essential** and very **effective**, especially for HOCs for issues regarding **content** and **coherence/cohesion** (T1).*

*Teacher feedback improves **content development, organization of the paragraphs** (T2).*

*Teachers focus a lot on helping students **organize their ideas, develop their arguments** and ensure **coherence in their writing** (T3).*

In terms of additional strengths, it is believed that teacher feedback is ‘improving writing quality and writing skill’ (n=9), ‘personalized and detailed’ (n=6), ‘effective’ (n=4), and ‘well-organized’ (n=2). Extracts below support these:

*As an EFL teacher, I find my feedback significantly **enhances students’ writing ability** (T2).*

*In my opinion, teacher-generated feedback plays a crucial role in **improving writing quality** because it **provides personalized and detailed feedback** to students’ specific needs (T3).*

Teacher feedback fosters a long-term improvement in their writing skills (T3).

Unlike automated tools, my (teacher) feedback is detailed and context-specific, which makes it more meaningful for students (T4).

The interview data highlighted that teacher feedback is valued for enhancing writing quality, addressing HOCs like organization and content, and correcting LOCs such as grammar and mechanics. Teachers emphasized its personalized, detailed, and context-specific nature, which supports both immediate writing improvements and long-term skill development. Overall, teacher feedback is seen as critical for fostering effective and meaningful progress in students' writing abilities.

Strengths of AI feedback:

Semi-structured interviews showed that AI feedback has strengths in terms of LOCs (n=26) and 'additional strengths' (n=6). EFL students did not state any strengths regarding HOCs. Thus, AI feedback has strengths regarding LOCs.

'Grammar' (n=16), 'vocabulary' (n=2) and 'mechanics' (n=1) aspects were praised in terms of LOCs. Also, EFL teachers stated that AI feedback is 'reliable' (n=4) and 'fast and accurate' (n=3) in terms of LOCs. Extracts below support this view:

ChatGPT is really great when it comes to identifying LOCs quickly and accurately. For example, in one essay from the study, it detected a subject-verb agreement issue: "She play guitar on weekends" was corrected to "She plays guitar on weekends". This kind of feedback helps students improve grammatical accuracy without my direct involvement (T1).

I've realized that ChatGPT also highlights redundant phrases such as "basic and simple", suggesting one to be removed for conciseness (T1).

ChatGPT is excellent for LOCs, such as identifying grammatical errors, correcting punctuation and finding spelling mistakes (T2).

ChatGPT is especially helpful with lower-order concerns, like grammatical errors and spelling (T3).

ChatGPT is great for spotting grammatical errors in sentence structure and even offering alternatives for vocabulary (T3).

ChatGPT quickly identifies and corrects grammar, vocabulary and mechanics issues, which saves me time and allows students to immediately see areas of improvement (T4).

Regarding additional strengths, EFL teachers stated that AI feedback is ‘quick’(n=4) and ‘time-saving’ (n=2). Extracts below indicate teachers’ these views:

ChatGPT feedback is great for providing quick, immediate responses (T3).

When I first use ChatGPT to provide feedback, I appreciated how it could instantly identify and correct errors (T4).

The interview data suggested that AI feedback, particularly from ChatGPT, excels in addressing LOCs such as grammar, vocabulary, and mechanical issues. Teachers highlighted its efficiency in quickly identifying and correcting errors, such as subject-verb agreement and redundancy, with a focus on improving grammatical accuracy and sentence structure. Teachers also praised AI feedback for its speed and accuracy, emphasizing its ability to save time and provide immediate, actionable feedback for students. However, the data revealed no significant strengths for AI feedback in addressing HOCs, indicating that while AI is effective for technical improvements, it may lack the depth required to handle more complex writing aspects.

Strengths of hybrid feedback:

Interview data showed that hybrid feedback has strengths in terms of LOCs (n=10), HOCs (n=16) and additional strengths (n=25).

In terms of LOCs, ‘grammar’ (n=9) is the most favored part by the teachers. ‘Vocabulary’ (n=1) was also mentioned. Extracts below indicate favor for LOCs corrections:

I think hybrid feedback is practical because it combines speed and depth.

During the study, ChatGPT identified LOC errors like incorrect propositions “He is interested football” corrected to “He is interested in football”. This freed up my time to focus on HOCs (T1).

Hybrid feedback creates a balance by dividing the workload effectively.

ChatGPT enhances grammatical accuracy and eliminates redundancy

*while I guide students in **improving the logical flow and depth of their arguments** (T2).*

*ChatGPT helps with **grammar, vocabulary and surface-level issues**, while teacher feedback focuses on **idea development, coherence and organization** (T3).*

*From my experience, hybrid feedback is a powerful approach. ChatGPT handles **repetitive and time-consuming LOC corrections**, allowing me to focus on **higher-level skills like argument development** (T4).*

‘Organization/coherence/cohesion’ (n=13) was the most favored part of the HOCs, ‘content’ (n=3) followed. Extracts below indicated favor for HOCs:

*ChatGPT freed up my time to focus on HOCs, such as **explaining why the student’s argument lacked coherence or suggesting a better paragraph order** (T1).*

*With the help of hybrid feedback, I can focus on **helping students develop a strong conclusion or organize their essay more effectively without being distracted by surface-level errors** (T2).*

*ChatGPT helps with **grammar, vocabulary and surface-level issues**, while teacher feedback focuses on **idea development, coherence and organization** (T3).*

*Hybrid feedback effectively balances HOCs and LOCs corrections. ChatGPT ensures **linguistic and grammatical accuracy**, while I focus on improving **organization, transitions and content accuracy** (T4).*

In terms of additional strengths, according to the EFL teachers, hybrid feedback is ‘effective and comprehensive’ (n=14), ‘time-saving’ (n=5), ‘complementary’ (n=4) and ‘quick’ (n=2). The extracts below indicate these views:

*Hybrid feedback ensures that students receive **comprehensive feedback** without being overwhelmed. Other than this research, I already employed ChatGPT feedback as **complementary** to my feedback which we can say “hybrid feedback” (T1).*

*Hybrid feedback is **highly effective** because it combines the strengths of both ChatGPT and teacher feedback. ChatGPT’s ability to handle LOCs **frees up my time** to focus on HOCs (T2).*

*Honestly, I found the hybrid feedback approach to be **the most effective**. ChatGPT handled the LOCs **quickly and efficiently**, which **freed up time for teachers** to focus on HOCs. The combination allows teachers to **detect mistakes very fast and easy way** which makes the way **very useful** (T3). This combination (ChatGPT and Teacher feedback) enhances students' **overall writing** by ensuring both **surface-level accuracy** and **deeper content engagement** (T4).*

The interview data highlighted the strengths of hybrid feedback in addressing both HOCs and LOCs and having additional strengths. Hybrid feedback was praised for its effectiveness and balance, combining the precision and speed of ChatGPT in correcting LOCs such as grammar and vocabulary with the depth and contextual focus of teacher feedback on HOCs, particularly organization and content. Teachers emphasized that this approach allows ChatGPT to handle time-consuming surface-level corrections, enabling them to focus on deeper aspects of students' writing, such as coherence, argument development, and content structure. Overall, the hybrid method is seen as an effective and efficient way to enhance both surface-level accuracy and meaningful content engagement in students' writing.

Challenges of teacher, AI and hybrid feedback:

This part discussed the challenges of feedback sources as teacher, AI, and hybrid feedback in line with research questions. Three sub-themes were created under teacher, AI and hybrid feedback to be able to categorize the codes meaningfully. *LOCs*, *HOCs*, and *additional challenges* were the sub-themes under each feedback source.

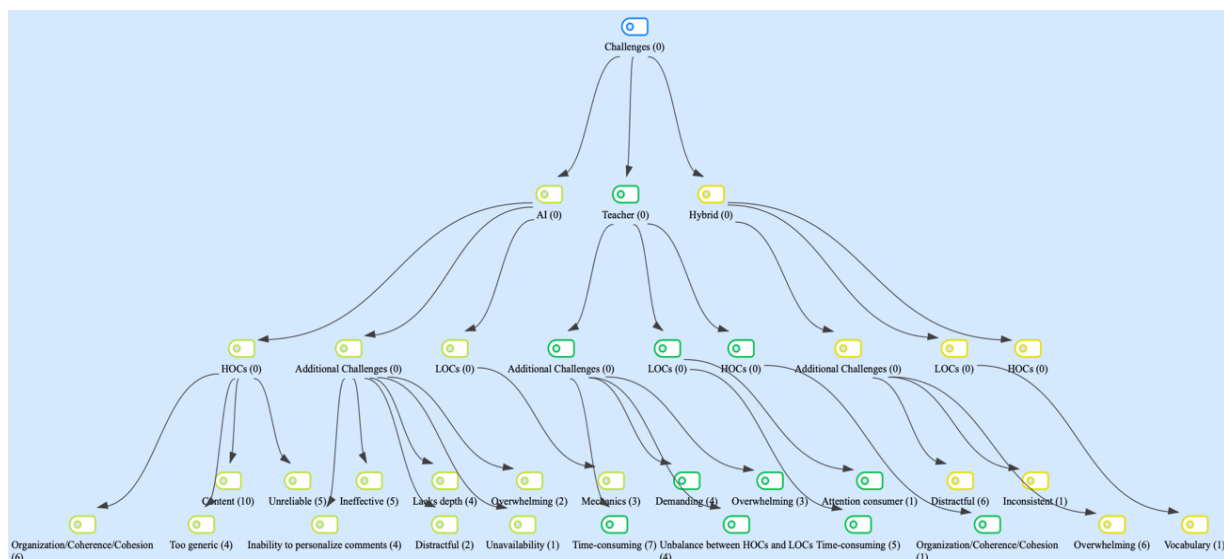


Figure 12. Challenges of teacher, AI and hybrid feedback in MaxQDA

Challenges of teacher feedback:

Interview data showed that teacher feedback has challenges in terms of LOCs (n=6), HOCs (n=1), and additional challenges (n=18). Thus, the analysis showed that EFL teachers believe that teacher feedback has mostly additional challenges and LOCs.

In terms of LOCs, teacher feedback corrections regarding LOCs are ‘time-consuming’ (n=5) and ‘attention-consumer’ (n=1). Extracts below support this view:

*Addressing LOCs, such as grammar mistakes, **takes up a lot of time**. I spent so much time correcting LOCs like missing determiners (T1).*

*LOCs can be **time-consuming** which sometimes hinders my attention from giving more detailed feedback on HOCs. (T2).*

In terms of HOCs, EFL teachers stated only one challenge regarding HOCs, which is about ‘organization/coherence/cohesion’ (n=1) Extract below indicates EFL teacher’s views on challenges regarding HOCs:

*If I focus on heavily on LOCs, students might improve their grammar but fail to develop their **ideas** or **organization**. On the other hand, prioritizing HOCs can leave persistent grammar issues unaddressed (T4).*

Additional challenges were stated by teachers as ‘time-consuming’ (n=7), ‘demanding’ (n=4), ‘unbalance between HOCs and LOCs’ (n=4) and ‘overwhelming’ (n=3). Extracts below indicate these challenges:

*One major challenge is **time**. I have 3-4 classes every term and this makes a 80-120 students for a semester. When it comes to provide written feedback to this number of students, it is really **demanding** and **burdensome** for me and my colleagues (T1).*

***Balancing HOCs and LOCs is tough**. For instance, when students receive detailed grammar corrections, they might **overlook my comments** for serious mistakes like **the flow of paragraphs** or **issues related to content of the essay**. On the other hand, if I focus mostly on HOCs this time **surface-level errors can left undetected** (T1).*

*One challenge is **the time** required to address both HOCs and LOCs. For instance, while I am comfortable providing detailed feedback on HOCs, I often find myself **spending too much time** correcting grammar or syntax issues (T2).*

*Addressing both types of concerns can be **overwhelming** for both me and the students (T2).*

*One big challenge is **the time it takes** for teachers to provide detailed feedback (T3).*

*The biggest challenge I face is **the time required** to provide comprehensive feedback. Providing detailed feedback is really **tiring** and **burdensome** (T4).*

*Prioritizing HOCs can leave persistent grammar issues **unaddressed** (T4).*

*It can be **overwhelming** to address repetitive LOC errors in many papers while also focusing on HOCs (T4).*

The interview data highlighted several challenges associated with teacher feedback. The most significant issues relate to writing quality and LOCs, with teachers emphasizing that providing detailed feedback is both time-consuming and demanding. Corrections for LOCs, such as grammar, often require substantial time and attention, which can detract from addressing HOCs like organization and idea development. Balancing HOCs and LOCs was noted as a persistent challenge, with teachers expressing concerns about students overlooking critical comments on HOCs when too much focus is placed on LOCs, and vice versa. Additionally, the burden of providing comprehensive feedback to

many students was described as overwhelming and exhausting. These findings suggested that while teacher feedback is highly effective, its labor-intensive nature presents a significant barrier to its consistent application, particularly in managing both surface-level accuracy and deeper content-related issues simultaneously.

Challenges of AI feedback:

Semi-structured interview data indicated that AI feedback has challenges in terms of LOCs(n=3), HOCs (n=25) and additional challenges (n=18). EFL teachers indicated that AI feedback has the most challenges on HOCs.

In terms of LOCs, ‘mechanics’ (n=3) was the only aspect mentioned by EFL teachers. Extracts below indicate EFL teachers’ comments on LOCs in terms of AI feedback:

*It (ChatGPT) flagged **every missing coma** in a student’s essay, such as “After school I go home and relax” suggestions a comma after “school.” While technically correct, this **overemphasis on punctuation** can make the students feel they have too many serious errors (T1).*

*It (ChatGPT) might **overemphasize punctuation issues**, such as suggesting unnecessary commas (T2).*

*It (ChatGPT) sometimes **overemphasizes minor mechanical issues** like punctuation (T4).*

In terms of HOCs, ‘content’ (n=10), ‘organization/coherence/cohesion’ (n=6), ‘unreliable’ (n=5) and ‘too generic’ (n=4) were the aspects stated by the EFL teachers. Extracts below indicate EFL teachers’ views on the challenges of HOCs in AI feedback:

*While ChatGPT is reliable for LOCs, its handling of HOCs is **limited** (T1).*

*ChatGPT’s HOCs feedback often lacks **depth**. For example, in one essay, it said “improve transitions”, without suggesting how or where to add them (T1).*

*When it comes to HOCs, such as analyzing whether **a thesis statement is clear** or whether **ideas are logically placed**, I don’t fully trust its (ChatGPT) feedback because it often **lacks depth and contextual understanding** (T2).*

*ChatGPT is **reliable for LOCs** but **falls short with HOCs**. For instance, it might point out that a **paragraph lacks a topic sentence** but its suggestions*

are mostly **too generic and not detailed to the student's intended meaning** (T2).

While it (ChatGPT) performs well at correcting issues like verb tense or redundant phrases, it **struggles to provide reliable advice on how to improve coherence or develop ideas further** (T2).

Its (ChatGPT) impact on higher order concerns such as **developing ideas or ensuring accuracy of information is limited** (T3).

I'd say ChatGPT struggles most with HOCs, like **evaluating the accuracy of claims or the logical flow of ideas**. It can identify that a paragraph might lack coherence, but it doesn't provide **in-depth guidance** on how to fix it (T3).

The main challenge with ChatGPT is that it lacks **context**. It can identify errors and suggest changes, but it **doesn't always understand the specific goals of an assignment**. During the research, I noticed it sometimes **offered generic suggestions** that didn't align with what the student needed. This was particularly evident in addressing HOCs, where its feedback felt superficial. Also, especially in Türkiye's context, these kinds of tools **are not accessible** everywhere (T3).

It struggles with HOCs like **evaluating the coherence of ideas or the logical flow of arguments** (T4).

It (ChatGPT) can comment on the structure of an essay but often misses issues like the **development of ideas or the accuracy of information** (T4).

AI feedback has some additional challenges. They are 'ineffective (n=5), 'inability to personalize comments' (n=4), 'lacks depth' (n=4), 'distracting' (n=2), 'overwhelming' (n=2) and 'unavailability' (n=1). Extracts below indicate these challenges:

For example, it might indicate that a paragraph is "too short" but **fail to suggest how to expand it meaningfully** (T1).

ChatGPT can **overwhelm students** by correcting LOCs (T1).

It (ChatGPT) can detect a lack of cohesion between paragraphs but doesn't provide concrete, actionable steps to improve the logical flow or develop ideas in a better way. This limitation makes it **non-effective** for addressing the deeper, more complex aspects of writing (T2).

The biggest challenge with ChatGPT is its **inability to personalize its comments** (T2).

Especially in Türkiye's context, these kinds of tools (ChatGPT) are not accessible everywhere (T3).

ChatGPT cannot evaluate whether a claim is accurate or provide meaningful explanations or suggestions for improving the flow of the paragraph (T4).

While ChatGPT is really helpful at identifying surface-level errors, it often lacks the critical understanding needed for deeper feedback (T4).

*It (ChatGPT) sometimes overemphasizes minor mechanical issues like punctuation, which can **distract students** from more significant concerns. This makes it less reliable for addressing deeper aspects of writing (T4).*

The interview data highlighted that AI feedback, particularly from ChatGPT, faces significant challenges in addressing HOCs. They also face some additional challenges while having fewer challenges with LOCs. Teachers noted that AI feedback tends to overemphasize minor mechanical issues, such as punctuation, which can distract students from more important aspects of their writing. For HOCs, AI feedback was criticized for being too generic, lacking depth, and failing to provide actionable, context-specific suggestions for improving coherence, content development, or organization. Regarding additional challenges, AI feedback was seen as ineffective in personalizing comments and lacking the ability to adapt feedback to specific contexts or student needs. These limitations, combined with occasional accessibility issues in certain contexts, highlight that while AI feedback is efficient for LOCs, it falls short in offering meaningful guidance for deeper aspects of writing.

Challenges of hybrid feedback:

The interview data showed that hybrid feedback has challenges in terms of LOCs (n=1) and additional challenges (n=15). EFL teachers did not state any challenge on HOCs regarding hybrid feedback.

In terms of LOCs, 'vocabulary' (n=1) was the only aspect mentioned by EFL teachers. The extract below indicates the challenge regarding LOCs:

*ChatGPT might recommend a **word choice** that the teacher didn't find suitable for that context (T3).*

Hybrid feedback has some additional challenges like ‘overwhelming’ (n=6), ‘distracting’ (n=6) and ‘inconsistent’ (n=1). Extracts below indicate the additional challenges of hybrid feedback stated by the teachers:

*The main challenge is the **volume of feedback**. During the study, one student's essay received corrections from ChatGPT for every minor issue (e.g., extra spaces and punctuation), which can be **distracting**, while my comments focused on major HOCs. This **overwhelmed the student**, making it hard for them to prioritize revisions (T1).*

*The main is ensuring that the feedback volume is not **overwhelming** for students (T2).*

*The biggest challenge with hybrid feedback is ensuring **consistency**. Sometimes, ChatGPT's suggestions conflicted with the teacher's comments, which **can distract students** (T3).*

*Students can occasionally feel **overwhelmed** by the volume of feedback when using the hybrid approach (T3).*

*If ChatGPT focuses too many minor issues, **students may focus on fixing those while skipping larger, more important HOC-related concerns** (T4).*

Students sometimes find it difficult to prioritize hybrid feedback, especially when ChatGPT's LOC-focused comments are more detailed than my HOC-focused ones. This can make it **harder for them to understand which changes are most critical** (T4).

The interview data highlighted that the main challenges of hybrid feedback lie in writing quality, with minimal issues noted for LOCs and no challenges reported for HOCs. In terms of LOCs, a minor challenge was related to vocabulary, where teachers did not sometimes approve ChatGPT's word choice suggestions. Regarding additional challenges stated by the teachers, hybrid feedback was reported to be occasionally overwhelming and distracting due to the volume of feedback, particularly when ChatGPT provided extensive corrections on minor issues. These findings suggest that while hybrid feedback is effective, ensuring a balanced and advanced approach to feedback delivery is crucial to avoid overwhelming students and to maintain clarity in the revision process.

5.2.1.2. What suggestions do teachers have for improving each feedback type to better support EFL students' writing feedback categories, specifically in relation to HOCs and LOCs?

In this part of the study, suggestions for each feedback source were discussed as teacher feedback, AI feedback and hybrid feedback, respectively. Three sub-themes were created under teacher, AI and hybrid feedback to be able to categorize the codes meaningfully. *LOCs*, *HOCs*, and *additional suggestions* were the sub-themes under each feedback source.

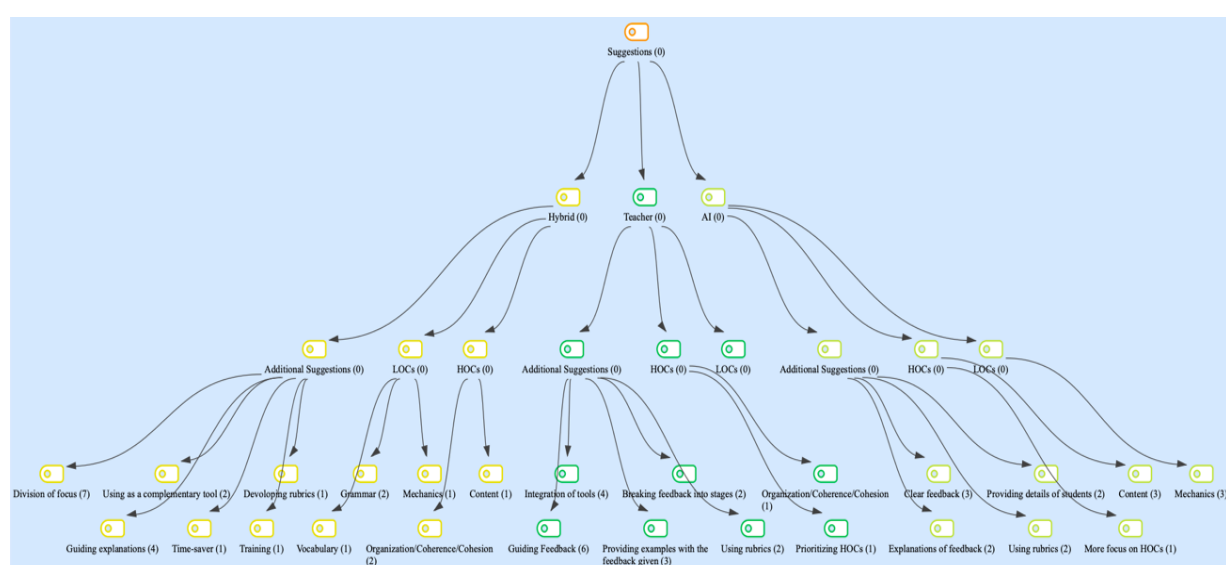


Figure 13. Suggestions for teacher, AI and hybrid feedback in MaxQDA

Suggestions for teacher feedback:

Interview data showed that EFL teachers stated suggestions in terms of HOCs (n=2) and additional suggestions (n=17) for teacher feedback. In contrast, no suggestion was made regarding LOCs by EFL teachers.

In terms of HOC, 'prioritizing HOCs' (n=1) and 'organization/coherence/cohesion' (n=1) were the aspects stated by EFL teachers. Extract below indicates the suggestions of EFL teachers regarding HOCs in teacher feedback:

*Creating a detailed guidelines for areas like **coherence** and **paragraph organization** could help students understand the importance of these aspects (T2).*

*To improve teacher feedback, I suggest focusing more **on prioritizing HOCs** during the feedback process (T2).*

EFL teachers made several additional suggestions. They were ‘guiding feedback’ (n=6), ‘integration of tools’ (n=4), ‘providing examples with the feedback’ (n=3), ‘breaking the feedback into stages’ (n=2) and ‘using rubrics’ (n=2). Extracts below indicate the suggestions of EFL teachers regarding teacher feedback:

*Another refinement could be using **guiding**, actionable comments. Instead of saying, "Your argument is unclear," a more effective approach might be, "Add an example or statistic to support this argument." This kind of **specific guidance** makes revisions more manageable for students (T1).*

*Teacher feedback could be refined by **using technology** for repetitive LOC corrections allowing more time for detailed HOC guidance (T1).*

***Using examples or templates** can clarify expectations, **such as showing how to write a clear thesis statement or provide a strong transition between paragraphs** (T1).*

***Dividing comments into two sections**, like HOCs and LOCs, helps students focus on the most critical areas first (T1).*

Using consistent terminology for HOCs, like topic sentences, transitions, can help students better interpret and apply the feedback (T2).

*A possible refinement is to **integrate tools like rubrics or checklists** into the feedback process (T2).*

***Using consistent terminology** can help students better interpret and apply the feedback (T2).*

*One way is to **integrate technology like using tools** to address LOCs before the teacher focuses on HOCs (T3).*

***Guiding students** how to rewrite a confusing sentence or rephrase a weak argument would make the feedback more practical (T3).*

*I think teacher feedback could be more effective if it includes **examples alongside comments** (T3).*

Breaking feedback into stages, starting with HOCs and moving to LOCs later, could also help students revise more effectively (T3).

Providing examples or model sentences can help students better understand abstract HOC concepts (T4).

Using clear rubrics, like the one provided in the study, helps organize the feedback and ensures all aspects are covered (T4).

The interview data highlighted several suggestions for enhancing teacher feedback, particularly regarding HOCs and teachers emphasized the importance of prioritizing these concerns and providing detailed guidelines for areas like coherence and paragraph organization to help students better understand and address these aspects. Regarding additional suggestions, teachers suggested making feedback more actionable and specific, such as offering examples or templates to clarify expectations and using guiding comments to make revisions more manageable for students. Integrating tools to handle repetitive LOC corrections was also recommended, allowing teachers to focus on HOCs. Additional suggestions included breaking feedback into stages (e.g., addressing HOCs first, then LOCs), using rubrics or checklists for consistency, and employing clear terminology to enhance students' understanding and application of feedback. These refinements aim to make teacher feedback more practical, focused, and effective in supporting students' writing development.

Suggestions for AI feedback:

The analysis of interview data indicated that EFL teachers stated suggestions for LOCs(n=3), HOCs (n=6) and additional suggestions (n=9).

In terms of LOCs, the only aspect mentioned by EFL teachers was 'mechanics' (n=3). Extracts below include suggestions for AI feedback:

*During the study, students often felt overwhelmed by **excessive comments on minor punctuation errors**, which distracted them from addressing larger issues. **Reducing the focus on minor issues, like overcorrecting commas**, would make the feedback less overwhelming and more targeted (T1).*

*ChatGPT could **focus less on mechanical corrections like punctuation, especially commas** (T2).*

In terms of HOCs, ‘content’ (n=3) and ‘more focus on HOCs’ (n=3) stated by EFL teachers. Extracts below indicate of EFL teachers’ suggestions regarding AI feedback:

*To improve HOCs feedback, ChatGPT could provide **more detailed, context-specific** suggestions (T1).*

*When identifying **a weak thesis statement**, it could suggest how to make it more argumentative or aligned with the **essay's content** (T2).*

*If ChatGPT could provide more **context-sensitive** feedback, it would be a game-changer (T3).*

*ChatGPT could be improved by **focusing more on HOCs** such as **idea development and organization of the paragraphs** (T4).*

EFL teachers made several additional suggestions. They were ‘clear feedback (n=3), ‘explanations of feedback (n=2), ‘providing details of students’ (n=2) and ‘using rubrics’ (n=2). Extracts below indicate the additional suggestions regarding AI feedback:

*ChatGPT could be improved by letting it know **more information about students and different proficiency levels** (T1).*

*To make ChatGPT feedback more effective, it should provide **a more detailed understanding of student proficiency levels and cultural contexts** (T2).*

*Instead of suggesting overly advanced vocabulary or generic phrases, it (ChatGPT) could **advise clear suggestions** to learners’ level (T2).*

*ChatGPT could improve its handling of HOCs by offering **more clear, feasible advice** (T2).*

***Using rubrics**, ChatGPT could categorize errors by importance so students can prioritize corrections (T4).*

The interview data highlighted several suggestions for improving AI feedback, focusing on LOCs, HOCs, and writing quality. For LOCs, teachers suggested reducing the emphasis on minor mechanical issues, like excessive corrections of punctuation, as this can overwhelm students and distract them from addressing more critical writing concerns. Regarding HOCs, teachers recommended that ChatGPT provide more detailed and context-specific feedback, particularly on content development and paragraph organization, to support higher-level writing improvements better. Additional suggestions included adjusting feedback to students' proficiency levels and cultural contexts, providing clearer and more actionable advice, and incorporating rubrics to help students

prioritize corrections effectively. These suggestions aim to make AI feedback more targeted, practical, and supportive of both surface-level and deeper aspects of student writing.

Suggestions for hybrid feedback:

The interview data showed that EFL teachers stated suggestions for LOCs (n=4), HOCs (n=3) and additional suggestions (n=16) regarding hybrid feedback.

In terms of LOCs, ‘grammar’ (n=2), ‘vocabulary’ (n=1) and ‘mechanics’ (n=1) were mentioned by EFL teachers. Extracts below indicate the suggestions in terms of LOCs regarding hybrid feedback:

*After ChatGPT **corrects grammar and syntax errors**, I can guide students on how to strengthen and improve paragraphs in terms of HOCs (T2).*

*For EFL students, creating a clear division of labor between ChatGPT and teacher feedback could help. ChatGPT could address **grammar, vocabulary and mechanics**, while teachers focus on clarity, argumentation, and organization (T3).*

In terms of HOCs, ‘organization/coherence/cohesion’ (n=2) and ‘content’ (n=1) were the aspects for suggestions regarding HOCs in hybrid feedback. Extracts below indicate suggestions for hybrid feedback:

*For EFL students, creating a clear division of labor between ChatGPT and teacher feedback could help. ChatGPT could address grammar, vocabulary and mechanics, while teachers focus on **clarity, argumentation and organization** (T3).*

There were also some additional suggestions. ‘Division of focus’ (n=7), ‘guiding explanations’ (n=4), ‘using as a complementary tool’ (n=2), ‘time-saver’ (n=1), ‘developing rubrics’ (n=1) and ‘training teachers on hybrid feedback’ (n=1) were the suggestions stated by EFL teachers. Extracts below indicate these suggestions:

*To enhance hybrid feedback, I recommend integrating it in **a staged process** like in the research: ChatGPT first handles LOCs, such as grammar and punctuation, followed by teacher feedback focusing on HOCs. During the*

*research, I think this method worked effectively because ChatGPT's ability to detect LOCs were enough and it **freed up my time** (T1).*

***Including a feedback summary** at the end of the document that highlights major strengths and weaknesses can help EFL students navigate and apply hybrid feedback efficiently (T1).*

*One recommendation is to establish a clear **division of focus**: ChatGPT handles LOCs, while teacher feedback concentrates on HOCs (T2).*

***A clear workflow** would make hybrid feedback more effective. For example, using ChatGPT **for a first round of corrections on LOCs, then following up with teacher feedback for HOCs**, would facilitate the process (T3).*

***Training sessions** on how to interpret and combine both feedback types would be beneficial for EFL teachers (T3).*

*For instance, **ChatGPT could handle LOCs entirely, while I focus solely on HOCs** (T4).*

*Teachers should use it (ChatGPT) **as a complementary tool** for feedback rather than completely replacing teacher feedback (T4).*

***Developing clear rubrics** for combining ChatGPT and teacher feedback would improve the hybrid approach (T4).*

The interview data suggested several ways to enhance hybrid feedback, focusing on LOCs, HOCs, and writing quality. For LOCs, teachers recommended creating a clear division of labor where ChatGPT addresses grammar, vocabulary, and mechanics, freeing up teachers to focus on higher-level skills. Regarding HOCs, suggestions emphasized prioritizing aspects like organization, coherence, and content development, with ChatGPT and teacher feedback complementing each other effectively. Teachers also proposed integrating hybrid feedback into a staged process, where ChatGPT first corrects LOCs, followed by teacher input on HOCs. Providing a summary of major strengths and weaknesses at the end of feedback was also recommended to help students navigate revisions more effectively. Teachers also suggested using hybrid feedback as a complementary tool rather than a replacement. Training sessions for teachers on interpreting and combining feedback types were also highlighted as essential for maximizing the benefits of hybrid feedback.

6. DISCUSSION

This study examined the interaction among teacher, AI and hybrid feedback types in terms of main-feedback categories, HOCs and LOCs, and sub-feedback categories, content, organization/coherence/cohesion, vocabulary, grammar and mechanics, in the quantitative part of the study. For the qualitative part of the study, semi-structured interviews were conducted to see EFL teachers' perceptions regarding the strengths and challenges of different feedback types. In addition, EFL teachers' suggestions for improving each feedback type were gathered during the interviews. The findings will be presented in this chapter in accordance with each research question.

6.1. Interaction of Teacher, AI, and Hybrid Feedback across Feedback Categories

The findings of this study reveal significant differences in the quality of essays rated under teacher, ChatGPT, and hybrid feedback conditions, shedding light on the strengths and limitations of each feedback type. These results contribute to the growing discourse on the role of AI in writing education, emphasizing the potential of hybrid feedback systems to enhance both teaching efficiency and student outcomes.

6.1.1. Variations in feedback categories across feedback types

The findings of this study reveal significant differences in the distribution of feedback categories (HOCs and LOCs) across teacher, ChatGPT, and hybrid feedback types. The statistical analysis demonstrated that hybrid feedback resulted in the highest number of total feedback instances, aligning with previous literature that highlights the benefits of combining automated and human feedback mechanisms (Wilson & Czik, 2016). The results indicate that hybrid feedback is particularly effective in addressing both HOCs and LOCs, which supports the argument that integrating AI-generated feedback with teacher input provides a more comprehensive evaluation of student writing (Jiang et al., 2020; Link et al., 2020).

The logit values from the results section indicate that feedback was generally provided at a moderate level. However, essays receiving hybrid feedback exhibited a significantly higher concentration of comments. This suggests that the hybrid model successfully

leverages the immediacy of AI-generated corrections for LOCs while allowing teachers to focus on discourse-level revisions. This supports Warschauer and Grimes (2008), who argued that automated tools enhance feedback efficiency by enabling teachers to allocate more time to higher-order concerns such as organization and argumentation.

Teacher feedback primarily targeted HOCs, including idea development, logical flow, and coherence. These findings are consistent with Ferris (2006), who emphasizes the essential role of teacher feedback in addressing discourse-level writing aspects. However, Truscott (1996) argues that the time-intensive nature of detailed teacher feedback may limit its frequency and consistency, particularly regarding LOCs. The results of this study support this perspective, as teacher-evaluated essays exhibited lower accuracy in surface-level mechanics compared to those reviewed by ChatGPT or hybrid methods. This highlights the inherent challenge of providing extensive grammar and mechanics corrections while also addressing conceptual writing aspects.

Conversely, ChatGPT feedback exhibited a strong focus on LOCs, reinforcing previous findings that AWE tools are highly effective in detecting and correcting grammatical and syntactical errors (Chapelle et al., 2015; Dikli, 2010). However, as Link et al. (2020) noted, such systems often lack the contextual understanding necessary for guiding students in idea development and coherence. The results confirm this limitation, as essays reviewed exclusively by ChatGPT received fewer comments related to HOCs and organization, despite demonstrating improved surface-level accuracy.

The presence of statistical outliers (Essays 5, 17, 26, and 31) suggests some variability in rater evaluations, which may be influenced by the nature of the essays themselves or differences in how feedback was interpreted. This aligns with research indicating that human raters introduce subjectivity into feedback practices, which can lead to inconsistencies in evaluation (Wilson & Czik, 2016). The significant chi-square results confirm that the distribution of feedback categories varied significantly across feedback types, further reinforcing the need for a structured approach to integrating AI in writing evaluation.

The pedagogical implications of these findings suggest that hybrid feedback offers an optimal balance, supporting Swain's (1985) output hypothesis by providing students with

targeted and diverse feedback. This aligns with studies that emphasize the importance of combining AI efficiency with human expertise to enhance both writing accuracy and conceptual depth (Jiang et al., 2020; Link et al., 2020). The ability of hybrid systems to provide comprehensive feedback while reducing the burden on instructors presents a promising direction for writing pedagogy, ensuring that students receive immediate corrective feedback while also benefiting from deeper discourse-level guidance.

This study contributes to the growing discourse on AI integration in language education by demonstrating that feedback types differ not only in their focus but also in their effectiveness. The superior performance of hybrid feedback in addressing both HOCs and LOCs suggests that future writing instruction should incorporate AI-assisted feedback mechanisms alongside traditional teacher evaluation. This confirms previous research that highlights the complementary nature of automated and human feedback systems (Link et al., 2020; Wilson & Czik, 2016).

6.1.2. Variations in feedback categories and patterns

The findings of this study indicate significant variability in feedback patterns and categories among teacher, ChatGPT and hybrid feedback types. These differences are particularly evident when comparing single raters (teachers or ChatGPT) to combined rater types (hybrid), shedding light on the strengths and limitations of each approach.

The statistical results reveal distinct feedback patterns across the three rater types. ChatGPT's feedback was mostly focused on LOCs, such as grammar and mechanics. In contrast, teacher feedback focused more on HOCs, such as content development and organizational structure. Hybrid feedback achieved the most balanced distribution, addressing both LOCs and HOCs. This balance highlights the ability of hybrid feedback systems to combine the efficiency of AI tools with the comprehensive understanding provided by teacher expertise. These results align with findings from Link et al. (2020), who demonstrated that automated feedback systems excel in addressing surface-level issues. However, they fall short in providing meaningful guidance for complex writing tasks. Similarly, Ferris (2006) emphasized the strength of teacher feedback in fostering critical thinking and idea development while noting its limitations in addressing time-

intensive LOC corrections. The hybrid model in this study exemplifies the complementarity of these feedback types, as it effectively balances the strengths of each.

This study confirmed that the AI tool provided quick, consistent, and repetitive feedback on surface-level errors. This outcome supports prior findings indicating that AWE tools address grammatical issues and mechanical accuracy (Dikli, 2010; Ranalli et al., 2016; Shaikh et al., 2023). Additionally, the study demonstrated that AI systems like ChatGPT are highly effective in reducing the time spent on surface-level error detection while offering reliable and instant corrections, consistent with previous findings (Allehyani & Algamdi, 2023; Athanassopoulos et al., 2023; Chang et al., 2024; Chen & Cheng, 2008; Nugroho et al., 2023).

Another key finding was that ChatGPT mostly focused on LOCs, providing significantly more feedback on grammar and mechanics than on content, organization, or coherence. This observation aligns with earlier studies by Dikli and Bleyle (2014), Ranalli et al. (2016), and Chukharev-Hudilainen and Saricaoglu (2014), which reported that AWE systems prioritize accuracy over discourse-level concerns. Conversely, this finding contrasts with Wilson and Czik (2016), who observed that AWE tools in hybrid feedback conditions offered significantly more feedback on higher-level writing subskills than the teacher-feedback-only condition. Nevertheless, consistent with the broader literature, this research found that the hybrid feedback model provided HOC feedback the most.

However, contrasting findings also emerged when comparing this study to Link et al. (2020), where teachers in the AWE + teacher feedback condition provided less feedback than those in the teacher-only feedback condition. In Link et al.'s study, AWE tools did not lead to an increase in HOC feedback. In contrast, the current study revealed that EFL teachers provided more feedback in the hybrid condition than in the teacher-only condition. Notably, hybrid feedback led to an increase in HOC feedback from teachers, while ChatGPT primarily addressed LOCs. A logical explanation for this divergence is the advanced capabilities of ChatGPT as a chatbot compared to earlier AWE tools.

While ChatGPT excelled in providing immediate and accurate corrections for LOCs such as grammar, mechanics, and syntax, its ability to address HOCs was notably limited. This aligns with findings from Jiang et al. (2020) and Chapelle et al. (2015), which suggest

that automated systems often lack the contextual and interpretative capacity to provide meaningful feedback on complex writing aspects. The design of ChatGPT, as reflected in its emphasis on surface-level corrections, prioritizes accuracy in LOCs (Dikli, 2010; Chukharev-Hudilainen & Saricaoglu, 2014). However, as Warschauer and Grimes (2008) emphasized, the inability of AWE systems to engage deeply with argumentation and idea development limits their effectiveness in writing instruction. Interestingly, student perceptions of ChatGPT align with its observed strengths and limitations. As Mohamed (2023) noted, students appreciated ChatGPT's immediacy and accessibility while expressing concerns regarding its ability to provide meaningful feedback on HOCs. This contrast underscores the necessity of a complementary approach, integrating automated tools with teacher expertise to maximize the benefits of both systems.

That said, ChatGPT demonstrated limited effectiveness in addressing HOCs such as content development and coherence. This limitation reflects findings by Shen et al. (2023), which indicated that AI tools lack the contextual understanding and critical reasoning required for evaluating HOCs effectively. Similarly, Wilson and Czik (2016) highlighted the tendency of AI tools to focus on mechanical corrections rather than idea flow or argumentation.

Taken together, these findings align with research by Grimes & Warschauer (2010) and others, emphasizing the complementary yet distinct roles of teacher and AI feedback. Specifically, teachers contribute contextual insights, humanistic evaluation and depth of critique, while AI systems excel at delivering rapid, consistent feedback on surface-level errors. Lastly, while ChatGPT effectively addresses LOCs with speed and consistency, teacher feedback remains essential for fostering writing development, particularly in relation to HOCs. These distinctions reaffirm the potential for a complementary relationship between human and AI feedback systems (Liao, 2016).

On the other hand, teacher feedback was limited by practical challenges, including time demands and variability in feedback quality, as noted by Hattie and Timperley (2007). These limitations sometimes delay the feedback process, which is crucial for fostering writing development. In this context, this study highlighted that integrating AI tools allows teachers to save time and concentrate on higher-level concerns when providing feedback to students.

Nevertheless, teacher feedback was particularly effective in addressing HOCs and fostering improvements in content development, organizational structure and coherence. Ferris (2006) argued that teacher feedback is often more personalized and detailed, allowing students to engage deeply with their writing and improve critical thinking skills. However, the time-intensive nature of teacher feedback, as noted by Truscott (1996), can limit its ability to comprehensively address LOCs, especially when dealing with large numbers of student submissions. Despite these constraints, the role of teachers in addressing HOCs remains undeniable, as their feedback often extends beyond error correction to guide students in developing rhetorical and argumentative skills. These findings also resonate with Ranalli (2018), who observed that students perceive teacher feedback as more credible and impactful than automated alternatives. This perception likely contributes to increased engagement with teacher-provided suggestions, particularly in areas requiring creative and analytical input.

Furthermore, the adjusted mean scores for hybrid feedback were significantly higher than those for teacher-only and ChatGPT-only feedback. This suggests that combining teacher guidance with AI-generated feedback enhances both the quantity and quality of feedback. Such findings also align with Liao (2016), who emphasized the utility of hybrid systems in improving feedback coverage and efficiency without compromising depth.

6.1.3. Distinctiveness of feedback categories in the rubric

The results of this study demonstrate that the rubric criteria are highly distinctive across subcategorized benchmark facets, rater facets, and rater-type facets. This aligns with previous research highlighting the effectiveness of well-defined rubrics in capturing variability in rater tendencies and feedback focus (Barkaoui, 2010; Wilson & Czik, 2016).

The analysis indicates that the criteria in the rubric are significantly distinctive across subcategorized benchmark facets, rater facets and rater-type facets. The multifaceted Rasch measurement model confirmed high separation, strata, and reliability values, underscoring the rubric's ability to differentiate between various feedback conditions with considerable accuracy. This indicates that the rubric is a reliable tool for distinguishing feedback quality and focus across different conditions. Specifically, grammar emerged as the criterion receiving the most feedback, while content received the least. This aligns

with Chapelle et al. (2015), who argued that automated systems like ChatGPT often focus on surface-level issues due to their algorithmic design, while HOCs like idea development and coherence require more detailed human judgment. This pattern was consistent across teacher, AI, and hybrid feedback types, reflecting inherent differences in their feedback tendencies. AI feedback type, for instance, showed a strong preference for addressing grammar and mechanics, whereas teacher and hybrid feedback types distributed their feedback more evenly across HOCs, such as content and organization.

Teachers displayed a clear emphasis on HOCs, such as content and organizational coherence, aligning with Ferris (2006), who emphasized the role of teacher feedback in fostering critical thinking and deeper engagement with student writing. However, as Truscott (1996) observed, time constraints can hinder the ability of teachers to provide equally thorough feedback on LOCs, such as grammar and mechanics. In this study, this limitation became apparent, as teachers demonstrated some inconsistency in addressing LOCs comprehensively. In contrast, AI feedback type, such as ChatGPT, consistently provided detailed feedback on LOCs, particularly grammar and mechanics. Studies by Dikli (2010) and Jiang et al. (2020) have highlighted the efficiency of AWE systems in addressing repetitive surface-level issues. However, these systems often lack the contextual awareness needed to guide students in more complex aspects, such as argumentation and idea organization. This contrast between AI and teachers illustrates the complementary strengths and limitations of each approach.

Hybrid feedback type achieved the most balanced distribution of feedback, excelling across both HOCs and LOCs. The hybrid model's effectiveness reflects its ability to integrate the precision and efficiency of AI tools with the interpretative depth of teacher feedback. This finding aligns with the work of Link et al. (2020) and Wilson and Czik (2016), who highlighted that hybrid feedback systems provide a comprehensive approach to writing instruction by effectively addressing the full range of writing concerns. The interaction analysis between rater types and rubric criteria further supports this observation, as hybrid feedback type consistently outperformed single raters in reliability and coverage. Liao (2016) similarly noted that combining automated tools with teacher feedback enhances the quantity and quality of feedback, making it a sustainable and scalable solution for writing assessments.

The findings of this study underscore the importance of leveraging hybrid feedback systems to address the inherent limitations of single-rater approaches. By balancing the strengths of teacher and AI feedback, hybrid systems can provide a comprehensive evaluation of writing, fostering both learner autonomy and instructional efficiency. These results align with recent studies by Hong (2023) and Kim et al. (2023), which demonstrated that hybrid feedback models improve writing quality and enhance student engagement and motivation by providing timely and meaningful feedback. Moreover, the high reliability and distinctiveness of the rubric criteria validate its effectiveness as a tool for assessing writing across diverse feedback conditions. As suggested by Ferris and Roberts (2001), using well-defined rubrics can promote structured revisions and foster better learning outcomes, particularly when integrated with hybrid feedback systems. Future research should explore how such systems can be further refined to adapt to different learner needs and educational contexts.

6.1.4. Variations in the interaction between feedback types and frequency across feedback categories

The analysis of the interaction between rater types and feedback frequency reveals statistically significant variations across both the main categories (HOCs and LOCs) and their subcategories (content, organization, vocabulary, grammar, and mechanics). Hybrid feedback demonstrated a more balanced distribution of feedback across all categories, while AI feedback concentrated primarily on LOCs, and teachers focused predominantly on HOCs.

AI provided the highest amount of feedback in addressing grammar and mechanics. This mirrors findings from Jiang et al. (2020) and Ranalli et al. (2016), who noted the strength of AWE tools in delivering consistent feedback on grammatical and mechanical accuracy. However, AI feedback's limited capacity to address content and organization highlights its inability to engage with more complex, discourse-level concerns.

In contrast, teachers excelled in providing feedback on HOCs, especially organization and content, reflecting their ability to assess ideas and coherence. This aligns with Ferris (2006), who emphasized the role of human judgment in addressing higher-order concerns. However, the results also revealed a relative insufficiency in their focus on mechanics, as

noted in previous critiques by Truscott (1996), who argued that time constraints can limit the breadth of teacher feedback.

Hybrid feedback emerged as the most comprehensive approach, addressing both HOCs and LOCs with near-equal distribution. This balanced interaction is consistent with studies by Wilson and Czik (2016), who highlighted the advantages of integrating automated feedback with teacher input to provide more holistic support to learners.

The analysis of subcategories revealed further distinctions. Hybrid and teachers performed significantly better, with hybrid feedback particularly excelling in delivering targeted content feedback. In terms of organization, teacher and hybrid feedback demonstrated comparable effectiveness in addressing organizational aspects, with teacher feedback slightly outperforming in terms of depth. This supports Ferris and Roberts (2001), who noted the importance of teacher feedback in guiding students' rhetorical and structural development. Regarding grammar and mechanics, AI rater dominated these subcategories, reflecting their algorithmic design for surface-level error correction. Hybrid feedback, however, managed to provide sufficient feedback in these areas without compromising their focus on HOCs, which aligns with the findings of Link et al. (2020).

The findings highlight the significance of utilizing the unique strengths of various rater types to meet the diverse requirements of students. AI tools effectively provide immediate and consistent feedback on LOCs, thereby reducing teachers' workloads. However, their limitations in addressing HOCs necessitate teacher involvement. Hybrid feedback systems, which integrate AI precision with human contextual understanding, represent a promising approach to comprehensive writing instruction.

Overall, the statistical significance of the interactions between rater types and feedback categories highlights the need for educators to adopt strategies that maximize the strengths of both AI and teachers. Training programs for teachers should emphasize integrating AI tools like ChatGPT to enhance feedback efficiency while focusing on higher-order concerns.

6.2. EFL Teachers' Perceptions

The findings of this study provided a detailed understanding of how EFL teachers perceive the three feedback types in relation to their strengths, challenges and potential for improvement. These insights reveal the practicality and pedagogical implications of adopting teacher, AI and hybrid feedback in the EFL context.

6.2.1. Strengths and challenges in feedback categories across feedback types

The qualitative findings provided deeper insights into teachers' perspectives on the strengths and challenges of teacher, AI and hybrid feedback.

Regarding strengths, teachers identified their role as essential for improving HOCs, such as content clarity, argument development, and idea organization. Teachers favored personalized feedback, and this aligns with the literature as personalized teacher feedback fosters a deeper understanding of writing concepts (Ferris, 2006) and ensures targeted revisions to address specific weaknesses (Ellis et al., 2009). Teachers also valued ChatGPT's ability to identify LOCs like grammatical and mechanical errors efficiently and instantly. These results agree with those obtained by Dikli (2010) and Ranalli et al. (2016). EFL teachers viewed hybrid feedback as comprehensive and time-efficient. These are consistent with those of Wilson & Czik (2016) and Link et al. (2020), who stated that by delegating surface-level issues to ChatGPT, teachers could focus on improving students' higher-order writing skills.

Additionally, teachers stated that AI technologies in language learning can enhance the ecological perspective of teaching and offer personalized and interactive learning experiences. Teachers also stated that by providing real-time responses to questions of students, AI tools can create a more personalized learning atmosphere for students. These results match those observed in earlier studies (Haristiani, 2019; He, 2023; Holmes & Tuomi, 2022; Hong & Zhang, 2023; Lin & Lin, 2023).

EFL teachers also stated that chatbots can provide personalized feedback, consequently improving their writing skills. EFL teachers mostly favored the AI tool's accessibility and ease of use when providing quick feedback, as well as the AI tool's ability to lessen their

workload and save time. These results are consistent with the findings of a great deal of the previous work (Hew et al., 2022; Kohnke, 2023; Rahman & Watanobe, 2023; Qadir, 2023; Quasem et al., 2023; Su & Yang, 2023). Similarly, Hashem et al. (2023) stated that AI tools can reduce teachers' workload and facilitate the tracking of student development, and these findings were consistent with this current study.

Similarly, Mohamed (2023) and Nguyen (2023) interviewed teachers who viewed ChatGPT as an affordable and convenient tool for providing feedback. Furthermore, Han et al. (2023) and Schmidt-Fajlik (2023) reported positive student experiences with ChatGPT's feedback through questionnaire surveys. However, these studies rely on self-reported outcomes and lack statistical evidence. Thus, this study combined quantitative analysis with qualitative analysis to complement each other.

Feedback provided by ChatGPT was investigated in terms of teaching and learning. Although ChatGPT's output may not always be informative or original, ChatGPT can provide teachers with helpful feedback for improving classroom instruction (Chang et al., 2024). Dai et al. (2023) also stated in their study that ChatGPT's feedback is typically more thorough and consistent for students, especially regarding student reports. Thus, EFL teachers in this study favor the hybrid feedback model the most. Similarly, Wang and Han (2022) explored the potential of integrating ChatGPT with teacher feedback and suggested the two sources of feedback types can complement each other effectively. In accordance with the present results, participants in their study believed that while ChatGPT can address LOCs and provide initial feedback, teachers can then focus on HOCs, such as content and organization, and offer more personalized guidance (Wang & Han, 2022).

Teachers acknowledged ChatGPT as a practical tool for addressing lower-order concerns such as grammar, vocabulary, and punctuation errors. Teachers favored its efficiency, immediacy, and consistent feedback as major strengths. These qualities made it particularly valuable for reducing their workload and providing quick corrections that students could act on independently. Such strengths align with findings in the literature (Ranalli et al., 2016; Wilson & Czik, 2016), emphasizing AI-driven systems' ability to enhance linguistic accuracy. However, the limitations of ChatGPT in addressing higher-order concerns, such as content development, coherence, and organization, emerged as a

significant challenge. Teachers observed that ChatGPT often failed to interpret the detailed aspects of academic writing or provide context-sensitive suggestions. This limitation is similarly highlighted by Link et al. (2020). Furthermore, occasional inaccuracies or irrelevant recommendations undermined its reliability in handling more complex writing tasks. To address these limitations, teachers proposed advancements in ChatGPT's algorithms to enable more accurate evaluation of HOCs, as well as the incorporation of context-specific knowledge and rhetorical analysis to enhance its contextual sensitivity.

Participant teachers stated that chatbots' ability to provide meaningful feedback according to context is limited, and it sometimes can provide misinformation and improvised references. When engaging in complex conversations, sometimes it can suggest irrelevant corrections as well. Similarly, Yang et al. (2022) underlined these concerns in their study, as AI chatbots often lack the process of natural language and provide context-specific responses. In accordance with the present results, previous studies have demonstrated similar concerns regarding misinformation and improvised references (Annamalai et al., 2023; Choi et al., 2023; Fergus et al., 2023; Sun & Hoelscher, 2023; Tlili et al., 2023), academic integrity (Khan et al., 2023; Su & Yang, 2023). In addition, EFL teachers had concerns regarding the lack of ChatGPT's human-like understanding of context-specific feedback, which ultimately affects students' language-learning process. These results are consistent with data obtained in the previous studies (Kohnke et al., 2023; Koubaa et al., 2023).

Teacher feedback, by contrast, was consistently identified as the most effective for addressing HOCs. Teachers' ability to provide personalized, detailed and context-specific suggestions allowed them to guide students in developing arguments, improving structure and refining rhetorical strategies. Such personalized input is well-supported in the literature, with Hyland (2019) emphasizing the role of teacher feedback in fostering critical engagement and deep revisions. However, challenges related to workload and time constraints were undeniable. Teachers highlighted the difficulty of balancing the need for detailed feedback with the pressures of large class sizes and diverse student needs, which sometimes resulted in delays or inconsistencies in feedback delivery. These results are in accord with prior research (Hattie & Timperley, 2007; Hyland & Hyland, 2006), indicating the tension between feedback quality and teacher capacity. Suggestions

for improvement included institutional support, such as reduced teaching loads and professional development opportunities to enhance feedback techniques. Teachers also recommended leveraging technological tools to foster the feedback process, particularly for LOCs, which saves time for them to focus more on HOCs.

Conversely, ChatGPT was viewed as efficient for LOCs yet limited in addressing HOCs. Teachers appreciated ChatGPT's efficiency in identifying and correcting LOCs, particularly grammar and spelling errors, which helped teachers save significant time. These perceptions support the findings of Liao (2016) and Dikli (2010), who emphasized the speed and accuracy of AWE systems for surface-level issues. As a limitation, AI tools like ChatGPT struggle to address contextual and global concerns, particularly content generation and idea development. Teachers noted that AI feedback often lacked depth and could not offer meaningful suggestions for improving argumentation or coherence, which is similar to the findings of Shen et al. (2023).

The hybrid feedback model was regarded as the most comprehensive approach, combining the strengths of ChatGPT and teacher feedback to address both HOCs and LOCs effectively. Teachers praised the efficiency of ChatGPT in handling surface-level corrections, freeing them to focus on higher-order revisions. This integration of automated and teacher feedback created a balanced feedback environment, fostering both linguistic accuracy and critical thinking in student writing. Similar benefits have been observed in previous studies, such as Chen and Cheng (2008), which highlighted the complementary roles of AI tools and human expertise in writing instruction. However, teachers also noted challenges in aligning ChatGPT-generated feedback with their own pedagogical objectives. Ensuring consistency and coherence between the two feedback sources often required additional effort, reflecting findings by Jiang et al. (2020) regarding the need for seamless integration of AI and teacher feedback. To enhance the hybrid model, teachers suggested the development of platforms that unify automated and teacher-generated feedback, providing greater control and alignment with instructional goals. They also emphasized the importance of training programs to equip teachers with the skills needed to effectively implement hybrid feedback in their teaching practices. Teachers suggested refining AI tools to improve their feedback on content and organization while maintaining their accuracy for grammar and mechanics. They recommended professional development programs to train teachers on effectively

integrating AI-based tools like ChatGPT into writing instruction, similar to the findings of Güneyli et al. (2024).

For hybrid feedback, teachers emphasized the need for further improvements in AI capabilities, particularly in enhancing cohesion, logical development and vocabulary use. Teachers favored hybrid feedback for its comprehensive nature, as it allowed them to focus on higher-order writing skills while delegating grammar and mechanical corrections to AI. This validates the findings of Wang and Han (2022), who found hybrid approaches to be the most effective for managing both HOCs and LOCs. While teachers appreciated hybrid feedback, they stressed the need for careful coordination and quality control between AI tools and teacher input to ensure effectiveness.

In summary, while each feedback type offers unique strengths and faces distinct challenges, the hybrid model emerges as a practical and effective solution for addressing the diverse needs of EFL writing instruction. By leveraging the complementary capabilities of ChatGPT and teacher feedback, the hybrid approach provides a balanced framework for supporting both linguistic and rhetorical development.

6.2.2. Suggestions for enhancing teacher, ChatGPT, and hybrid feedback

Teachers proposed several strategies to improve feedback types to better support students' writing development and emphasized the need to address both LOCs and HOCs more effectively. While ChatGPT excels in handling LOCs such as grammar and spelling, its limited ability to provide meaningful feedback on HOCs, including argument development, coherence, and organization, remains a critical limitation (Nugroho et al., 2023; Shaikh et al., 2023). This supports the findings of Warschauer and Grimes (2008), who argued that while automated tools are effective for addressing surface-level errors, they still require teacher intervention to guide students through more complex aspects of writing. Teachers recommended enhancing ChatGPT's natural language processing and contextual analysis capabilities to generate feedback on thesis statements, argument flow, and paragraph transitions. Such refinements would align with the suggestions of Harvey-Scholes (2017) and Liao (2016), who highlighted the importance of incorporating critical thinking elements into AI-driven feedback systems.

In comparison, teacher feedback continues to be regarded as more targeted and detailed, particularly in addressing HOCs (Ferris, 2006; Truscott, 1996). Teachers proposed the use of structured rubrics prioritizing HOCs to ensure their efforts focus on the areas where students most need support. This reflects Ferris' (1999) perspective that effective feedback requires balancing detailed commentary with actionable suggestions to sustain student motivation. For example, teachers suggested incorporating templates, guiding questions, or examples within feedback to help students address specific weaknesses more effectively. The inclusion of targeted feedback strategies would also allow students to engage more deeply with revision processes which aligns with the findings of Ferris and Roberts (2001) that selective and focused feedback is more effective for student learning outcomes.

In addition, professional development for teachers was identified as a critical need to ensure the effective integration of AI tools like ChatGPT into classroom practices. Teachers emphasized the value of workshops and training programs to improve their ability to manage hybrid feedback systems effectively. This is consistent with the findings of Dincer (2018) and Güneyli et al. (2024), who emphasize the critical need for enabling educators to develop the skills required to effectively utilize AI-assisted feedback systems while maintaining pedagogical standards. Such training would also address the challenges associated with balancing ChatGPT-generated feedback with instructional objectives, as discussed by Jiang et al. (2020).

The hybrid feedback model emerged as a particularly promising approach leveraging the strengths of both ChatGPT and teacher feedback to create a comprehensive and balanced feedback system. In this model, ChatGPT efficiently handles LOCs, allowing teachers to focus on HOCs, thereby enhancing the depth and quality of writing feedback. This aligns with Wilson and Czik (2016) and Link et al. (2020), who noted that combining automated and teacher feedback improves writing instruction by enabling teachers to concentrate on higher-order writing skills while AI tools address surface-level issues. Teachers recommended clearly defining the roles of ChatGPT and themselves to ensure that both feedback sources work together effectively without redundancy or conflict. They also recommended conducting regular quality checks to ensure consistent and reliable feedback delivery, as highlighted by Liao (2016) and Harvey-Scholes (2017).

However, challenges in managing hybrid feedback were also noted. One major concern was the overwhelming volume of feedback generated by ChatGPT, which could confuse students and hinder effective revisions. To address this, teachers recommended limiting ChatGPT's feedback to the most critical errors and providing a concise feedback summary to help students prioritize their revisions. This approach reflects the findings of Ferris and Roberts (2001), who demonstrated that selective and focused feedback leads to better student engagement and more meaningful revisions.

Overall, this study highlighted the complementary roles of ChatGPT and teacher feedback in writing instruction. While ChatGPT excels in addressing surface-level errors, teacher feedback remains essential for fostering higher-order writing skills, such as argument development, coherence, and organization. The hybrid feedback model combines these strengths, offering an efficient and impactful solution to support both LOCs and HOCs. Teachers' insights highlight the importance of improving AI tools to better address content-related issues and prioritizing professional development to optimize hybrid feedback systems. Future research should explore the long-term effectiveness of hybrid feedback, with particular attention to overcoming challenges related to ethical considerations, AI refinement, and teacher training.

7. CONCLUSION

This study investigated the comparative effectiveness of teacher, AI and hybrid feedback on tertiary-level EFL students' writing development, focusing on higher-order concerns such as content and organization and lower-order concerns such as grammar and mechanics as well as EFL teachers' perceptions regarding the strengths, challenges and suggestions for these feedback types. The findings offer valuable insights into the unique contributions of each feedback type, their limitations and the potential of integrating artificial intelligence with traditional pedagogical practices to enhance EFL writing instruction.

The results revealed that ChatGPT excels in addressing LOCs by providing quick, consistent and detailed grammatical corrections. However, its limited ability to address HOCs, including content and coherence, highlights its reliance on predefined structures and its lack of contextual understanding. Teacher feedback, on the other hand, demonstrated a comprehensive ability while addressing HOCs. Teachers were particularly effective in guiding students in content development and improving overall writing coherence, which validates the irreplaceable role of human judgment in written feedback.

Hybrid feedback, combining ChatGPT's efficiency with teacher expertise, was found to be the most effective in addressing both HOCs and LOCs. ChatGPT handled LOCs efficiently, like mechanical errors, which allowed teachers to deliver more time to critical writing concerns. This approach not only reduced teachers' workload but also provided students with more holistic and impactful feedback. The findings highlighted the value of hybrid systems in maximizing the potential of AI tools without compromising the quality and depth of teacher feedback.

This study contributes to the growing body of research on integrating artificial intelligence into EFL writing instruction, emphasizing the importance of balancing automation with teachers' expertise. The findings suggest that while ChatGPT and teacher feedback serve distinct purposes, their combination in the hybrid model offers a promising solution to the challenges of EFL writing development. By addressing the

identified challenges and building on the strengths of these feedback types, educators can better support EFL learners in achieving their writing goals.



8. SUGGESTIONS FOR FURTHER STUDIES

While this study provides valuable insights regarding the feedback nature of teacher, AI and hybrid in EFL writing instruction, several areas require further exploration.

First, longitudinal studies could be conducted to assess the long-term impact of hybrid feedback on students' writing development. Investigating whether the constant use of hybrid feedback models leads to measurable improvements in both higher-order and lower-order concerns over time would provide critical evidence for its pedagogical value.

Second, future research could expand the scope of this study by examining student perceptions of teacher, AI and hybrid feedback. Understanding how students perceive and engage with these feedback types could reveal important aspects of learner motivation, satisfaction and writing autonomy.

Third, further research should address how hybrid feedback can be effectively implemented in classroom settings. Future research could explore strategies for effectively integrating AI and teacher input to minimize redundancy and optimize the combined benefits of both approaches.

Lastly, given the ethical and technical limitations of AI tools, such as data privacy concerns and contextual inaccuracies, future studies should explore ways to refine AI systems to provide more contextually relevant feedback. Interdisciplinary research involving educators, linguists and AI developers could result in the development of tools that effectively address the needs of both language learners and educators.

Exploring these areas will enhance our understanding of feedback systems in EFL writing instruction and promote the development of innovative methods for improving writing proficiency across diverse educational contexts.

REFERENCES

- Agyei, D. D., & Voogt, J. M. (2011). Exploring the potential of the will, skill, tool model in Ghana: Predicting prospective and practicing teachers' use of technology. *Computers & Education*, 56(1), 91–100. <https://doi.org/10.1016/j.compedu.2010.08.017>
- Ahangari, S., Hejazi, M., & Razmjou, L. (2014). The impact of scaffolding on content retention of Iranian post-elementary EFL learners' summary writing. *Procedia - Social and Behavioral Sciences*, 98, 83–89. <https://doi.org/10.1016/j.sbspro.2014.03.392>
- Akgun, S., & Greenhow, C. (2021). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI And Ethics*, 2(3), 431–440. <https://doi.org/10.1007/s43681-021-00096-7>
- Aktay, S., Gök, S., & Uzunoğlu, D. (2023). ChatGPT in education. *Türk Akademik Yayınlar Dergisi (TAY Journal)*, 7(2), 378–406. <https://dergipark.org.tr/en/pub/tayjournal/issue/76658/1259832>
- Al-Abdullatif, A. M., & Alsubaie, M. A. (2022). Using digital learning platforms for teaching arabic literacy: A post-pandemic mobile learning scenario in Saudi Arabia. *Sustainability*, 14(19), 11868. <https://doi.org/10.3390/su141911868>
- AlAfnan, M. A., Dishari, N. S., Jovic, N. M., & Lomidze, N. K. (2023). ChatGPT as an educational tool: opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology*. <https://doi.org/10.37965/jait.2023.0184>
- Alier, M., García-Peñalvo, F., & Camba, J. D. (2024). Generative artificial intelligence in education: From deceptive to disruptive. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5), 5. <https://doi.org/10.9781/ijimai.2024.02.011>

- Alkan, A. (2024). Artificial intelligence: Its role and potential in education. *İnsan Ve Toplum Bilimleri Araştırmaları Dergisi*, 13(1), 483–497. <https://doi.org/10.15869/itobiad.1331201>
- Allehyani, S. H., & Algamdi, M. A. (2023). Digital competences: Early childhood teachers' beliefs and perceptions of ChatGPT application in teaching English as a second language (ESL). *International Journal of Learning Teaching and Educational Research*, 22(11), 343–363. <https://doi.org/10.26803/ijlter.22.11.18>
- Andrei, E. (2016). Technology in teaching English language learners: The case of three middle school teachers. *TESOL Journal*, 8(2), 409–431. <https://doi.org/10.1002/tesj.280>
- Annamalai, N., Zakarneh, B., Mangaleswaran, S., & Zyoud, S. H. (2023). Exploration of undergraduate students lived experiences of using a chatbot for English language learning. *International Journal of English Language and Literature Studies*, 12(3), 254–268. <https://doi.org/10.55493/5019.v12i3.4876>
- Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom. *Advances in Mobile Learning Educational Research*, 3(2), 818–824. <https://doi.org/10.25082/amler.2023.02.009>
- Ayala-Pazmiño, M. (2023). Artificial intelligence in Education: Exploring the potential benefits and risks. *593 Digital Publisher CEIT*, 8(3), 892–899. <https://doi.org/10.33386/593dp.2023.3.1827>
- Baehr, J. (2013). Educating for intellectual virtues: From theory to practice. *Journal of Philosophy of Education*, 47(2), 248–262. <https://doi.org/10.1111/1467-9752.12023>

- Bai, B., Wang, J., & Nie, Y. (2020). Self-efficacy, task values and growth mindset: what has the most predictive power for primary school students' self-regulated learning in English writing and writing competence in an Asian Confucian cultural context? *Cambridge Journal of Education*, 51(1), 65–84.
<https://doi.org/10.1080/0305764x.2020.1778639>
- Bii, P. K., Too, J. K., & Mukwa, C. W. (2018). Teacher attitude towards use of chatbots in routine teaching. *Universal Journal of Educational Research*, 6(7), 1586-1597.
- Bloom, B. S. (1974). *Taxonomy of educational objectives: The Classification of Educational Goals*.
- Bond, T., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge. <https://doi.org/10.4324/9781315814698>
- Bozkurt, A., Karadeniz, A., Banares, D., Guerrero-Roldán, A. E., & Rodríguez, M. E. (2021). Artificial intelligence and reflections from educational landscape: A review of AI studies in half a century. *Sustainability*, 13(2), 800.
<https://doi.org/10.3390/su13020800>
- Brennan, K. (2006). The managed teacher: Emotional labour, education, and technology *Educational Insights*, 10(2), 55–65.
<http://einsights.ogpr.educ.ubc.ca/v10n02/pdfs/brennan.pdf>
- Buitrago, C. R., & Díaz, J. (2017). Flipping your writing lessons: Optimizing time in your EFL writing classroom. In *Springer eBooks* (pp. 69–91).
https://doi.org/10.1007/978-981-10-6968-0_6
- Çam, M. B., Çelik, N. C., Güntepe, E. T., & Durukan, Ü. G. (2021). Öğretmen adaylarının yapay zekâ teknolojileri ile ilgili farkındalıklarının belirlenmesi. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 18(48), 263–285.
<https://dergipark.org.tr/en/download/article-file/1764677>

- Çayır, A. (2023). A literature review on the effect of artificial intelligence on education. *Journal of Human and Social Sciences*, 6(2), 276–288. <https://doi.org/10.53048/johass.1375684>
- Cen, Y., & Ross, H. (2009). Chinese higher education and evaluation in context. *Chinese Education & Society*, 42(1), 3–7. <https://doi.org/10.2753/ced1061-1932420100>
- Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3), 1–45. <https://doi.org/10.1145/3641289>
- Chapelle, C. A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, 32(3), 385–405. <https://doi.org/10.1177/0265532214565386>
- Chen, C. E., & Cheng, W. E. C. (2008). Beyond the design of automated writing evaluation: pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12(2), 94–112. <http://elc.msu.edu/llt/vol12num2/chencheng.pdf>
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, 20(4), 328–338. <https://doi.org/10.1016/j.learninstruc.2009.08.006>
- Choi, E. P. H., Lee, J. J., Ho, M., Kwok, J. Y. Y., & Lok, K. Y. W. (2023). Chatting or cheating? The impacts of ChatGPT and other artificial intelligence language models on nurse education. *Nurse Education Today*, 125, 105796. <https://doi.org/10.1016/j.nedt.2023.105796>

- Choi, J. (2010). *The impact of Automated Essay scoring (AES) for improving English language learner's essay writing* (pp. 1–208) [Doctoral dissertation, University of Virginia]. <https://www.learntechlib.org/p/116200/>
- Chuah, K., & Kabilan, M. K. (2021). Teachers' views on the use of chatbots to support English language teaching in a mobile environment. *International Journal of Emerging Technologies in Learning (iJET)*, 16(20), 223. <https://doi.org/10.3991/ijet.v16i20.24917>
- Chukharev-Hudilainen, E., & Saricaoglu, A. (2014). Causal discourse analyzer: Improving automated feedback on academic ESL writing. *Computer Assisted Language Learning*, 29(3), 494–516. <https://doi.org/10.1080/09588221.2014.991795>
- Clarizia, F., Colace, F., Lombardi, M., Pascale, F., & Santaniello, D. (2018). Chatbot: an education support system for student. In *Lecture notes in computer science* (pp. 291–302). https://doi.org/10.1007/978-3-030-01689-0_23
- Copulsky, J. (2019). Do conversational platforms represent the next big digital marketing opportunity? *Applied Marketing Analytics*. <https://doi.org/10.69554/hiur2549>
- Cotos, E. (2011). Potential of automated writing evaluation feedback. *CALICO Journal*, 28(2), 420–459. <https://doi.org/10.11139/cj.28.2.420-459>
- Crawford, J., Cowling, M., & Allen, K. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching and Learning Practice*, 20(3). <https://doi.org/10.53761/1.20.3.02>
- Creswell, J. W. (2014). *Research design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE.

- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. SAGE Publications.
- Crompton, H., & Song, D. (2021). The potential of artificial intelligence in higher education. *Revista Virtual Universidad Católica Del Norte*, 62, 1–4.
<https://doi.org/10.35575/rvucn.n62a1>
- Cui, Y., & Li, H. (2022). Evaluation system of mobile English learning platform by using deep learning algorithm. *Mobile Information Systems*, 2022, 1–9.
<https://doi.org/10.1155/2022/3849079>
- Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication*, 7(4), 482–511.
<https://doi.org/10.1177/0741088390007004003>
- Dai, Y., Liu, A., & Lim, C. P. (2023). Reconceptualizing ChatGPT and generative AI as a student-driven innovation in higher education. *Procedia CIRP*, 119, 84–90.
<https://doi.org/10.1016/j.procir.2023.05.002>
- Darmaji, N., Astalini, N., Kurniawan, D. A., Minarsih, N., & Perdana, R. (2023). The positive impact of digital literacy on science process skills. *Jurnal Pendidikan Dan Pengajaran*, 56(2), 241–250. <https://doi.org/10.23887/jpp.v56i2.58032>
- Deng, J., & Lin, Y. (2023). The benefits and challenges of ChatGPT: An overview. *Frontiers in Computing and Intelligent Systems*, 2(2), 81–83.
<https://doi.org/10.54097/fcis.v2i2.4465>
- Dewi, F., Lengkanawati, N. S., & Purnawarman, P. (2019). Teachers' consideration in technology-integrated lesson design – A case of Indonesian EFL teachers. *International Journal of Emerging Technologies in Learning (iJET)*, 14(18), 92.
<https://doi.org/10.3991/ijet.v14i18.9806>

- Dikli, S. (2010). The nature of automated essay scoring feedback. *CALICO Journal*, 28(1), 99–134. <https://doi.org/10.11139/cj.28.1.99-134>
- Dikli, S., & Bleyle, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, 22, 1–17. <https://doi.org/10.1016/j.asw.2014.03.006>
- Dinçer, S. (2018). Are preservice teachers really literate enough to integrate technology in their classroom practice? Determining the technology literacy level of preservice teachers. *Education and Information Technologies*, 23(6), 2699–2718. <https://doi.org/10.1007/s10639-018-9737-z>
- Dörnyei, Z., & Griffee, D. T. (2010). Research methods in applied linguistics. *TESOL Journal*, 1(1), 181–183. <https://doi.org/10.5054/tj.2010.215611>
- Dung, P. T. K. (2020). Teachers’ written feedback: How to make it work more effectively in a language classroom? *BVNU Journal of Foreign Studies*, 36(3). <https://doi.org/10.25073/2525-2445/vnufs.4554>
- Durall, E., & Kapros, E. (2020). Co-design for a competency self-assessment chatbot and survey in science education. In *Lecture notes in computer science* (pp. 13–24). https://doi.org/10.1007/978-3-030-50506-6_2
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., . . . Wright, R. (2023). Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>

- Ellis, R., Loewen, S., Elder, C., Reinders, H., Erlam, R., & Philp, J. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Multilingual Matters.
- 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>
- Farrokhi, F., Esfandiari, R., & Schaefer, E. (2012). A many-facet Rasch measurement of differential rater severity/leniency in three types of assessment. *JALT Journal*, 34(1), 79-101.
- Farrokhi, F., Esfandiari, R., & Vaez Dalili, M. (2011). Applying the many-facet Rasch model to detect centrality in self-assessment, peer-assessment and teacher assessment. *World Applied Sciences Journal*, 15(11), 76-83.
- Fergus, S., Botha, M., & Ostovar, M. (2023). Evaluating academic answers generated using ChatGPT. *Journal of Chemical Education*, 100(4), 1672–1675. <https://doi.org/10.1021/acs.jchemed.3c00087>
- Ferris, D. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly*, 29(1), 33. <https://doi.org/10.2307/3587804>
- Ferris, D. (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](https://doi.org/10.1016/s1060-3743(99)80110-6)
- Ferris, D. (2003). *Response to student writing: Implications for second language students*. <http://ci.nii.ac.jp/ncid/BA62000794>
- Ferris, D. (2004). The “grammar correction” debate in L2 writing: Where are we, and where do we go from here? (And what do we do in the meantime . . .?). *Journal of Second Language Writing*, 13(1), 49–62. <https://doi.org/10.1016/j.jslw.2004.04.005>

- Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In *Cambridge University Press eBooks* (pp. 81–104). <https://doi.org/10.1017/cbo9781139524742.007>
- Ferris, D., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing*, 22(3), 307–329. <https://doi.org/10.1016/j.jslw.2012.09.009>
- Ferris, D., Pezone, S., Tade, C. R., & Tinti, S. (1997). Teacher commentary on student writing: Descriptions & implications. *Journal of Second Language Writing*, 6(2), 155–182. [https://doi.org/10.1016/s1060-3743\(97\)90032-1](https://doi.org/10.1016/s1060-3743(97)90032-1)
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes. *Journal of Second Language Writing*, 10(3), 161–184. [https://doi.org/10.1016/s1060-3743\(01\)00039-x](https://doi.org/10.1016/s1060-3743(01)00039-x)
- Foltz, P. W., Yan, D., & Rupp, A. A. (2020). The past, present, and future of automated scoring. In *Chapman and Hall/CRC eBooks* (pp. 1–10). <https://doi.org/10.1201/9781351264808-1>
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and Evaluate research in Education*. McGraw-hill.
- Furqoni, N. I. (2019). The effect of peer feedback technique in teaching writing on students' writing achievement of Writing II course at English Department FKIP Lambung Mangkurat University. *Journal of English Teaching Applied Linguistics and Literatures (JETALL)*, 2(1), 28. <https://doi.org/10.20527/jetall.v2i1.7374>
- Geckin, V. (2020). Yabancı dil olarak ingilizce öğretilen ortamda çok taslaklı kompozisyonlara verilen yazılı düzeltici geribildirime öğrenci tepkileri. *Bartın University Journal of Faculty of Education*, 9(3), 577–597. <https://doi.org/10.14686/buefad.685527>

- Gikandi, J., Morrow, D., & Davis, N. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, 57(4), 2333–2351. <https://doi.org/10.1016/j.compedu.2011.06.004>
- Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9, e45312. <https://doi.org/10.2196/45312>
- Gleason, N. (2022). ChatGPT and the rise of AI writers: How should higher education respond. *Times Higher Education*. <https://www.timeshighereducation.com/campus/chatgpt-and-rise-ai-writers-how-should-higher-education-respond>
- Gocen, A., & Aydemir, F. (2020). Artificial intelligence in education and schools. *Research on Education and Media*, 12(1), 13–21. <https://doi.org/10.2478/rem-2020-0003>
- González-Lloret, M. (2016). Technology and task-based language teaching. In *Springer eBooks* (pp. 1–13). https://doi.org/10.1007/978-3-319-02328-1_16-1
- Graham, S., MacArthur, C. A., & Fitzgerald, J. (2013). *Best practices in writing instruction*. Guilford Press.
- Grainger, P. R., Purnell, K., & Zipf, R. (2008). Judging quality through substantive conversations between markers. *Assessment & Evaluation in Higher Education*, 33(2), 133–142. <https://doi.org/10.1080/02602930601125681>
- Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: a multi-site case study of automated writing evaluation. *Journal of Technology, Learning, and Assessment*, 8(6), 1–43. <http://files.eric.ed.gov/fulltext/EJ882522.pdf>

- Güneyli, A., Burgul, N. S., Dericioğlu, S., Cenkova, N., Becan, S., Şimşek, Ş. E., & Güneralp, H. (2024). Exploring teacher awareness of artificial intelligence in education: a case study from Northern Cyprus. *European Journal of Investigation in Health Psychology and Education*, 14(8), 2358–2376. <https://doi.org/10.3390/ejihpe14080156>
- Hamid, H. A., & Romly, R. (2021). Teachers' perception on giving feedback to students' online writing assignment during Movement Control Order (MCO). *Advances in Social Science, Education and Humanities Research/Advances in Social Science, Education and Humanities Research*. <https://doi.org/10.2991/assehr.k.210312.076>
- Haristiani, N. (2019). Artificial intelligence (AI) Chatbot as language learning medium: An inquiry. *Journal of Physics Conference Series*, 1387(1), 012020. <https://doi.org/10.1088/1742-6596/1387/1/012020>
- Harmer, J. (2006). *How to teach writing*. Pearson Education India.
- Harvey-Scholes, C. (2017). Computer-assisted detection of 90% of EFL student errors. *Computer Assisted Language Learning*, 31(1–2), 144–156. <https://doi.org/10.1080/09588221.2017.1392322>
- Haseski, H. I. (2019). What do Turkish pre-service teachers think about artificial intelligence? *International Journal of Computer Science Education in Schools*, 3(2), 3–23. <https://doi.org/10.21585/ijcses.v3i2.55>
- Hashem, R., Ali, N., Zein, F. E., Fidalgo, P., & Khurma, O. A. (2023). AI to the rescue: Exploring the potential of ChatGPT as a teacher ally for workload relief and burnout prevention. *Research and Practice in Technology Enhanced Learning*, 19, 023. <https://doi.org/10.58459/rptel.2024.19023>

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hew, K. F., Huang, W., Du, J., & Jia, C. (2022). Using chatbots to support student goal setting and social presence in fully online activities: learner engagement and perceptions. *Journal of Computing in Higher Education*, 35(1), 40–68. <https://doi.org/10.1007/s12528-022-09338-x>
- Hibbert, K. M., Van Deven, T., & Ros, S. (2012). Fundamentals of assessment and evaluation: Clarifying terminology. In *Springer eBooks* (pp. 11–19). https://doi.org/10.1007/978-3-642-27600-2_2
- Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. *Proceedings of the National Academy of Sciences*, 115(27), 6631–6632.
- Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542–570. <https://doi.org/10.1111/ejed.12533>
- Hong, Y., & Zhang, J. (2023). An attempt of construction of English language intelligent flipped classroom teaching mode from an ecological perspective. *Asia-Pacific Journal of Humanities and Social Sciences*, 3(1), 195–208. <https://doi.org/10.53789/j.1653-0465.2023.0301.022>
- Hu, L., & Wang, Y. (2023). The predicting role of EFL teachers' immediacy behaviors in students' willingness to communicate and academic engagement. *BMC Psychology*, 11(1). <https://doi.org/10.1186/s40359-023-01378-x>
- Hwang, G., & Chen, N. (2023). Editorial position paper: Exploring the potential of generative artificial intelligence in education: Applications, challenges, and future research directions. *Educational Technology & Society*, 26(2), 1–18. [https://doi.org/10.30191/ETS.202304_26\(2\).0014](https://doi.org/10.30191/ETS.202304_26(2).0014)

- Hyland, K. (2019). *Second language writing*. Cambridge University Press.
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83–101. <https://doi.org/10.1017/s0261444806003399>
- Idham, A. Z., Rauf, W., & Rajab, A. (2024). Navigating the transformative impact of artificial intelligence on English language teaching: Exploring challenges and opportunities. *Jurnal Edukasi Saintifik*, 4(1), 8–14. <https://doi.org/10.56185/jes.v4i1.620>
- Ioannou-Georgiou, S., & Pavlou, P. (2007). *Assessing young learners*. Oxford University Press.
- Javier, D. R. C., & Moorhouse, B. L. (2023). Developing secondary school English language learners' productive and critical use of ChatGPT. *TESOL Journal*, 15(2). <https://doi.org/10.1002/tesj.755>
- Jiang, L., Yu, S., & Wang, C. (2020). Second language writing instructors' feedback practice in response to automated writing evaluation: A sociocultural perspective. *System*, 93, 102302. <https://doi.org/10.1016/j.system.2020.102302>
- Jiang, X., & Wang, D. (2023). Enhancing journal reputation and academic socialization: Review feedback matters beyond its gatekeeping function. *Learned Publishing*, 36(4), 506–516. <https://doi.org/10.1002/leap.1577>
- Kafyulilo, A., Fisser, P., Pieters, J., & Voogt, J. (2015). ICT use in science and mathematics teacher education in Tanzania: Developing technological pedagogical content knowledge. *Australasian Journal of Educational Technology*, 31(4). <https://doi.org/10.14742/ajet.1240>

- Karabıyık, Ü. (2024). Matematik eğitiminde yenilikçi bir yaklaşım: ChatGPT'nin rolü. *Uşak Üniversitesi Eğitim Araştırmaları Dergisi*, 10(1), 26–46. <https://doi.org/10.29065/usakead.1393487>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kassim, A.N.L. (2007). *Exploring rater judging behaviour using the many-facet Rasch model*. Paper Presented in the Second Biennial International Conference on Teaching and Learning of English in Asia: Exploring New Frontiers (TELiA2), Universiti Utara, Malaysia.
- Khan, R. A., Jawaid, M., Khan, A. R., & Sajjad, M. (2023). ChatGPT - Reshaping medical education and clinical management. *Pakistan Journal of Medical Sciences*, 39(2). <https://doi.org/10.12669/pjms.39.2.7653>
- Kim, N. (2019). A study on the use of artificial intelligence chatbots for improving English grammar skills. *Journal of Digital Convergence*, 17(8), 37–46. <https://doi.org/10.14400/jdc.2019.17.8.037>
- Kim, S., Shim, J., & Shim, J. (2023). A study on the utilization of OpenAI ChatGPT as a second language learning tool. *Journal of Multimedia Information System*, 10(1), 79–88. <https://doi.org/10.33851/jmis.2023.10.1.79>
- Kohnke, L. (2023). L2 learners' perceptions of a chatbot as a potential independent language learning tool. *International Journal of Mobile Learning and Organisation*, 17(1/2), 214. <https://doi.org/10.1504/ijmlo.2023.128339>

- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550.
<https://doi.org/10.1177/00336882231162868>
- Koubaa, A., Boulila, W., Ghouti, L., Alzahem, A., & Latif, S. (2023). Exploring ChatGPT capabilities and limitations: A survey. *IEEE Access*, 11, 118698–118721.
<https://doi.org/10.1109/access.2023.3326474>
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon.
- Küçükkara, M. F., Ünal, M., & Sezer, T. (2024). Okul öncesi eğitimi öğretmenlerinin yapay zekâya ilişkin görüşleri. *Temel Eğitim Araştırmaları Dergisi*, 4(1), 17–28.
<https://doi.org/10.55008/te-ad.1431142>
- Lee, H. (2023). The rise of ChatGPT: Exploring its potential in medical education. *Anatomical Sciences Education*. <https://doi.org/10.1002/ase.2270>
- Li, B., Bonk, C. J., & Kou, X. (2023). Exploring the multilingual applications of ChatGPT. *International Journal of Computer-Assisted Language Learning and Teaching*, 13(1), 1–22. <https://doi.org/10.4018/ijcallt.326135>
- Li, J., 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>
- Li, Y. (2022). Teaching mode of oral English in the age of artificial intelligence. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.953482>
- Li, Z. (2021). Teachers in automated writing evaluation (AWE) system-supported ESL writing classes: Perception, implementation, and influence. *System*, 99, 102505.
<https://doi.org/10.1016/j.system.2021.102505>

- Li, Z., Feng, H., & Saricaoglu, A. (2017). The short-term and long-term effects of AWE feedback on ESL students' development of grammatical accuracy. *CALICO Journal*, 34(3), 355–375. <https://doi.org/10.1558/cj.26382>
- Liao, H. (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System*, 62, 77–92. <https://doi.org/10.1016/j.system.2016.02.007>
- Lidén, A., & Nilros, K. (2020). Perceived benefits and limitations of chatbots in higher education. *Kalmar: Linnaeus University*. <http://www.diva-portal.org/smash/record.jsf?pid=diva2:1442044>
- Lin, Y. L. Y., & Lin, J. Y. Y. (2023). Development of an educational chatbot system for enhancing students' biology learning performance. *Journal of Internet Technology*, 24(2), 275–281. <https://doi.org/10.53106/160792642023032402006>
- Linacre, J. M. (1993). Rasch-based generalizability theory. *Rasch Measurement Transaction*, 7(1), 283–284.
- Linacre, J. M. (2017). *FACETS (Version 3.80.0) [Computer Software]*. MESA Press.
- Link, S., Mehrzad, M., & Rahimi, M. (2020). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 35(4), 605–634. <https://doi.org/10.1080/09588221.2020.1743323>
- Liu, G., & Ma, C. (2023). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*, 18(2), 125–138. <https://doi.org/10.1080/17501229.2023.2240316>

- Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: how may AI and GPT impact academia and libraries? *Library Hi Tech News*, 40(3), 26–29. <https://doi.org/10.1108/lhtn-01-2023-0009>
- Manchón, R. M., & De Larios, J. R. (2007). Writing-to-learn in instructed language learning contexts. In *Springer eBooks* (pp. 101–121). https://doi.org/10.1007/978-1-4020-5639-0_6
- McCarthy, J., Minsky, M. L., Shannon, C. E., Rochester, N., & College, D. (1955). A proposal for the Dartmouth Summer Research Project on Artificial Intelligence August 31, 1955. *AI Magazine*, 27(4), 12.
- Memarian, B., & Doleck, T. (2023). ChatGPT in education: Methods, potentials, and limitations. *Computers in Human Behavior Artificial Humans*, 1(2), 100022. <https://doi.org/10.1016/j.chbah.2023.100022>
- Menekse, M. (2023). Envisioning the future of learning and teaching engineering in the artificial intelligence era: Opportunities and challenges. *Journal of Engineering Education*, 112(3), 578–582. <https://doi.org/10.1002/jee.20539>
- Meng-Yue, C., Dan, L., & Jun, W. (2020). A study of college English culture intelligence-aided teaching system and teaching pattern. *English Language Teaching*, 13(3), 77. <https://doi.org/10.5539/elt.v13n3p77>
- Mijwil, M. M., Aggarwal, K., Mutar, D. S., Mansour, N., & Singh, R. S. S. (2022). The position of artificial intelligence in the future of education: An overview. *Asian Journal of Applied Sciences*, 10(2). <https://doi.org/10.24203/ajas.v10i2.6956>
- Mohamed, A. M. (2023). Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: perceptions of EFL Faculty Members. *Education and Information Technologies*, 29(3), 3195–3217. <https://doi.org/10.1007/s10639-023-11917-z>

- Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4(4), 386-422.
- Nabiyev, V., & Erümit, A. K. (2020). Eğitimde yapay zeka kuramdan uygulamaya. In *Ankara Pegem Akademi Yayıncılık eBooks*.
<https://doi.org/10.14527/9786257052986>
- Nation, I. (2008). *Teaching ESL/EFL reading and writing*. Routledge.
- Nguyen, T. T. H. (2023). EFL teachers' perspectives toward the use of ChatGPT in writing classes: A case study at Van Lang University. *International Journal of Language Instruction*, 2(3), 1–47. <https://doi.org/10.54855/ijli.23231>
- Nugroho, A., Putro, N. H. P. S., & Syamsi, K. (2023). The potentials of ChatGPT for language learning: Unpacking its benefits and limitations. *Register Journal*, 16(2), 224–247. <https://doi.org/10.18326/register.v16i2.224-247>
- OpenAI. (2024, August 8). ChatGPT — Release Notes. <https://help.openai.com/en/articles/6825453-chatgpt-release-notes>
- Oshima, A., & Hogue, A. (2007). *Introduction to academic writing*. Longman Publishing Group.
- Öztürk, K., & Şahin, M. E. (2018). Yapay sinir ağları ve yapay zekâ'ya genel bir bakış. *Takvim-i Vekayi*, 6(2), 25-36.
<https://dergipark.org.tr/tr/pub/takvim/issue/40063/427526>
- Patiño, A. E. C., Montaña, V. M. V., Buri, P. E. C., Rojas, M. E. A., & Del Rocío Cabrera González, A. (2024). The improvement of oral communicative competence in English through the artificial intelligence. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 5(1). <https://doi.org/10.56712/latam.v5i1.1850>

- Pavlik, J. V. (2023). Collaborating with ChatGPT: Considering the implications of Generative Artificial intelligence for journalism and media education. *Journalism & Mass Communication Educator*, 78(1), 84–93. <https://doi.org/10.1177/10776958221149577>
- Perbawa, I. K. S. L. P. (2021). The application of artificial intelligence as evidence in the court. *International Journal of Social Science and Business*, 5(2). <https://doi.org/10.23887/ijssb.v5i2.36185>
- Pham, X. L., Pham, T., Nguyen, Q. M., Nguyen, T. H., & Cao, T. T. H. (2018). Chatbot as an intelligent personal assistant for mobile language learning. *Proceedings of the 2018 2nd International Conference on Education and E-Learning*, 16–21. <https://doi.org/10.1145/3291078.3291115>
- Phuket, P. R. N., & Othman, N. B. (2015). Understanding EFL students' errors in writing. *Journal of Education and Practice*, 6(32), 99–106. <http://files.eric.ed.gov/fulltext/EJ1083531.pdf>
- Popenici, S. a. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1). <https://doi.org/10.1186/s41039-017-0062-8>
- Qasem, F., Ghaleb, M., Mahdi, H. S., Khateeb, A. A., & Fadda, H. A. (2023). Dialog chatbot as an interactive online tool in enhancing ESP vocabulary learning. *Saudi Journal of Language Studies*, 3(2), 76–86. <https://doi.org/10.1108/sjls-10-2022-0072>
- Qu, X., Lin, H., & Liu, Y. (2023). Envisioning the future of transportation: Inspiration of ChatGPT and large models. *Communications in Transportation Research*, 3, 100103. <https://doi.org/10.1016/j.commtr.2023.100103>

- Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, 13(9), 5783. <https://doi.org/10.3390/app13095783>
- Ranalli, J. (2018). Automated written corrective feedback: how well can students make use of it? *Computer Assisted Language Learning*, 31(7), 653–674. <https://doi.org/10.1080/09588221.2018.1428994>
- Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2016). Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, 37(1), 8–25. <https://doi.org/10.1080/01443410.2015.1136407>
- Reid, J. M. (1993). *Teaching ESL writing*. Prentice Hall.
- Reuters. (2023, November 14). *ChatGPT Sets Record for Fastest-growing User Base - Analyst Note*. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>
- Richards, J. C. (2015). *Error Analysis: Perspectives on Second Language Acquisition*. Routledge.
- Rock, J. L., & Service, E. T. (2007). The impact of short-term use of Criterion on writing skills in Ninth grade. *ETS Research Report Series*, 2007(1). <https://doi.org/10.1002/j.2333-8504.2007.tb02049.x>
- Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26(2), 582–599. <https://doi.org/10.1007/s40593-016-0110-3>
- Ruane, E., Birhane, A., & Ventresque, A. (2019). Conversational AI: Social and ethical considerations. *AICS*, 104–115. http://ceur-ws.org/Vol-2563/aics_12.pdf

- Russell, S., & Norvig, P. (2016). *Artificial intelligence: A Modern Approach*. Createspace Independent Publishing Platform.
- Schleiss, J., Laupichler, M. C., Raupach, T., & Stober, S. (2023). AI course design planning framework: Developing domain-specific AI education courses. *Education Sciences*, 13(9), 954. <https://doi.org/10.3390/educsci13090954>
- Schmidt-Fajlik, R. (2023). ChatGPT as a grammar checker for Japanese English language learners: A comparison with Grammarly and ProWritingAid. *AsiaCALL Online Journal*, 14(1), 105–119. <https://doi.org/10.54855/acoj.231417>
- Scriven, M. (1991). *Evaluation Thesaurus*. SAGE.
- Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation Practice*, 17(2), 151–161. [https://doi.org/10.1016/s0886-1633\(96\)90020-3](https://doi.org/10.1016/s0886-1633(96)90020-3)
- Şenel, E. (2018). The integration of creative writing into academic writing skills in EFL classes. *International Journal of Languages Education*, 1(Volume 6 Issue 2), 115–120. <https://doi.org/10.18298/ijlet.2869>
- Shaikh, S., Yayilgan, S. Y., Klimova, B., & Pikhart, M. (2023). Assessing the usability of ChatGPT for formal English language learning. *European Journal of Investigation in Health Psychology and Education*, 13(9), 1937–1960. <https://doi.org/10.3390/ejihpe13090140>
- Shen, C., Shi, P., Guo, J., Xu, S., & Tian, J. (2023). From process to product: writing engagement and performance of EFL learners under computer-generated feedback instruction. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1258286>
- Shawar, B. A., & Atwell, E. (2007). Chatbots: Are they really useful? *Journal for Language Technology and Computational Linguistics*, 22(1), 29–49. <https://doi.org/10.21248/jlcl.22.2007.88>

- Shi, N., Zeng, Q., & Lee, R. (2020). Language chatbot–The design and implementation of English language transfer learning agent apps. In *2020 IEEE 3rd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*. <https://doi.org/10.1109/auteee50969.2020.9315567>
- Shumanov, M., & Johnson, L. (2021). Making conversations with chatbots more personalized. *Computers in Human Behavior*, 117, 106627. <https://doi.org/10.1016/j.chb.2020.106627>
- Silverman, D. (2016). *Qualitative Research*. SAGE.
- Smestad, T. L., & Volden, F. (2019). Chatbot personalities matters. In *Lecture Notes in Computer Science* (pp. 170–181). https://doi.org/10.1007/978-3-030-17705-8_15
- Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: a framework for applying generative AI in education. *ECNU Review of Education*, 6(3), 355–366. <https://doi.org/10.1177/20965311231168423>
- Sumo, D. S., & Bah, M. L. (2021). Chinese language education in the era of artificial intelligence: Innovation development, pedagogy, & the smart classroom. *Zenodo (CERN European Organization for Nuclear Research)*. <https://doi.org/10.31014/aior.1993.04.04.390>
- Sun, G. H., & Hoelscher, S. H. (2023). The CHATGPT Storm and what faculty can do. *Nurse Educator*, 48(3), 119–124. <https://doi.org/10.1097/nne.0000000000001390>
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in Second Language Acquisition*, 15(1), 165–179. <https://ci.nii.ac.jp/naid/10030942470>
- Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16(3), 371–391. <https://doi.org/10.1093/applin/16.3.371>

- Sweetman, D., Badiee, M., & Creswell, J. W. (2010). Use of the transformative framework in mixed methods studies. *Qualitative Inquiry*, 16(6), 441–454. <https://doi.org/10.1177/1077800410364610>
- Tawney, J. W., & Gast, D. L. (1984). *Single subject research in special education*. Merrill Publishing Company.
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1). <https://doi.org/10.1186/s40561-023-00237-x>
- 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. (1999). The case for “The case against grammar correction in L2 writing classes”: A response to Ferris. *Journal of Second Language Writing*, 8(2), 111–122. [https://doi.org/10.1016/s1060-3743\(99\)80124-6](https://doi.org/10.1016/s1060-3743(99)80124-6)
- Turing, A. M. (2009). *Computing machinery and intelligence* (pp. 23-65). Springer Netherlands.
- Uygun, D., Aktaş, I., Duygulu, İ., & Köseer, N. (2024). Exploring teachers’ artificial intelligence awareness. *Advances in Mobile Learning Educational Research*, 4(2), 1093–1104. <https://doi.org/10.25082/amlr.2024.02.004>
- Vo, T. T. M. (2022). EFL students’ attitudes towards teacher correction and peer correction in writing skills. *International Journal of Language Instruction*, 1(1), 155–173. <https://doi.org/10.54855/ijli.221113>
- Wali, F., & Huijser, H. (2018). Write to improve: Exploring the impact of an automated feedback tool on Bahraini learners of English. *Learning and Teaching in Higher*

- Education Gulf Perspectives*, 15(1), 14–34.
<https://doi.org/10.18538/lthe.v15.n1.293>
- Wang, Y., Shang, H., & Briody, P. (2013). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning*, 26(3), 234–257.
<https://doi.org/10.1080/09588221.2012.655300>
- Wang, Z., & Han, F. (2022). The effects of teacher feedback and automated feedback on cognitive and psychological aspects of foreign language writing: A Mixed-Methods Research. *Frontiers in Psychology*, 13.
<https://doi.org/10.3389/fpsyg.2022.909802>
- Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: an International Journal*, 3(1), 22–36.
<https://doi.org/10.1080/15544800701771580>
- Widianingtyas, N., Mukti, T. W. P., & Silalahi, R. M. P. (2023). ChatGPT in language education: Perceptions of teachers - a beneficial tool or potential threat? *Voices of English Language Education Society*, 7(2), 279–290.
<https://doi.org/10.29408/veles.v7i2.20326>
- Williams, C. (2011). Research methods. *Journal of Business & Economics Research (JBER)*, 5(3). <https://doi.org/10.19030/jber.v5i3.2532>
- 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>
- Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo, and potential future development. *IEEE/CAA*

- Journal of Automatica Sinica, 10(5), 1122–1136.
<https://doi.org/10.1109/jas.2023.123618>
- Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, 8(3), 212.
<https://doi.org/10.3390/languages8030212>
- Yagiz, O., Yigiter, K., & Genc, G. (2009, June). Academic writing in English: Approaches, processes and challenges in higher education. In *1st International Symposium on Sustainable Development*.
<https://eprints.ibu.edu.ba/files/original/a34c298042a03cdf153e30b36f8a874a.pdf>
- Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, 34(3), 327–343.
<https://doi.org/10.1017/s0958344022000039>
- Yetişensoy, O. (2024). Tomorrow's teachers and artificial intelligence: Exploring attitudes and perceptions of Turkish prospective social studies teachers. *Eurasian Journal of Teacher Education*, 5(1), 1–31.
<https://dergipark.org.tr/en/pub/ejte/issue/84261/1408227>
- Ying, Z. (2018). Exploring construction of college English writing course from the perspective of output-driven hypothesis. *English Language Teaching*, 11(2), 188.
<https://doi.org/10.5539/elt.v11n2p188>
- Zhan, Y. (2022). Developing and validating a student feedback literacy scale. *Assessment & Evaluation in Higher Education*, 47(7), 1087-1100.
- Zhang, L., Zhang, X., Yang, L., Chen, Y., Zhang, Y., & Wang, Q. (2022). Teaching reform of computer science in combination with artificial intelligence.

International Journal of Education and Humanities, 5(3), 152–155.

<https://doi.org/10.54097/ijeh.v5i3.2792>

Zhang, P., & Tur, G. (2023). A systematic review of ChatGPT use in K-12 education.

European Journal of Education. <https://doi.org/10.1111/ejed.12599>

Zhu, C., Sun, M., Luo, J., Li, T., & Wang, M. (2023). How to harness the potential of

ChatGPT in education? *Knowledge Management & E-Learning an International*

Journal, 15(2), 133–152. <https://doi.org/10.34105/j.kmel.2023.15.008>





APPENDICES

Appendix A. Ethics Committee Approval



T.C.
SÜLEYMAN DEMİREL ÜNİVERSİTESİ
Üniversite Etik Kurulu



Sayı :E-87432956-050.99-662920
Konu :Bilimsel Çalışma Onayı.

Tarih: 10.01.2024

Ali Dinçay AYAN

Yürütücülüğünü yaptığınız "Analysis of Teacher, ChatGPT and Teacher and ChatGPT Feedback for EFL Writing" başlıklı bilimsel çalışmanız Sosyal ve Beşeri Bilimler Etik Kurulu'nun 08.01.2024 tarihli ve 144/28 sayılı kararı ile uygun bulunmuştur.

Kararın bir sureti gönderilmiş olup, başvuruza cevaben bilgilerinizi rica ederim.

Prof. Dr. Alim KOŞAR
Rektör a.
Rektör Yardımcısı

Bu evrak 5070 sayılı Elektronik İmza Kanununun 5. maddesi gereğince güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu: 921C0D78
SDÜ Batı Yerleşkesi Çünür/İSPARTA
Tel No:(246) 211-8052 Faks No:(246) 237-0431
E-Posta:zuhalbardak@sdu.edu.tr İnternet Adresi: www.sdu.edu.tr
Kep Adresi: sdu@hs01.kep.tr

Belge Takip Adresi: <https://cbys.sdu.edu.tr/EvrakDogrula.html?921C0D78>

Bilgi İçin: Zühal Sema BARDAK

Tel No: 2118052



T.C.
SÜLEYMAN DEMİREL ÜNİVERSİTESİ
Sosyal ve Beşerî Bilimler
Etik Kurul Kararları

TOPLANTI TARİHİ	TOPLANTI SAYISI	KARAR SAYISI
08.01.2024	144	28

28- Üniversitemiz Eğitim Bilimleri Enstitüsü İngiliz Dili Eğitimi Bölümü yüksek lisans öğrencisi Ali Dinçay AYAN tarafından yürütülen "Analysis of Teacher, ChatGPT and Teacher and ChatGPT Feedback for EFL Writing" başlıklı bilimsel çalışma hk.

Üniversitemiz Eğitim Bilimleri Enstitüsü İngiliz Dili Eğitimi Bölümü yüksek lisans öğrencisi Ali Dinçay AYAN tarafından yürütülen "Analysis of Teacher, ChatGPT and Teacher and ChatGPT Feedback for EFL Writing" başlıklı bilimsel çalışmanın kapsam ve uygulama açısından etik ilkelere ve insan haklarına uygun olduğuna,

Mevcudun oybirliği ile karar verildi.

Prof.Dr.Alın KOŞAR
Başkan

Prof.Dr.Hüseyin GÜL
Üye

Prof.Dr.Talat SAKALLI
Üye

Prof.Dr.Serpil SENAL
Üye

Prof.Dr.Gülcan SARP
Üye

Prof.Dr.Hülya ALTUNYA
Üye

Prof.Dr.Cevdet YILMAZ
Üye

08.01.2024
Zühal BARDAK
Raportör

Appendix B. Writing Tasks

Writing Task 1: Daily Life Hobbies

Instructions:

Write a short essay describing your hobbies and how you spend your free time. Write 250–300 words. Include information about the following:

- What are your hobbies?
- How often do you do them, and why do you enjoy them?
- Where you usually practice your hobbies.
- Who do you share your hobbies with, if anyone?

Writing Task 2: Daily Life Relations

Instructions:

Write a short essay about your relationships with people in your daily life. Write 250–300 words in four paragraphs. Include information about the following:

- Who do you spend most of your time with (friends, family)?
- What do you usually do together?
- How you maintain these relationships.
- Why are these relationships important to you?

Appendix C. Rubric for Main- and Sub-feedback Categories

HOCs (Higher order concerns)	- Content - Organization - Coherence - Cohesion	-Clarity and understandability -Development or lack of development -Accuracy of information, truth value of claim, accuracy of interpretation -Transitions -Thesis statement -Topic sentence -Coherence, cohesion -Idea placement -Paragraph order
LOCs (Lower order concerns)	-Vocabulary -Grammar/Syntax /Morphology -Mechanics	-Word choice, collocations, phrasing -Sentence structure -Word choice -Verb tense -Verb form -Noun endings (singular/plural) -Word form -Articles/determiners -Pronouns -Preposition -Conjunctions -Subject-verb agreement -Punctuation -Missing word -Extra word, redundancy, or repetition -Overall quality of grammar -Spelling -Documentation or attribution - Formatting and style

Appendix D. Semi-structured Interview Questions

Theme 1: Strengths

1. Teacher Feedback

- How does teacher-generated feedback contribute to the overall writing of students' work?
- In what ways does teacher feedback effectively address both higher-order concerns (HOCs) like content and organization, and lower-order concerns (LOCs) like grammar and vocabulary?

2. ChatGPT Feedback

- How does ChatGPT feedback support students' overall writing?
- Does ChatGPT feedback effectively address HOCs (such as structure and coherence) and LOCs (such as grammar and vocabulary)? If so, how?

3. Hybrid Feedback

- How does hybrid feedback (combining ChatGPT and teacher feedback) contribute to students' writing?
- In your experience, does the hybrid feedback approach effectively balance addressing both HOCs and LOCs? If yes, what specific aspects does it enhance?

Theme 2: Challenges

1. Teacher Feedback

- What challenges do you encounter with teacher-generated feedback in terms of enhancing students' writing?
- Are there challenges in teacher feedback when addressing HOCs and LOCs? If so, how do these affect students' ability to improve their writing?

2. ChatGPT Feedback

- What challenges have you observed with ChatGPT feedback in supporting overall writing quality?
- Are there limitations to ChatGPT's ability to address HOCs and LOCs effectively? Which areas seem most affected?

3. Hybrid Feedback

- What challenges arise in using hybrid feedback to improve students' writing?
- Are there specific difficulties in the way hybrid feedback handles HOCs and LOCs? How do students respond to the combined feedback on these aspects?

Theme 3: Further Suggestions

1. Teacher Feedback

- What suggestions do you have for enhancing the effectiveness of teacher-generated feedback to better support students' writing?
- How could teacher feedback be refined to more effectively address both HOCs and LOCs?

2. ChatGPT Feedback

- What improvements could make ChatGPT feedback more effective for fostering students' writing quality?
- Are there any adjustments that could improve ChatGPT's handling of HOCs and LOCs in feedback?

3. Hybrid Feedback

- What recommendations would you make to enhance the hybrid feedback approach in supporting students' writing?
- How might hybrid feedback be better tailored to balance and clearly address HOCs and LOCs for EFL students?