AN INVESTIGATION ON THE EFFECTIVENESS OF CHATBOTS IN EVALUATING WRITING ASSIGNMENTS IN EFL CONTEXTS

YABANCI DİL OLARAK İNGİLİZCE BAĞLAMINDA YAZMA ÖDEVLERİNİ DEĞERLENDİRMEDE SOHBET BOTLARININ ETKİNLİĞİ ÜZERİNE BİR ARAŞTIRMA

Saliha TOSCU[1]

Başvuru Tarihi: 24.01.2024 Yayına Kabul Tarihi: 29.07.2024 DOI: 10.21764/maeuefd.1425384

(Araştırma Makalesi)

Abstract: This study aims to explore the effectiveness of chatbots, specifically ChatGPT, in evaluating writing assignments. 51 English as a foreign language (EFL) learners from Türkiye took part in the study voluntarily. Based on a quasi-experimental research design, the study involved dividing the participants into two non-equivalent groups. The participants in the control group received feedback on their writing from a teacher, while the experimental group was given feedback by the ChatGPT. Thus, the research indicated how Artificial Intelligence (AI)-generated feedback was influential in evaluating writing assignments compared to human-generated feedback. The feedback given by the teacher and the ChatGPT was analyzed via a content analysis, and the results showed that the ChatGPT and the teacher feedback had differences in the language style, register, content, and accuracy. Besides, students' perceptions of the feedback were revealed through the Student Perceptions of Writing Feedback Scale of Marss (2016) and analyzed using Independent Samples t-Tests on the SPSS. The statistical tests indicated that the students' perceptions did not vary whether the feedback was AI or human-generated. This study suggests that AI-generated feedback can match human feedback in terms of students' perceptions. Still, the differences regarding feedback features show the strengths of a teacher and a bot for giving feedback by suggesting a more complementary role for AI technology in educational settings.

Keywords: *Chatbots, ChatGPT, EFL writing, feedback, evaluation*

Özet: Bu çalışma, sohbet botlarının, özellikle ChatGPT'in, yazma ödevlerini değerlendirmedeki etkinliğini ortaya çıkarmayı hedeflemektedir. Türkiye'deki özel bir okuldaki 51 yabancı dil olarak İngilizce öğrencisi çalışmaya gönüllü olarak katılmıştır. Yarı deneysel araştırma desenine dayalı olan çalışmada katılımcılar eşit olmayan iki gruba ayrılmıştır. Kontrol grubundaki katılımcıların yazma ödevlerine öğretmen, deney grubundakilere ChatGPT tarafından dönüt verilmiştir. Böylece, çalışma yapay zekâ tarafından üretilen geri bildirimin insan tarafından üretilen geri bildirimle karşılaştırıldığında yazma ödevlerini değerlendirmede etkinliğini göstermiştir. Öğretmen ve ChatGPT tarafından verilen dönüt içerik analizi yoluyla incelenmiştir ve sonuçlar ChatGPT ve öğretmen dönütü arasında dil stili, dil biçimi, içerik ve dil bilgisi doğruluğu konularında farklılıklar olduğunu göstermiştir. Ayrıca, çalışmada öğrencilerin geri bildirime yönelik algıları Marss'ın (2016) Öğrenci Yazma Geri Bildirim Algısı Ölçeği yoluyla ortaya çıkarılmıştır ve SPSS'te Bağımsız Grup t-Test kullanılarak analiz edilmiştir. İstatistiksel testler, öğrencilerin algılarının geri bildirimin yapay zekâ veya insan tarafından üretildiğinde değişmediğini göstermiştir. Bu çalışma, öğrenci algısı bakımından yapay zekâ teknolojisi ve insan geri bildiriminin eşleşebileceğini önermektedir, ancak dönütün özelliklerine ilişkin farklılıklar yapay zekâ teknolojisinin eğitim ortamlarında daha tamamlayıcı bir rol oynamasını önererek öğretmenin ve botun geri bildirim vermede güçlü yönlerini göstermektedir.

Anahtar Sözcükler: *Sohbet Robotları, chatgpt, yabancı dil olarak İngilizce yazma, geri bildirim, değerlendirme*

[1] Dr., Çankaya University, Department of Foreign Languages, Ankara, Türkiye, salihatoscu@cankaya.edu.tr, ORCID ID: 0000-0002-8179-5444

Introduction

Artificial Intelligence (AI) is "the creation and analysis of intelligent agents (software and machines)," and it may be operated in any work by functioning in numerous ways (Khanna et al., 2015, p. 277). Recent advancements in AI have become a reality in almost all areas. Among the significant developments in AI, the integration of "artificial intelligence and virtual reality" has remarkably expanded the concept of communication, and this extension goes beyond human-to-human interactions, now encompassing interactions between humans and machines (Çakmak, 2022, p. 114).

Chatbots, as one of the prominent applications of AI, use natural language processing, which has rendered it possible for machines to understand, produce, and communicate with human beings in a unique way (Khanna et al., 2015). Different kinds of chatbots with various strengths and potential are emerging to address different audiences. They encompass "artificial conversation entities, interactive agents, smart bots, and digital assistants" (Adamopoulou & Moussiades, 2020, p. 1). Their prevalence has grown across various sectors, including customer service, health care, and education (Adamopoulou & Moussiades, 2020).

In educational settings, chatbots allow learners to communicate with smart technology agents for diverse purposes, varying from information delivery to the management of extensive class activities (Çakmak, 2022; Huang, Hew, & Fryer, 2022). Chatbots can be beneficial for issues such as learning support and maintaining administration in education by transforming traditional language classrooms (Adamopoulou & Moussiades, 2020). Especially in foreign language learning, the use of chatbots encapsulates learners' involvement in natural language practice, conduct of assessment, and provision of feedback (Huang et al., 2022). They also offer benefits for the users such as motivation, content integration, easiness of fast access, enabling multiple users, and quick and immediate support from the application, opportunity for personalized learning (Okonkwo & Ade-Ibijilo, 2021). Çakmak (2022) explains that personalized learning has emerged as an effect of students' individual needs, and artificial intelligent systems have brought novel approaches in pedagogy and program developments in learning settings. Chatbots are regarded as having the potential to support language learners' personalized learning since they enable them to personalize their time and practice by giving them interaction opportunities in foreign language learning settings.

Additionally, the studies indicate its positive effects on motivation and confidence. Çakmak (2022) states that using chatbots for language practice on learning impacts learners' motivation positively since learners are likely to feel more comfortable and involved while practicing the target language by using chatbots. In a different study, Huang et al. (2022) noted that chatbot self-closure positively impacts learners' confidence and enjoyment by creating an open and social atmosphere for language learning. This is because the students are encouraged to give long answers and can express their emotions on the topics in "a friendly thread, expressing agreement, and addressing students by names" (Huang et al., 2022, p. 253).

As explained by Fryer and Carpenter (2006), teachers' lack of time to create group or pair work activities in the classroom and to provide valuable feedback and the students' lack of opportunities to practice the foreign language and confidence are the factors affecting foreign language learning. The integration of chatbots into foreign language learning enables teachers to employ them to improve their teaching practices and allows learners to improve their interaction skills (El Shazly, 2021). Due to their "timeliness, personalization, and ease of use" features, chatbots effectively improve learners' communication in the language they are learning (Fryer & Carpenter, 2006; Huang et al., 2022, p. 253). Hence, it allows one to practice the target language without the boundaries of time and place with a language partner, never tired of responding to questions (Adamopoulou & Moussiades, 2020). Learners who interact with chatbots can decide when to start and end the interaction by controlling their learning pace (Çakmak, 2022). Its use in foreign language learning settings makes learning student-oriented, collaborative, and not restricted by boundaries (El Shazly, 2021).

Regardless of their advantages, chatbots have limitations. One limitation is related to privacy and security when the data collected in a service are shared with third parties (Adamopoulou & Moussiades, 2020). Additionally, misunderstandings in a conversation can cause a hindrance to the conversation and end in a person's renunciation of the use of the agent (Adamopoulou & Moussiades, 2020). Moreover, Okonkwo and Ade-Ibijilo (2021) report that chatbots may have some disadvantages due to "ethical, evaluation, user attitudes, and maintenance issues" (p.8).

Chatbots differ from human-to-human conversations since conversation with chatbots lacks empathy (Adamopoulou & Moussiades, 2020). Fryer and Carpenter (2006) explain that a chatbot's purpose is to continue a conversation with a human being, which is crucial for foreign language

learners. Despite this, they are still inadequate because they cannot communicate in a human-like style by giving machine-produced responses (Fryer & Carpenter, 2006). Nevertheless, Fryer and Carpenter (2006) posit that because of the future advancements in AI, they may become more sensitive to human feelings. Chantarotwong (2005), as cited in Yin and Satar (2020), points out that the discourse with chatbots is easily anticipated, unnecessary, has no character, and has no memory of preceding replies. Furthermore, chatbots have technological limitations and may pose a novelty effect like all the other modern technologies, adding to the cognitive load of knowledge for the students in the learning process (Huang et al., 2022). All in all, chatbots offer promising advantages in foreign language learning settings, as indicated in the literature. However, understanding their limitations is crucial for the optimal and responsible use of chatbots.

Automated Writing Evaluation or Feedback Tools and Theoretical Background

Writing has become an essential skill to develop in educational realms because English is a global language; however, its instruction is not effortless because it compels time and skills to assess learners' writings (Warschauer & Ware, 2006). The emergence of educational technologies has significantly had a considerable impact on teaching practices and learning. One of these technologies is Automated Writing Evaluation (AWE) (also referred to as Automated Writing Feedback (AWF)) tools, some examples of which are represented by Warschauer and Ware (2006) as Grammarly, Criterion, My Access, and e-rater. Such tools have started to be included in education programs because of the assets they possess. To illustrate, students' revision of their writing with AWE feedback is effective (Zhang, 2020); AWE tools help to decrease linguistic errors by improving accuracy in writing (Saricaoğlu & Bilki, 2021); they complement teacher feedback and help reduce the workload of evaluating multiple drafts of writings (Link, Mehrzad, & Rahimi, 2022); they enable teachers to facilitate their instruction (Li, 2021); they have the potential to improve students' self-efficacy of writing (Wilson & Roscoe, 2020). Despite the existing research suggesting its benefits, AWE tools have also received criticism regarding correcting more than wanted, not focusing on individual differences, and not being aligned with teacher's teaching practices.

Research on using AWE relies on various theories and approaches in second language learning. One of those theories is Long's (1996) Interaction Hypothesis Theory. In this theory, Long (1996) explains that negotiation of meaning fosters learners' language development. During negotiation,

learners engage in interaction and are given input, which has a significant role in language learning and fosters learners' language development through interaction. During the interaction, learners negotiate meaning in which they could get positive or negative feedback until the linguistic form or message becomes acceptably understandable (Long, 1996). Additionally, the negotiation process provides learners feedback, thereby encouraging learners to produce an output. Whether explicit or implicit, feedback is valuable because it shows the discrepancy between what is produced and what is targeted. The Interaction Hypothesis Theory stems from Krashen's input hypothesis (Krashen, 1985), which involves the notion that acquisition is possible when learners receive comprehensible or understandable input, and Swain's output hypothesis (Swain, 1995), in which Swain (1995) contends that learners can notice the linguistic problems in the output they have produced. In this process, learners may ask a teacher, for example, or a friend; thus, the learners can obtain new knowledge or associate the prevailing knowledge. Modifying a language output based on interaction and feedback effectively supports learners' language development process (Swain, 1985). Interaction, input, feedback, and output are vital concepts in language development. Likewise, AWE systems facilitate them by providing language users with opportunities to interact and receive input and feedback, thereby assisting learners' production process in language learning.

In addition to the Interaction Theory, the AWE systems are also supported by constructivist theory, in which a learner learns through active interaction with the physical and social world (Fosnot, 1996). Constructivist theories explain that language users actively build and demonstrate knowledge based on their practices (Szabó & Csépes, 2023). Szabó and Csépes (2023) expound that the knowledge constructed by individuals and realization of self-sufficiency, in other words, autonomy in learning feature differentiation as a key term, which suggests differences among learners with respect to learners' language development levels, assessment methods, which makes the feedback meaningful and significant for the individuals. An approach adopted by constructivism stimulates teachers professionally because the teachers look for ways to find solutions for learners with different learning styles at the right pace and with the optimum methods (Szabó & Csépes, 2023).

Besides, AWE is based on the Socio-Cultural Theory, which explains a language learner's cognitive development through social actions. It has critical concepts such as Zone of Proximal

Development (ZPD), mediator, and scaffolding. ZPD refers to a zone between actual and potential development (Vygotsky, 1978). For a language learner to move from the actual one to the zone of proximal development, scaffolding is essential with the help of a mediator. The notion of constructivism positively influences the integration of technology in education because it aims to foster reflective and cooperative learning. Individuals build knowledge by interacting with their environment, which is based on social constructivism. AWE tools create opportunities for language users to collaborate and build knowledge at their own pace and based on their needs.

In summary, AWE tools positively influence language learning processes by employing the principles of the hypotheses and the theories explained in this section. The systems offer an interactive learning environment by providing input and active, constant feedback to the learners. This process enables learners to build their understanding of effective writing. When social support such as peer review or collaboration with others is integrated into the process, it also supports learning on social terms.

Significance of the Study

As technology evolves, AI technologies also advance at the same pace, simultaneously opening new investigation areas. The effectiveness of using chatbots in foreign language learning has been investigated in numerous studies in the literature. The studies indicated their positive outcomes, such as diminishing preposition and article problems (Ahn, 2022), improving vocabulary skills, and increasing motivation and confidence (Kim, 2018), giving the possibility for negotiation for meaning in interactions, thereby improving communication skills (Yin & Satar, 2020), including writing. The studies involved the use of different sorts of chatbots such as Elbot (Kim, 2018), Replika (Çakmak, 2022), Mike (Yin & Satar, 2020), Mondly (El Shazly, 2021), Argumate (Guo et al., 2022), and ChatGPT. Among all these chatbots, ChatGPT, one of the most popular chatbots these days, is widely used to perform various deeds such as searching for information, responding to questions, doing research, engaging in discussion and conversation, writing reports and essays in learning environments (Halaweh, 2023). Halaweh (2023) mentions that ChatGPT is valuable in higher education for developing writing skills because it produces texts and describes the main facts of particular information, which is time-saving and effective for the quality of work. For improving writing, chatbots have the potential to be a supreme partner to cooperate with (Guo, Wang, & Chu, 2022). Guo et al. (2022) explain that chatbots are companions whose language

proficiency is better than language learners interacting with them, providing immediate responses to writing and creating a positive learning atmosphere because interaction with them is less stressful for the learners.

Writing is a challenging and sophisticated skill in language learning. It poses difficulties for learners and teachers because it requires regular practice and well-timed feedback (Burstein, Chodorow, & Leacock, 2004). Automated writing evaluation tools or automated evaluation tools provide ease in providing feedback and scoring. Different kinds of chatbots have been examined for writing assessments, and one of those chatbots is Chat Generative Pre-Trained Transformer (ChatGPT). Although it has been in use for a relatively brief time, there is a growing body of research on it. Previous research has indicated that ChatGPT serves well as an AI-based AWE tool. It differs from the earlier AWE systems because it is cost-free (or is low-cost) and available. Also, it can provide "timely, targeted, adaptive, and useful feedback" (Steiss et al., 2024, p.2). ChatGPT can potentially improve students' writing skills with rapid feedback and assess the students' papers based on language use (vocabulary, grammar, sentence form, etc. (Kovačević, 2023). Literature shows the studies conducted in the EFL context to examine ChatGPT use for writing skills. The studies indicated that ChatGPT provided feedback on content, organization, and language of writing effectively (Guo & Wang, 2024); language learners could improve their written outputs through revisions based on ChatGPT feedback (Tsai & Brown, 2024); the bot was effective for error analysis (Algaraady & Mahyoob, 2024); also, its efficacy for improving accuracy and spelling as well as idea generation for writing was explored in the previous research (Harunasari, 2022). Even though existing research provides valuable understanding, there is still essential space for research on ChatGPT to understand how learners perceive the feedback given by this relatively new and fast-evolving tool and how effective ChatGPT feedback is compared to teacher feedback. Thus, a clearer understanding of the strengths and weaknesses of this tool could be revealed. This article aims to explore the effectiveness of the use of ChatGPT for assessing writing in foreign language learning settings. The investigation into its effectiveness as a tool to assess EFL learners' writing is significant in educational assessment since it aims to present solid findings regarding how its use for writing evaluation could be benefitted and the extent to which it could be relied on. Concerning this purpose, the study addresses the research questions below:

- To what extent does the use of ChatGPT affect students' perceptions of the feedback they receive for their writing?
- To what extent does the feedback from ChatGPT on writing vary compared to teacher feedback?

Method

The present study intends to reveal the students' perceptions of the feedback given to writings by the ChatGPT and a teacher, as well as the differences in the feedback given by both assessors. To achieve this, both qualitative and quantitative data collection methods were used for data collection. This section details the research design, setting and participants, data collection and data analysis procedures followed in the study.

Research Design

The study was based on a quasi-experimental research design. 51 Turkish EFL students taking a weekly grammar course at university were divided into two non-equal groups: experimental and control. The groups were alike with respect to the course content given to them, the instructor giving the course, and the assignments. The only difference was that AI-generated feedback was given to the assignments submitted by the participants in the experimental group, whereas teacher feedback was given in the control group. At the end of the course, a questionnaire was administered to all students to reveal the students' perceptions. The research design is illustrated in Figure 1.

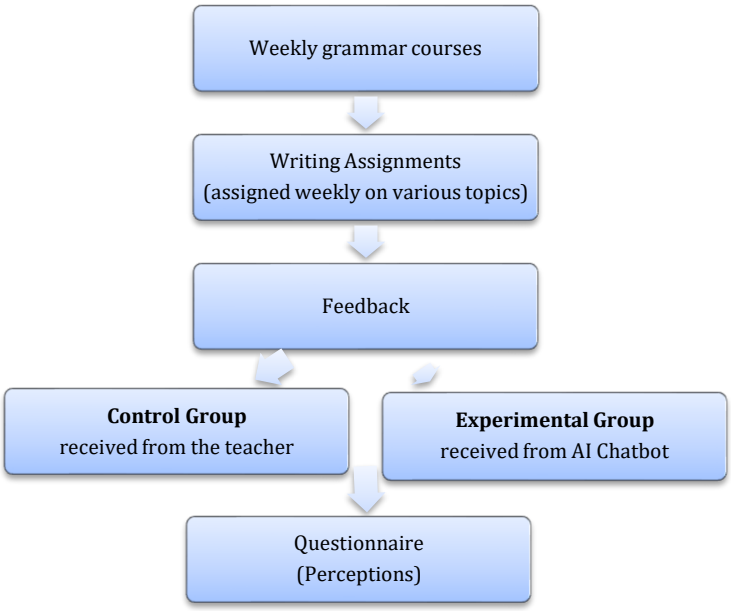


Figure 1. Research Design

Setting and Participants

The study was carried out in a private university in Türkiye. The participants were 51 Turkish first-year students enrolled in the Interpretation and Translation BA program in the 2022-2023 academic year in the Spring Semester. The students took the English Grammar in Context Course as a department must-course. The students' ages ranged from 18 to 26, and the time they were involved in learning English varied from 2 to 14 years at most. The participants took part in the study based on convenience sampling, and they were acknowledged that their participation in the study had no violation of ethics in research.

Table 1

Demographic Features of the Participants

	Control Group	*Experimental Group*
Number of participants	25	26
Mean of ages	20	21
Mean of the length of learning English	7.8 years	8.3 years
Departments	Interpretation and Translation	Interpretation and Translation

Instrument

In order to explore the first research question, a 5-point Likert scale, *Development of the Student Perceptions of Writing Feedback Scale*, developed by Marrs (2016), was employed. The questionnaire involved thirty-one items, which were grouped as related to four themes. Namely, the items aimed to reveal students' perceptions/ expectations of feedback, experiences with feedback, use/ value of feedback, and affect associated with feedback. The internal consistency of the questionnaire was at an acceptable level, with a score of 0.87 Cronbach's alpha (Marss, 2016). The Cronbach's alpha value was also calculated for this study on the SPSS 25, and the results indicated an acceptable level with a score of 0.76. The survey was administered to the students online so that the students' thoughts regarding the feedback given by either the teacher or the chatbot could be revealed. Each student took the survey individually.

Data Collection

For data collection, students' weekly writings were used. Also, the students were administered a survey online to get information regarding their opinions on the feedback they received for their writing.

Writing Assignments. In line with the objectives of the English Grammar in Context Course given in the Spring Semester, the students were assigned a writing task each week. The assignment process lasted five weeks. It started with the first writing assignment on 01.05.2023 and ended on 02.06.2023. The details of the procedure order and the topics assigned each week are shown in Figure 2 in detail. The entire number of writing assignments collected and analyzed in the present research was 194. Throughout the weeks, some students failed to send their assignments (n=61) because of personal problems such as missing the deadline and the weekly load of other courses at university. The missing assignments were not particular to a specific group of students, but various students failed to submit their assignments in diverse weeks.

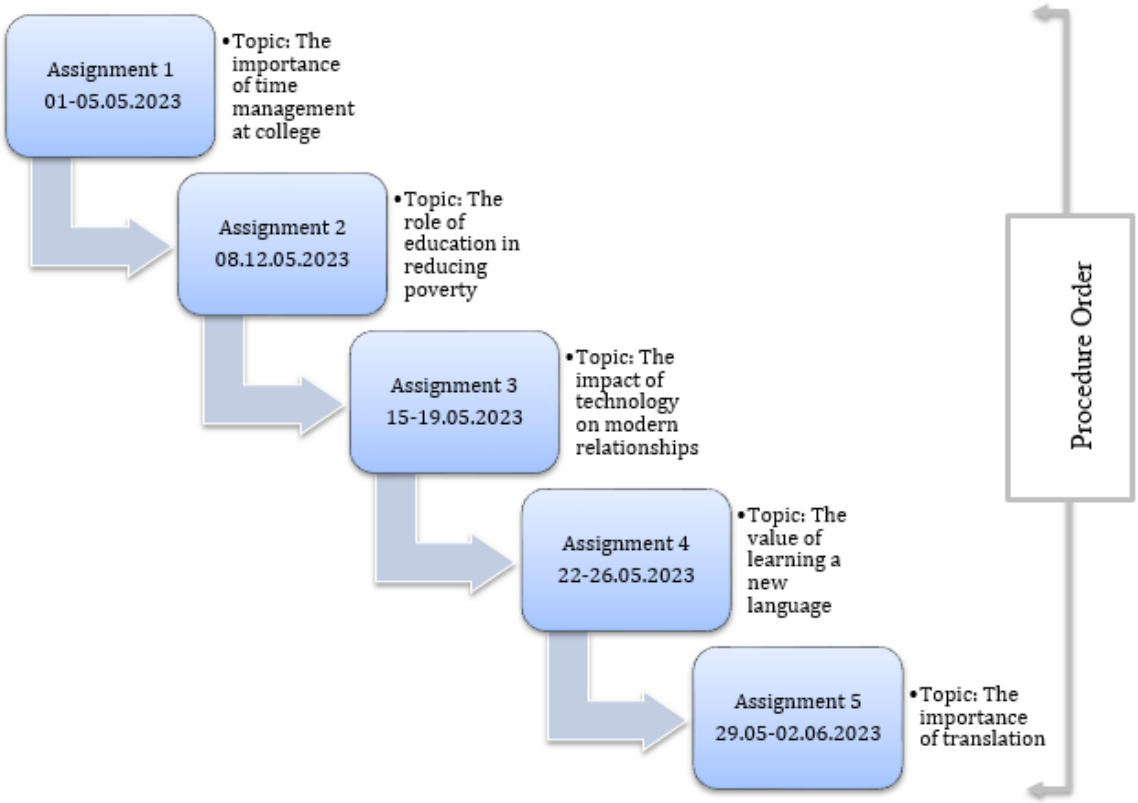
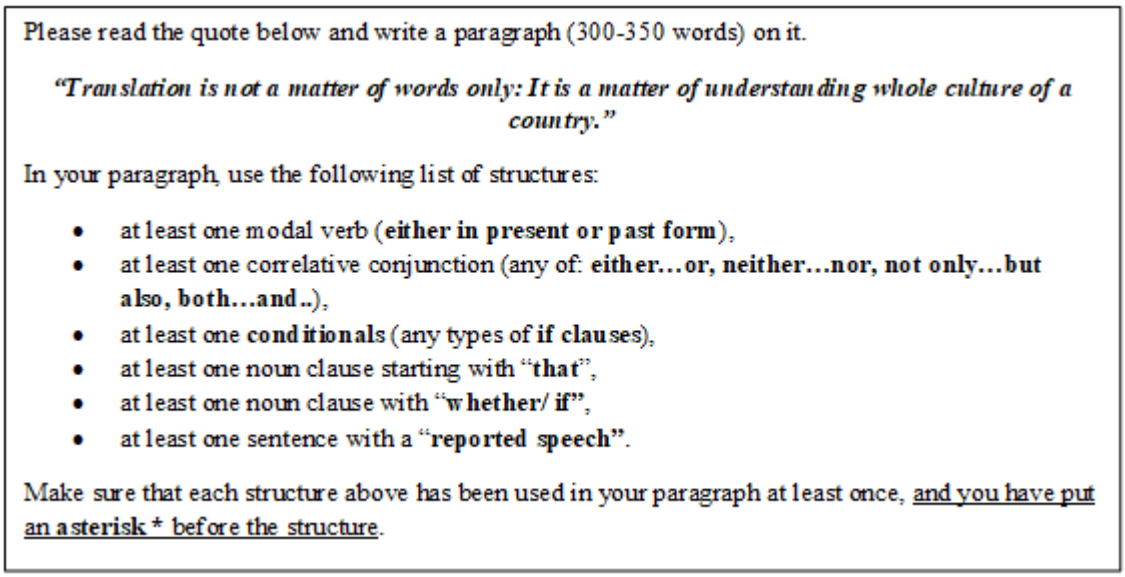


Figure 2. The details of the procedure order followed in data collection

The assignments were based on a weekly topic and involved specific grammatical structures determined according to the weekly writing topics. The word count for the assignments was limited to 350 words, but the papers were evaluated even if they exceeded that limit. The students were asked to write their papers in a WORD document and submit them to the educational platform Moodle. The papers submitted to the platform were later prepared for evaluation.

The topics of the writings were based on the use of general knowledge regarding the topic, and the students were asked to use the particular grammar structures correctly in their papers. The topics were various, such as the importance of time management in college, the role of education in reducing poverty, the impact of technology on modern relationships, the value of learning a new language, and the importance of translation. Each topic was assigned to the students throughout the research. The writing instruction involved the topic, the assignment's word count, and a list of grammar structures the students focused on the previous week and were responsible for using in their writings. An example of an instruction given is shared in the figure below:


Please read the quote below and write a paragraph (300-350 words) on it.

"Translation is not a matter of words only: It is a matter of understanding whole culture of a country."

In your paragraph, use the following list of structures:

- at least one modal verb (**either in present or past form**),
- at least one correlative conjunction (any of: **either…or, neither…nor, not only…but also, both…and..**),
- at least one **conditionals** (any types of **if clauses**),
- at least one noun clause starting with "**that**",
- at least one noun clause with "**whether/ if**",
- at least one sentence with a "**reported speech**".

Make sure that each structure above has been used in your paragraph at least once, and you have put an asterisk * before the structure.

Figure 3. Sample writing instruction.

The Evaluation Process of Writing. In the evaluation process, all the papers were assessed by taking the content and accuracy of the papers into consideration as criteria. Accordingly, the content involved the papers' analyses of idea generation, organization, and fluency. The evaluation of the accuracy part encapsulated the effectiveness of grammar and vocabulary used in the papers. Also, the grammar structures assigned to the students were analyzed by considering whether or not they were used effectively in the papers.

The teacher assessed each paper and gave feedback to them by considering the predetermined criteria. The teacher's feedback involved bullet point notes at the bottom of each paper. The teacher also used the 'Review' function of the Office WORD program to give feedback and then shared the document with the students by uploading the feedback files back to the platform.

To receive feedback from the ChatGPT, the students' papers were copied from Moodle and pasted to the chatbot by giving the evaluation criteria used by the teacher. To be precise, the chatbot was asked to evaluate each paper one by one, considering the content and accuracy of the papers. Additionally, the students' profile information was entered into the system so that the chatbot could evaluate each paper accordingly. The feedback produced by the chatbot was shared with the students on Moodle.

Data Analysis

The data analysis involved qualitative and quantitative methods. For research questions aiming to reveal the students' perceptions, quantitative data analysis methods were used through the SPSS. Specifically, the descriptive statistics and an Independent Pairs t-Test were performed so that the students' perceptions of the feedback given by the teacher and the ChatGPT could be compared.

The qualitative data analysis was based on grounded theory. The source of the content used in this study was the written feedback given to the students' writings (n= 194). Therefore, all the data were transferred to the qualitative data analysis software MAXQDA and read carefully for the thematic analysis. As a result of the content analysis, codes and categories were determined. The analysis process was carried out by two researchers so that interrater reliability could be ensured. The set of codes was created, and the data were systematically analyzed by two coders using spreadsheets. The researchers coded the prepared documents for the analysis individually first. Later, they brought the coded data together and checked the extent to which the codes were similar,

and both coders were consistent in their analysis. Two coders agreed with each other regarding the coded segments in the data. In case of a disagreement between the coders, the coders negotiated and compromised related to their analyses.

Findings

The findings from the quantitative and qualitative data analyses are presented in line with the research questions of the current study. This section first presents the findings from the quantitative data analysis and then the qualitative data analysis.

The effect of the ChatGPT on students' perceptions of the feedback. The first research question aimed to indicate the learners' perceptions of writing feedback. The perceptions specifically investigated learners' views/ expectations of feedback, experiences with feedback, usefulness/ value of feedback, and affect/ emotions associated with feedback. The results from the statistical tests indicated that the students' perceptions of feedback did not change whether the teacher or the ChatGPT gave it. The results are shared in Table 2 below.

Table 2

Differences in Perceptions of Feedback between the Groups

	Group	N	M	SD	Sig (two-tailed)
Views/ expectations of feedback	Control	25	32.04	5,75	.64
	Experimental	26	32.65	3,29	
Experiences with feedback	Control	25	27.96	2.99	.68
	Experimental	26	27.65	2.26	
Usefulness of feedback	Control	25	29.84	5.84	.59
	Experimental	26	30.65	4.89	
Affect/ emotions with feedback	Control	25	19.00	3.24	.08
	Experimental	26	20.30	1.87	
Total	Control	25	108.84	10.97	.40
	Experimental	26	111.26	9.50	

An Independent Samples t-Test was performed to compare the learners' perceptions of feedback in the control group (M= 108, SD= 10.9) and the experimental group (M= 111, SD=9.5). When the total statistics was compared, a statistically significant difference was not found between them t (49) = -.84, p= .40 (two-tailed) with a small effect size (eta squared= .01). Also, each sub-theme in the scale was calculated separately.

As shared in Table 2, the levels of views and expectations of feedback for the control group (*M*= 32.04, *SD*=5.75) did not significantly differ from the experimental group (*M*= 32.65, *SD*=3.29); *t* (49) = -.47, *p*= .64 (two-tailed). The magnitude of the differences in the means (mean difference= -.61, 95% CI: -3.24 to 2.01) was very small (eta squared= .004). This result shows that whether the students received the feedback from the ChatGPT, or the teacher did not make a difference in the students' views and expectations of feedback. Particularly, when the questionnaire items on a scale from 1 (strongly disagree) to 5 (strongly agree) are considered, it is understood that the students mostly had positive views and expectations of feedback, whether from the ChatGPT or the teacher. Although the students in the control group were given feedback by the teacher and in the experimental group by the ChatGPT, the results indicated no difference between the groups. Also, the results suggested that the students agreed that the feedback made them believe they were good at writing, they thought the feedback was necessary, they felt impatient about the feedback they would receive, the feedback encouraged them to write better next time, they thought the feedback was essential and explained their score of the writing. The students disagreed on the importance of feedback only when they got a good score and did not think that it led them to write worse.

Table 3

Views/ Expectations of Feedback

	M	SD	Min.	Max.
Feedback makes me feel like I am a good writer.	4.18	.93	1.00	5.00
I think I should get feedback even if I don't try very hard in my writing.	3.57	1.10	1.00	5.00
Feedback is not important if I get a good grade.	1.65	.82	1.00	5.00
I look forward to feedback on my writing.	4.12	1,05	1.00	5.00
Feedback I get on writing makes me want to become a better writer.	4.11	.97	1.00	5.00
Feedback on my writing encourages me to do better next time.	4.37	91	1.00	5.00
Feedback on my writing makes me feel like I am a bad writer.	1.82	.99	1.00	5.00
Feedback on my writing is important.	4.43	.83	1.00	5.00
Feedback on my writing explains my grade.	4.09	.85	2.00	5.00

N= 51

In addition, the results in Table 2 indicated that the students' experiences of feedback did not make a change whether they received feedback from the ChatGPT or not. There was not a statistically significant difference in the levels of experiences with the feedback between the learners in the control group (*M*= 27.96, *SD*=2.99) and the experimental group (*M*= 27.65, *SD*=2.26); *t* (49) = .41, *p*= .68 (two-tailed) with a very small effect size (eta squared= .003). The questionnaire items in Table 4 made it clear that the students in the control and experimental groups received feedback on their writing. Also, they agreed that the feedback was positive and helped them to understand

what they did well or poorly in their writing without any differences between the groups. The feedback was not considered too judgmental or challenging for the students in the control and experimental groups to understand.

Table 4

Experiences with Feedback

	M	*SD*	Min.	Max.
I get feedback on my writing.	4.80	.44	3.00	5.00
Feedback I get on my writing is too critical.	1.84	.90	1.00	5.00
Feedback is very specific.	2.47	1.37	1.00	5.00
Feedback on my writing is positive.	4.25	.82	3.00	5.00
Feedback on my writing is confusing.	1.86	.96	1.00	4.00
Feedback tells me what I did badly in my writing.	4.31	.93	1.00	5.00
Feedback tells me what I did well in my writing.	4.47	.70	1.00	5.00
I receive feedback soon after I turn in a writing assignment.	3.78	1.0	1.00	5.00

N= 51

Another theme investigated in the study was the usefulness of feedback. As can be seen in Table 2, the analysis did not indicate a statistically significant difference in the usefulness of the feedback when the statistics from the learners in the control group (*M*= 29.84, *SD*=5.84) and the experimental group were compared (*M*= 30.65, *SD*=4.89); *t* (49) = -.54, *p*= .59 (two-tailed) with a very small effect size (eta squared= .005). The result suggested that the students thought the feedback they received was helpful. The analysis of the items in Table 5 suggests that the students read their feedback and thought it helped them write better.

Table 5

Usefulness of Feedback

	M	*SD*	Min.	Max.
Feedback helps me write better next time.	4.14	1.00	3.00	5.00
Feedback on my writing is useful.	4.45	.87	1.00	5.00
Feedback makes me a better writer.	4.04	1.02	1.00	5.00
I read the feedback on my writing.	4.63	.69	3.00	5.00
I use feedback to help me write better next time.	4.22	.92	1.00	4.00
Feedback on my writing is helpful.	4.41	.92	1.00	5.00
Feedback tells me how to make my writing better.	4.37	.87	1.00	5.00

N= 51

Finally, the affect and feelings with feedback were analyzed. As can be seen in Table 2, the comparison of the levels of affect and emotions with the feedback from the learners in the control group (*M*= 19, *SD*= 3.24) and the experimental group (*M*= 20.30, *SD*= 1.87) did not indicate a

statistically significant difference; $t(49) = 1.77$, $p = .08$ (two-tailed) with a moderate effect size (eta squared= .06). A detailed look at the questionnaire items revealed that the students in both groups disagreed that the feedback caused them to experience negative emotions such as disappointment, frustration, hopelessness, or nervousness while writing or towards writing. On the other hand, as Table 6 indicates, the students agreed on the positive effects of the feedback on their emotions. The results indicated that the feedback in the control and experimental groups made the students feel proud of their writing. Also, the students reported feeling confident in their writing performance.

Table 6

Affect & Emotions with Feedback

	M	*SD*	Min.	Max.
Feedback on my writing makes me want to give up.	1.61	.98	3.00	5.00
Feedback on my writing makes me feel hopeless.	1.76	1.05	1.00	5.00
Feedback on my writing makes me feel nervous.	1.94	1.05	1.00	5.00
Feedback on my writing makes me feel frustrated.	1.71	.97	1.00	5.00
Feedback on my writing makes me feel proud.	4.13	.89	1.00	5.00
Feedback on my writing makes me feel confident.	4.18	.97	1.00	5.00
Feedback on my writing makes me feel happy.	4.33	.71	1.00	5.00

N= 51

Overall, as the statistical tests displayed, the students in the control and experimental groups were alike concerning their perceptions of the feedback. When the feedback was given by an AI technology or a human being, the students' views, experiences, emotions, and thoughts regarding its usefulness showed no variations.

The differences and/ or similarities between the writing feedback of the ChatGPT and the teacher. The second research aim of this paper was to show how writing feedback given by the teacher and ChatGPT varied. The findings involved a qualitative data analysis, and the analysis results are presented in detail in this section.

The presentation of feedback. The analysis first involved analyzing the feedback presentation and then the content-related detailed analysis results. The ChatGPT was asked to assess each paper depending on the rubric and the students' English proficiency levels. The exact process was repeated for each paper written by the students.

The general look of the feedback showed that the feedback output from the ChatGPT was long and detailed. It starts with a summary of the main ideas in the students' papers. Then, an explanation regarding the accuracy and content was produced in two paragraphs automatically, and the feedback ended again with a summary of the feedback. The word count in the feedback was around 300 words.

The feedback given by the teacher involved two parts. It involved the teacher's evaluation of the paper in the Word document uploaded by the students and sent back with the edited version. Also, the feedback had a concise note written by the teacher. The note indicated the teacher's overall comments regarding the content and accuracy of the paper and the teacher's further comments. The word count in the teacher's feedback did not exceed fifty words at most.

Language style and register. The analysis of the feedback given by the ChatGPT showed that each feedback was detailed and written using elaborative language. It involved long explanations about the correct or incorrect uses of the forms and appropriateness or efficiency of the content utilizing the complex sentence structures. The feedback also involved the use of a wide variety of vocabulary. Its register has a neutral and formal style. It was written formally by addressing the student in the third person.

The teacher, on the other hand, kept the feedback short and did not use complex sentence structures in it. The use of vocabulary was not advanced or complex, either. The teacher's sentences were short and involved basic vocabulary. The teacher's feedback involved a formal and academic register, but it was also found to be encouraging and supportive. All the students were addressed with 'dear' together with their names in a friendly manner. Also, the teacher's language use involved words to show appreciation and empathy towards the students. Some feedback reflected the closeness between the teacher and the students. The teacher's feedback was thought to target interactivity between the students and the teacher since it involves questions the students are requested to answer in class. Thus, the feedback gave the impression that the teacher elaborated on the feedback in class on the condition that the student had a question or wanted to check the feedback with the teacher during class time.

Table 7

Samples of Written Feedback – Language Style and Register

Control Group	*Dear Batuhan,* *Well done! I liked your ideas very much. Ideas were expressed throughout the paper very effectively. I also appreciate your effort to stick to the topic;)* *Dear Ezgi,* *It seems that you submitted the paper before saving the conclusion part. However, it is not a big problem. No worries!* *Dear İrem,* *Thank you for the submission.* *Grammar needs attention. There are major problems with the basic structures. When you are more careful with the parts I indicated, I believe you can write a better one.* *Dear Aleyna,* *Thank you for your kind words here. You are one of the nicest students a teacher could wish for ;)* *Also, the paper lacks the use of reported speech. If I am wrong, please inform me.*
Experimental Group	*The paper's content explores the importance of learning multiple languages and highlights its benefits, such as enhanced job opportunities and cultural understanding. The writer also discusses personal experiences and the need for practical language practice. The paper touches on relevant points related to language learning and its impact on personal growth. However, the organization of ideas could be improved to ensure a smoother flow and more precise development of arguments. Some ideas could be further expanded and supported with additional examples or evidence.* *Having meticulously assessed the reflection provided by the B2 level student on the grammar course, I am pleased to offer an evaluation of their insights and expression. The student's reflections illuminate both strengths and areas meriting enhancement within the course, effectively employing three instances of inversion to underscore their points.*

Content evaluation in feedback. The feedback given to the content of the writings was analyzed with respect to the differences and similarities between the teacher's and the ChatGPT's evaluation. The analysis indicated that the feedback given to the content by the ChatGPT was detailed and lengthy compared to the teacher's. It involved a summary of the writing. The summary depicted the main ideas in the writing by outlining its introduction, body, and conclusion parts concisely. Also, in the feedback provided by the ChatGPT, some examples to improve the content are provided in detail with explanations. The ChatGPT guides students in reforming their sentences to enhance their clarity by spotting the problem sentences to the point and showing the lack of clarity.

Table 8

Samples of Written Feedback – Content Evaluation

Control Group	*The content is okay but could have been expanded more with more details.* *You have good ideas here. However, they are a little bit disorganized. Not directly relevant to the topic.* *Overall, the content is good. You seem to have done your best to give ideas on the topic. More explanations and details could have been more effective.* *You have good ideas here. However, they are a little bit disorganized. Not directly relevant to the topic.*
Experimental Group	*The paper could be improved by providing more specific examples or instances that illustrate the difficulties faced by the author in learning English grammar or particular aspects of the grammar course. Additionally, the conclusion could be expanded to discuss the importance of grammar in the author's future job and how they plan to continue improving their skills.* *Additionally, providing specific examples of the challenges faced during exams would add depth to the author's reflections. Exploring strategies or steps to overcome exam difficulties would also contribute to a more comprehensive analysis of the author's experience.* *The sentence "It is important to manage time well in terms of finishing dozens of tasks and taking time for yourself" could be rephrased to improve clarity and flow. For example, "Effective time management is essential for completing numerous tasks and making time for personal activities."* *The sentence "The value of learning a new language, besides learning the grammar and word meanings of a different language in addition to your grammar..." could be revised for better clarity and coherence. For example, "The value of learning a new language goes beyond just understanding grammar and word meanings. It allows us to immerse ourselves in new cultures and perspectives."* *Vocabulary: Encourage the use of a wider range of vocabulary to enhance the paragraph's language and sophistication.*

The analysis of the feedback for the content by the teacher indicated that the teacher's feedback expressed an overall thought regarding the content of the students' writings. The feedback regarding the content was formed in short sentences, and the appropriateness and sufficiency of the details were explained.

Accuracy evaluation in feedback. The feedback from the ChatGPT and the teacher also involved the evaluation of the papers for accuracy. The students were required to employ various grammar structures as an expectation of the course requirement. Therefore, both the ChatGPT and the teacher gave the students feedback on using grammar and vocabulary in each paper. Later, the feedback given by both evaluators was compared to each other. Below are the results from the accuracy evaluation in feedback from the ChatGPT presented, followed by the teacher's feedback.

The results from the feedback of the ChatGPT gave an understanding of how the ChatGPT gave feedback regarding the use of the required structures and the overall grammar and vocabulary used in the writing. The findings specified that the ChatGPT feedback on the accuracy of the paper gave an overview of the efficiency of the grammar in the paper. There were explanations related to the use of grammar. The feedback provided information on whether there were any grammatical mistakes in the paper. The students were informed about the problems and missing grammatical structures. However, the feedback was not explicit since it did not spot or mark the problem parts clearly, especially when the problem did not generally hinder the overall meaning of the sentences. In the parts where the ChatGPT fixed or spotted a problem in a sentence, the feedback involved rewriting the sentence with the corrected form of problem parts.

The findings on how the ChatGPT assessed the structures the students were expected to use in their papers indicated some inconsistencies. The feedback gave the students explanations about whether the required structures were employed correctly and adequately in their papers. The feedback involved providing the corrected forms in restated sentences with the corrected forms. However, when the feedback and the papers were compared, it was revealed that there were some inconsistencies between how the required structures were used and how they were expected to function in the sentence. In some papers, the analysis was correct; however, the sample indicating the problem was wrong.

Table 9

Samples of Written Feedback – Accuracy Evaluation

Control Group	*The content is okay but could have been expanded more with more details.* *Not all the expected structures are used.* *Overall, the use of grammar is okay, but I would like to see more complicated forms.* *All the expected structures have been used correctly. Well done.* *Modal verb: +; correlative conjunction: +; conditional: +; noun clause with that: +; noun clause with whether/ if: +; reported speech: +* *In general, the use of grammar is problematic. Please check all my fixings carefully.*
Experimental Group	*The paper demonstrates a strong level of accuracy in presenting ideas. There are only a few minor grammatical errors, and the text effectively conveys the intended message. The vocabulary and sentence structures used are appropriate for the topic.* *The phrase "if a person wants to improve themself" should be revised to "if a person wants to improve themselves."* *The sentence "Professionally, knowing different languages opens up different job opportunities, because in an increasingly globalized world, businesses and organizations seeks people who can speaks and understands different languages" could be revised for better grammar and clarity. For example, "Professionally, knowing different languages opens up various job opportunities, as businesses and organizations seek people who can speak and understand different languages in an increasingly globalized world."* *Some sentences are repetitive and could be rephrased to improve the flow and readability.* *"Opportunities" is misspelled as "oppurtunities."* *In terms of fulfilling the required structures, the paper includes two causative forms ("learning a new language can improve" and "learning a new language will boost") and one reported speech with a wh-question word ("Some people may think what is very important about learning a new language"). However, it does not include a sentence with the passives formed with a modal verb.*

The teacher's feedback on the accuracy of the paper involved a short note giving the teacher's opinion of the overall accuracy of the paper and a detailed correction of the problem based on the use of the Review section of the Office Program Word. Using that program feature, the teacher could spot, correct, or explain the problem parts on the document the students submitted. The teacher's note on the accuracy was short and general, but the feedback in the Word Document was very detailed. The problem structures were highlighted, and the students were informed about the problems with a note in the document. Minor problems, such as the missing articles, were indicated even though they did not interfere with the meaning. The required structures were found and checked with a list of their uses. The analysis did not indicate any discrepancies between the analysis and the report of the required structures.

Discussion

Feedback is deemed potent for success in higher education, and feedback giving and receiving is not just a matter of "stimulus-response routine" (Hattie & Timperley, 2007, p. 103). Both teachers'

and students' skills are required for giving feedback and receiving feedback. As students themselves build their learning, it is the teachers' responsibility to understand that feedback is a part of the balance between students and teachers, and the feedback provided accurately will be helpful for the students to understand, engage, and cultivate the information that students are anticipated to learn (Hattie & Timperley, 2007). When its benefits, such as decreasing subjective evaluation, providing fast feedback, and enabling personalized learning, are considered, ChatGPT is promising for educational assessment. Dragging from the theories and hypothesis of the Interaction Hypothesis, Socio-Cultural Constructivist theory, ChatGPT is considered effective in the feedback process because of its potential to support interactive learning, which is rich in input and encourages practical outputs by making a socially interactive environment possible. In this section, the findings from the study are discussed by considering the previous research and within the frameworks of the Socio-Constructivist Theory and Interaction Hypotheses.

The students' perceptions of AI-generated and human-generated feedback were compared with respect to the students' views, experiences, and emotions of feedback as well as the usefulness of feedback. The results suggest that providing feedback, either employing the ChatGPT or the teacher, did not affect the students' insights of the feedback. Although there was no variation in the students' perceptions when the chatbot or teacher gave the feedback, the content analysis of the feedback provided by two different types of assessors indicated some differences. To begin with, the look of the feedback and the language styles used in the feedback were distinct from each other. The ChatGPT feedback had lengthy, detailed, and formal language; however, in the teacher's feedback, the language style was friendlier and formed of simpler language patterns. The teacher's feedback was found to be reflecting the closeness between the students and the teacher. With the definition of Shute (2008), feedback complexity refers to the amount and the kind of information to be given with the information in the feedback, and long feedback may cause the message to be lost. As Shute (2008) states, the complexity and length of feedback may be ineffective since students may not take heed of the feedback. However, according to Shute (2008), the research on the effectiveness of complexity and length of feedback varies with contrasting findings. While it has been supported to be ineffective in some studies, some other research suggests adverse results and indicates that complex feedback may yield positive results. In the present study, the students' perceptions of feedback did not indicate a difference when the ChatGPT or a human delivered the feedback. Such a result may be interpreted as the complex and lengthy feedback produced by the

ChatGPT did not lead to an ineffective impact on the students. By supporting this notion, Fryer and Carpenter (2006) state that when a chatbot responds to a question in a non-human style, this may challenge a learner with a high language level in a positive way and thereby be helpful for the learner. Carless (2020) underscores the impact of students' role in the feedback process. The learners' characteristics are essential to benefit from the feedback. Depending on a learner's motivation, skills, goals, or experience, the learner can make of the feedback (Chong, 2020). When the findings from the present research are considered, the students' responses to the questionnaire did not indicate differences in their perceptions of the feedback (given by the AI tool or the teacher). However, it should be reconsidered that the students in the present study had a high proficiency level in English. Therefore, understanding the lengthy, complex feedback may not hinder their uptake of the feedback. Aligned with the Input Hypothesis, it might be suggested that the ChatGPT modified the input depending on the students' English proficiency levels by making it comprehensible and challenging. Although the feedback from the ChatGPT was lengthy and complex, it did not affect the students' perceptions of the feedback they received. Still, this finding leaves the question of how AI-supported feedback would affect learners with lower English proficiency levels unanswered. Further studies may investigate individual differences in receiving feedback by comparing AI-supported and teacher-generated feedback.

The present study also indicated that the feedback regarding the content was more effective in the ChatGPT because the AI technology produced ideas to be included in the writing and suggested the students use them in their papers to improve. It provided a clear outline and summary of the students' writings by explaining new ideas that could be added to the paper. Also, there were restated sentences for the ones in the student's writing that lacked clarity. The findings here were also in line with the previous research, which suggested ChatGPT feedback was lengthier than the teachers' and involved summaries (Guo & Wang, 2024). On the other hand, the teacher's feedback provided an overall thought related to idea generation in the writing. New ideas were not given to the students directly by the teacher. Similar findings were also reported by Guo and Wang (2024), who suggested that ChatGPT feedback was documented as more direct than teachers' feedback. Precisely, the teachers led learners to form their writing by asking questions or requesting more explanations. Likewise, the teacher feedback in the study involved comments requiring the students to answer. From the Interaction Hypothesis, the feedback process has the potential to increase the negotiation of meaning to revise the paper and, hence, modify their output. This could also be

considered from the socio-constructivist theory because teacher feedback can enhance the collaboration between students and teachers to construct knowledge. On the other hand, while drawing the conclusion regarding the teacher feedback, stating that ChatGPT feedback would not give the same results would be incorrect. The previous research encapsulating the student and ChatGPT dialogue in the feedback process revealed that students perceive ChatGPT interaction similarly to talking to a friendly and smart peer from whom they receive feedback (Han et al., 2024); by referring to students' comfort in engaging in dialogue with ChatGPT. Therefore, the present study could not display the ChatGPT feedback in a dialogic process. However, considering the previous research, its value as a tool that can provide scaffolding and potentially enhance knowledge construction should not be disregarded.

Rad, Alipour, and Jafarpour (2023) mention in their study that the AI-provided feedback was trustworthy since it was objective to assess learners' writing; hence, the students could rely on that feedback to improve their writing. On the contrary, the results from the present study indicated an inconsistency with the feedback given by the chatbot. The feedback required to be controlled by the teacher so that it could become reliable for the students. As a result, the present study raises suspicion about the reliability of the chatbots for giving feedback, especially for accuracy. Despite the contrasting findings which suggest the effectiveness of ChatGPT for accuracy, reliability, and consistency in scoring writing (Demir, 2023; Mizumoto & Eguchi, 2023), the present study reveals that ChatGPT fails to be reliable and consistent with respect to accuracy checks. Hence, the study contradicts the existing literature. However, while making this conclusion, it is crucial to be careful since various factors such as research designs adopted in the previous ones and the present one, how the research was carried out, or how reliability is interpreted might have brought about the differences in the findings.

Feedback is significant as it offers supportive guidance for writing and provides interpersonal relationships that effectively improve writing (Hyland & Hyland, 2001). Rad et al. (2023) proposed that students' engagement also increases thanks to AI-supported feedback because of the interactivity and confidence the students would have with the dialogic process with a machine. However, Zhang and Mao (2023) explain that AI technology may have a negative effect on teacher-student interaction in the classroom since it may cause students to be less eager to contact their teachers. Regarding interactivity in the classroom, the present study indicated that teacher's

feedback reflected some patterns of interactivity with the students. The questions the teacher noted on the students' papers could have triggered the interaction by provoking students to reply to those questions in class in a follow-up manner. On the other hand, in the feedback produced by the ChatGPT, no similar patterns were found, suggesting that the AI technology might fall behind teacher feedback in this respect, which could lead to a reverse effect on improving writing. Since chatbots do not engage in a real dialogue with the users and cannot understand complex utterances, one may conclude that they lack the interaction aspect even in the written language the teachers use to give written feedback.

Teachers' praise statements tell more than the students' success level in the classroom. They indicate the positive emotional impact of the teacher, such as gladness and interest, and also inform the students about the value of their performance (Brophy, 1981). Brophy (1981) proposes the positive impact of reliable and frank feedback to encourage students to write better. When the appreciation words and sincere words, which indicate the closeness between the students and the teacher, are thought in the present study's findings, the study gives an insight regarding the lack of this fact between the machine-generated feedback and the students. Positive comments may have a positive effect on some students. As Hyland and Hyland (2001) also state, praising fosters language behaviors and increases students' self-confidence. Hyland and Hyland (2001) explain that the positive comments must be specific and connected to the actual text features instead of formulaic general comments. More importantly, they need to be sincere. In the feedback given by the teacher, the patterns showing the close relationship between the students and the teacher were revealed in the content analysis. With this respect, the study may give insights into machine-human relations and the lack of empathy, closeness, and interactivity. However, while making conclusions, it is essential to be careful since the findings here contradict Guo and Wang (2024). In Guo and Wang (2024), the ChatGPT feedback involved more praise than teachers' feedback. This might be because the assessors in Guo and Wang (2024) were not the students' teachers, whereas, in the present study, the teacher gave feedback to her own students. Thus, she was knowledgeable about the students' background and language development levels, which might have caused her to provide more personalized feedback to the students.

According to Shute (2008), the most frequent feedback types are based on elaboration, which involves providing elucidation, supervision using hints and cues, and correcting the problem in the

response. The present study found that elaboration was commonly used in the feedback given to the students' writings in the experimental and the control groups. The AI feedback and the teacher feedback both were based on elaboration. There were especially long explanations of the content of the message in the AI feedback. Unlike the AI feedback, the teacher feedback was shorter. Rather than lengthy explanations, short forms were used, and the success of the writing was mentioned with adjectives to appreciate the work. The teacher used cues and hints. Also, direct correction was more common in the teacher feedback. Thus, although the styles differed in both feedback types, the students received feedback based on elaboration whether the AI technology or the teacher gave the feedback.

Rad et al. (2023) state that AI feedback efficiently uses time. Accordingly, the students do not have to wait for the teacher to give feedback on their work. In the present study, an investigation of the use of AI tools by the teacher was done, and it was clear that using AI did not guarantee a decrease in the teachers' workload since each feedback should be analyzed well and checked for the goals of the writing task. Regarding the feedback produced by AI technology, the analysis indicated that the feedback was composed of formal language and had a format of a few paragraphs. Instead of showing the problem parts, an explanation of the mistakes was given to the students.

Conclusion

The current research investigated the effectiveness of the feedback from the ChatGPT on EFL students' writings compared to teacher feedback. With this respect, the study revealed EFL students' perceptions of the feedback provided by the ChatGPT and the teacher in terms of students' views and expectations of the feedback, experiences with the feedback, emotions, and thoughts regarding the usefulness of the feedback. Additionally, the study indicated the extent to which the ChatGPT feedback and the teacher feedback varied from each other. The findings indicated that the students' perceptions did not change significantly when the feedback was given by the ChatGPT or the teacher. However, the content analysis of the types of feedback revealed differences by proposing that the teacher and the bot had unique strengths over each other. Specifically, the feedback from the ChatGPT seemed to be more effective with respect to the feedback on the content of the assigned writings. The teacher's feedback was found to be more effective for evaluating the accuracy of the papers and having the potential to foster interactivity in the classroom. The findings from the study might suggest that while AI technology has value in being used for giving feedback

on EFL students' writings, it might fall short of replacing human feedback wholly. Therefore, it might be essential to bring the merits of AI tools and the effectiveness of teacher feedback together to help students improve their writing.

Limitations and Suggestions for Future Research

The present study suffers from particular limitations that need to be considered while interpreting its results. Firstly, the students' proficiency level in the present study was B2, according to the CEFR, and the students may not have had trouble understanding the feedback generated by the ChatGPT. When the study is conducted with language learners with different proficiency levels, and when their perceptions of the feedback are compared, results that are more different than those in this study could be derived in future studies. Also, the learners' perceptions were collected through a questionnaire administered online. On the other hand, if semi-structured face-to-face interviews had been carried out with the participants, more diverse results could have been obtained, and a deeper understanding of the participants' perceptions of the feedback could have been revealed, which could help to understand the differences between the feedback from an AI technology and a human better. Therefore, future studies may be conducted using a combination of research and triangulation of data through multiple means of data collection. In addition, only one teacher's feedback was involved in the study. Therefore, generalizing the results from the present study to all teachers and their process of giving feedback would yield erroneous results. Thus, doing research by involving more teacher feedback in the process may be effective in obtaining more reliable results in future studies. Also, the present study indicated that the teacher feedback involved patterns that could encourage teacher-student interaction in the classroom. On the other hand, although the ChatGPT feedback did not include any vivid patterns of requesting the students explain for elaboration, the feedback itself could have triggered such an interaction between the students and the teacher. It was not examined in the current study, so future research might also be conducted to reveal the effects of AI-generated feedback on teacher-student interaction in the classroom. In addition, the present research did not investigate the extent to which the feedback generated by an AI tool or a teacher helps language learners improve their writing. Therefore, future studies may be conducted to analyze the revisions of multiple drafts of writings that receive feedback from a teacher or an AI tool. Thus, the students' understanding of feedback

and how much they could fix the problems in the paper based on the feedback could be revealed more effectively.

Pedagogical Implications

The study provides an understanding of how the feedback processes in writing assessments differ when AI technology is employed compared to a human being. Ignoring the impact and integration of technology in education would be impossible. Chatbots used for educational purposes have the potential to improve learning as a result of activities depending on the learning objectives, so it is vital to integrate them into learning and assessing processes by reducing their limitations and deciding on which chatbot is the best to use in classes. It is significant to understand their potential in the writing evaluation process to support teacher feedback. A combination of AI and non-AI-supported feedback (provided by the instructor or teacher) seems optimal for effectively developing writing. Therefore, educators might benefit from the effectiveness of the bots in giving feedback related to the content of the paper and contribute to students' idea-generation process in writing classes, thereby spending more time on their feedback related to the accuracy check of the papers. It is also essential for curriculum designers and administrators to integrate AI tools to foster writing improvement in school curricula. In this vein, teacher trainers could organize training to equip teachers with the required skills to benefit from the AI tools to enhance learning and teaching processes.

Ethical Considerations

The Ethics Committee approval was received from Çankaya University Social and Human Sciences Scientific Research and Publication Ethics Committee on 07.07.2023 / Number: E-90705970-605.01-131514.

References

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications, 2*, 1-18.

Ahn, S. (2022). The effects of chatbot on grammar competence for Korean EFL college students. *Journal of Digital Convergence*. *20*(3), 53-61.

Algaraady, J., & Mahyoob, M. (2023). ChatGPTs capabilities in spotting and analyzing writing errors experienced by EFL learners. *Arab World English Journal, 9*, 3-17.

Brophy, J. (1981). Teacher praise: A functional analysis. *Review of Educational Research, 51*(1), 5–32. https://doi.org/10.3102/00346543051001005

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The criterion online writing service. *AI Magazine, 25(3), 27.*

Carless, D. (2020). From teacher transmission of information to student feedback literacy: Activating the learner role in feedback processes. *Active Learning in Higher Education*. https://doi.org/10.1177/1469787420945845

Çakmak, F. (2022). Chatbot-human interaction and its effects on EL students' L2 speaking performance and anxiety. *Novitas-ROYAL, 16*(2), 113-131.

Demir, S. (2023). Investigation of ChatGPT and real raters in scoring open-ended items in terms of inter-rater reliability. *International Journal of Turkish Educational Studies, 11*(21), 1072-1099.

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems, 38*, 1-15. DOI: 10.1111/exsy.12667

Fosnot, C. T. (1996). *Enquiring teachers enquiring learners: A constructivist approach to teaching*. Teachers College Press.

Fryer, L., & Carpenter, R. (2006). Emerging technologies: Bots as language learning tools. *Language Learning & Technology, 10*(3), 8-14.

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potent to support teacher feedback in EFL writing. *Education and Information Technologies, 29*, 8435-8463.

Guo, K., Wang, J, & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing, 54*, 1-6.

Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology, 15*(2), ep421. https://doi.org/10.30935/cedtech/13036

Han, J., Yoo, H., Myung, J., Kim, M., Lee, T. Y., Ahn, S., & Oh, A. (2024, May 20-25). *RECIPE4U: Student-ChatGPT interaction dataset in EFL writing education*. LREC-COLING 2024, Torino, Italy.

Harunasari, S. Y. (2022). Examining the effectiveness of AI-integrated approach in EFL writing: A case of ChatGPT. *International Journal of Progressive Sciences and technologies (IJPSAT), 39*(2), 357-368.

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

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning – Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38, 237-257.

Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing*, *10*, 185-212.

Khanna, A., Pandey, B., Vashishta, K., Kalia, K., Bhale, P., & Das, T. (2015). A study of today's A.I. through chatbots and rediscovery of machine intelligence. *International Journal of U- and e-Service, Science and Technology*, *8*, 277–284. http: //dx.doi.org/10.14257/ijunesst.2015.8.7.28.

Kim, N. (2018). Chatbots and Korean EFL students' English vocabulary learning. *Journal of Digital Convergence*, *16*(2), 1-7.

Kovačević, D. (2023, March 15-17). *Use of ChatGPT in ESP Teaching Process*. 2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. New York: Longman.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bathia (Eds.) *Handbook of research on second language*.

Marrs, S. A. (2016). Development of the student perceptions of writing feedback scale. Dissertation for the degree of Doctor of Philosophy. Virginia Commonwealth University. Richmond, Virginia.

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, *2*(2), 1-13.

Okonkwo, C. W., & Ade-Ibijilo, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education, 2*, 1-10.

Rad, H. S., Alipour, R., & Jafarpour, A. (2023). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application, *Interactive Learning Environments*, DOI: 10.1080/10494820.2023.2208170

Szabó, F., & Csépes, I. (2023). Constructivism in language pedagogy. *Hungarian Educational Research Journal*, *13*(3), 405–417.

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

Steiss, J., Tate, T., Graham, Cruz, S. J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, *91*, 1-15.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of William E. Rutherford* (pp. 125–144). Oxford: Oxford University Press.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House.

Tsai, C., Lİ, Y., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies*.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, *10*(2), 157-180.

Yin, Q. & Satar, M. (2020). English as a foreign language learner interaction with chatbots: Negotiation for meaning. *International Online Journal of Education and Teaching (IOJET)*, *7*(2), 390-410.

Zhang, B., & Mao, J. (2023). On the teaching and learning in the information age of "big data + internet?" – Some thoughts on the application of ChatGPT in teaching. In C. F. Peng, A. Asmawi, & C. Zhao (Eds.), *Proceedings of the 2023 2nd International Conference on Educational Innovation and Multimedia Technology (EIMT 2023)* (pp. 1005-1016). Atlantis Press.

Zhang, Z. V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, *43*.

Genişletilmiş Özet

Teknolojideki gelişmelerle birlikte yapay zekâ teknolojileri de gelişti. Bu bağlamda sohbet botlarının yabancı dil öğretimindeki etkinliği pek çok çalışmada araştırılmıştır. Bu çalışmalar, Mike, Elbot, Mondly, Argumate, Replika ve ChatGPT gibi farklı sohbet botlarının yabancı dil öğrenme süreçlerinde olumlu etkilerini ortaya koymuştur (Ahn, 2022; Kim, 2018; Yin & Satar,

2020). Birçok faydasının yanısıra, bu sohbet botlarının, yazma becerisini geliştirme, dil öğrencilerine anında geri bildirim verme, daha az stresli bir öğrenme ortamı sağlama ve kullanıcılardan daha ileri bir dil yeterliliğine sahip olduğundan özellikle iyi bir yazı partneri olma özellikleri dikkat çekmektedir (Guo, Wang, & Chu, 2022). Alan taramasında sohbet botlarının, olumlu etkilerini ortaya koyan çalışmalar bulunmaktadır. Ancak, dil öğrencilerinin yazdıklarını değerlendirmede sohbet botlarının etkinliğini ortaya koyan çok fazla çalışma yoktur. Bu nedenle, bu çalışma sohbet botlarının yazma ödevlerine ne derece etkili dönüt verdiğini ortaya koymayı hedeflemektedir.

Çalışma yarı deneysel bir araştırma desenine sahiptir ve araştırma hedefleri doğrultusunda aşağıdaki sorularını cevaplamayı hedeflemektedir.

- ChatGPT kullanımı öğrencilerin yazılarına verilen geri bildirim ile ilgili düşüncelerini ne ölçüde etkilemiştir?
- ChatGPT tarafından verilen geri bildirim ve insan tarafından verilen geri bildirim ne ölçüde farklılık göstermektedir?

Çalışmaya, 51 üniversite öğrencisi gönüllü olarak katılmıştır. Katılımcıların tamamı özel bir üniversitede Mütercim Tercümanlık bölümünde kayıtlı, 1. Sınıf öğrencileridir. Öğrenciler kontrol ve deney grubu olarak iki gruba ayrılmıştır. Araştırma beş hafta sürmüş ve bu süre boyunca her iki gruptaki öğrencilere de aynı ödevler haftalık verilmiştir. Kontrol grubundaki öğrencilerin ödevleri öğretmen tarafından değerlendirilip, ödevlere geri bildirim verilmiştir. Deney grubunda ise, öğrenci ödevleri ChatGPT tarafından değerlendirilip geri bildirim verilmiştir. Geri bildirimler öğrencilerle paylaşılmış ve beş haftanın sonunda öğrencilerin geri bildirimlerle ilgili görüşleri verilen anket yoluyla toplanmıştır. Toplanan veri istatistiksel yöntemlerle karşılaştırılmıştır ve öğrencilerin ChatGPT ve öğretmen tarafından verilen geri bildirim arasındaki farklar ile ilgili görüşleri ortaya çıkarılmıştır. Bununla birlikte, yazı ödevlerinin değerlendirmesinde ChatGPT ve öğretmen tarafından verilen geri bildirimin özelliklerinin ne derece farklılaştığını göstermek için geri bildirimler nitel yöntemler kullanılarak analiz edilmiştir.

İstatistiksel veri analizleri, geri bildirimin, ChatGPT veya insan tarafından verilmesinin öğrencilerin düşüncelerinde bir etkisi olmadığını göstermiştir. Buna rağmen, nitel veri analiz sonuçları geri bildirimle ilgili farkları ortaya koymuştur. Buna göre, öncelikle geri bildirimin

öğrenciye sunumu farklıdır. ChatGPT geri bildirimi uzun ve detaylıyken, öğretmen tarafından verilen geri bildirim, öğretmenin yazının içerik ve dilbilgisi ile ilgili genel görüşünü gösteren kısa bir nottan oluşmuştur. Buna ek olarak, öğretmen WORD dosyası üzerinde öğrencilerin hatalarını direk notlar düşerek gösterdiği bir dosyayı öğrencilerle paylaşmıştır. ChatGPT geri bildirimi, çok daha karmaşık dil yapılarıyla oluşturulmuş ve yanlış dil bilgisi kullanımıyla ilgili açıklamaları içermektedir. Stil olarak resmi ve nötrdür. Diğer taraftan öğretmen tarafından verilen geri bildirim kısadır ve basit dil yapıları içermektedir. Stil olarak, öğretmen geri bildirimi de resmi ve akademik bir yapıya sahip olmasına rağmen öğrenciyi teşvik edici, destekleyici bir ifade şekli de vardır. Geri bildirimde, yazının içeriğine yönelik ChatGPT'nin verdiği geri bildirim yazının özeti ve ana fikirlerini de gösteren kapsayıcı bir özelliğe sahipken, öğretmen tarafından içeriğe yönelik verilen geri bildirim öğretmenin genel görüşünü içermektedir. Ayrıca, yapılan analizler, ChatGPT ve öğretmen tarafından yazılardaki dilbilgisi yapılarına verilen geri bildirimin özelliklerini de ortaya koymuştur. Buna göre, ChatGPT'nin geri bildirimi dilbilgisi kullanımı ile ilgili açıklamaları içermektedir. Öğrencileri dilbilgisi yanlışlarıyla ilgili bilgilendirmiştir ve problemli yapıları özellikle anlamla ilgili sorun oluşturduğunda göstermiştir. Ancak, ChatGPT geri bildiriminde öğrencilerin kullanması beklenen yapıların sorgusunda tutarsızlıkların olduğu ortaya çıkmıştır. Öğretmenin dil bilgisi kullanımına yönelik geri bildirimi daha detaylı bulunmuştur ve ChatGPT'deki tutarsızlıklar öğretmen geri bildiriminde gözlenmemiştir.

Sonuç olarak, bu çalışma insan tarafından ve bir yapay zekâ teknolojisi tarafından verilen geri bildirimin birbirlerinden ne kadar farklı olduklarını ortaya koymuştur. İstatistiksel testlerden elde edilen sonuçlar, öğrencilerin algılarının geri bildirimin yapay zekâ tarafından ya da insan tarafından verildiğinde değişmediğini göstermiştir. Ancak içerik analizi, geri bildirimin dil stili, dil biçimi, içerik ve dil bilgisi doğruluğu açısından farklılıklar gösterdiğini ortaya koymuştur. Bu çalışma, dil öğrenme ortamlarında yazma geri bildirimi sağlamak için sohbet botlarının etkili olduğu yönleri göstererek, bu botların eğitime entegre edilmesi konusunda bir bakış açısı sunmaktadır.

ETİK BEYAN: "Yabancı Dil Olarak İngilizce Bağlamında Yazma Ödevlerini Değerlendirmede Sohbet Botlarının Etkinliği Üzerine Bir Araştırma" başlıklı çalışmanın yazım sürecinde bilimsel, etik ve alıntı kurallarına uyulmuş; toplanan veriler üzerinde herhangi bir tahrifat yapılmamıştır ve veriler toplanmadan önce Çankaya Üniversitesi Sosyal ve Beşerî Bilimler Bilimsel Araştırma ve Yayın Etiği Kurulu'ndan 07.07.2023 tarih ve E-90705970-605.01-131514 sayılı etik izin alınmıştır.

Karşılaşılacak tüm etik ihlallerde “Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi Yayın Kurulunun” hiçbir sorumluluğunun olmadığı, tüm sorumluluğun Sorumlu Yazara ait olduğu ve bu çalışmanın herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiş olduğunu taahhüt ederim."