ANADOLU UNIVERSITY | GRADUATE SCHOOL OF EDUCATIONAL SCIENCES
AJESI
ANADOLU JOURNAL OF EDUCATIONAL SCIENCES INTERNATIONAL

DOI: 10.18039/ajesi.1463503

ChatGPT: Is It Reliable as an Automated Writing Evaluation Tool?

Saliha TOSCU[1]

Date submitted: *02.04.2024* ***Date accepted:*** *16.10.2024* ***Type[2]:*** *Research Article*

Abstract

This study primarily aims to give an understanding of whether or not teachers could rely on AI technology, specifically ChatGPT, to score students' writings. The study was conducted with the participation of EFL university students. The students were assigned different writing tasks for five weeks, and the tasks were scored by a teacher and ChatGPT separately. Then, their scores were compared to see the extent to which ChatGPT and teacher scores differed on the SPSS. The test results indicated no statistically significant differences in the scores the bot or the teacher gave. Additionally, the results were supported by the qualitative analysis of the teachers' perception of ChatGPT use for automated writing evaluation. The teachers' perceptions indicated their positive attitudes towards its use for the evaluation process and general use for enhancing instruction and learning, together with the concerns and suggestions to make the most of ChatGPT. The study gives insights into the integration of ChatGPT into the assessment process and its effectiveness for class practices.

Keywords: artificial intelligence, assessment, ChatGPT, writing

Cite: Toscu, S. (2025). ChatGPT: Is it reliable as an automated writing evaluation tool? *Anadolu Journal of Educational Sciences International, 15*(1), 329-349. https://doi.org/10.18039/ajesi.1463503

[1] Dr., Çankaya University, Foreign Languages Department, Turkey, salihatoscu@cankaya.edu.tr , https://orcid.org/0000-0002-8179-5444

[2] This research study was conducted with Research Ethics Committee approval of Çankaya University, dated 11.07.2023 and issue number E-31115241-050.99-131681.

[3] Some of the research findings of this study were presented at the 15th METU Convention (May 2024).

Introduction

Artificial Intelligence (AI) has had a significant impact worldwide. For various purposes, it has been integrated into diverse sectors, such as business, health, and education. Numerous studies have discussed and revealed its positive impacts on education. Accordingly, AI-supported learning positively affects learners' performance and achievement in the course, learning motivation, and self-efficacy (Chiu, Hwang, Hsia, & Shyu, 2022). Because AI technologies have a positive influence on learning success and perception, their use is supported in learning at all education levels by considering "the sample size, learning domains, the types of organization, and AI software and hardware" (Zheng, Niu, Zhong, & Gyasi, 2021, p. 12). AI is used in language learning because it is considered to affect language acquisition positively by improving listening, speaking, and reading skills and having a salutary effect on students' feelings due to an authentic, meaningful interaction in a congenial atmosphere (Wang, Pang, Wallace, Wang, & Chen, 2022). Tai and Chen (2023) note that AI technologies, such as intelligent personal assistants, could benefit language learning by improving users' listening and speaking skills. Besides, they can contribute to users' willingness to communicate because they comply with the users' directives (Tai & Chen, 2023). Learners using AI technologies are able to benefit from personalized learning, the chance to practice interaction, and correct their mistakes thanks to the immediate feedback they receive from AI (Tai & Chen, 2023). Also, Moussalli and Cardoso (2020) explain that AI agents are indulgent, so learners using these agents do not get bored as a person would in human-to-human interactions, and they do not hesitate to ask questions continuously. Thus, it has become undeniable that AI technologies effectively offer prospects for useful, engaging, and encouraging learning.

One of the prospects has been related to the writing process. There is an increase in the use of automated writing evaluation systems, such as Grammarly, Criterion, and MY Access (Koltovskaia, 2020; Li, Link, & Hegelheimer, 2015). They have been noted as contributing to developing writing by giving opportunities to practice and revise a piece of paper more than once, providing more reliable and consistent ratings than a human, and giving more time to teachers to focus on content (Koltovskaia, 2020; Li, Link, & Hegelheimer, 2015). Attributable to their advantages, AI tools for improving writing performance have been quickly integrated into writing classes.

A number of AI technologies have been in use, and these days, Chat Generative Pre-Trained Transformer (ChatGPT) has spread far and wide. ChatGPT has become a new form of AI technology, a chatbot that can communicate with people due to its competencies in natural language processing and generation (Liu, Zhao, Sun, Zhang, Kou, & Gai, 2023). The bot has been integrated into educational settings and obviously has the capacity to change the present form and realization of teaching (Zhang & Mao, 2023). It takes commands from the users and does what they ask. It also allows users to prepare tests and assess and score performance. Today, in educational areas, students and teachers use it as a conversation partner, checker, and assessor. Lately, studies have investigated its use in language learning environments and indicated its positive impacts on diverse aspects of language learning. To exemplify, Qu and Wu (2024) revealed a positive effect of ChatGPT on students' intrinsic motivation to learn English as a second language through engaging and enjoyable interactions. In another recent study, Kucuk (2024) investigated the advantages and disadvantages of ChatGPT for teaching and learning grammar and explored the positive impact of using ChatGPT for grammar learning and teaching. Susanto et al. (2024) examined how ChatGPT could impact pragmatic instruction and revealed that it could be effective in designing teaching

materials. In Avsheniuk et al. (2024) study, the effects of ChatGPT on critical thinking and EFL learning were investigated; the researchers showed that learners' analytical thinking skills might be increased through the bot due to the bot's contextual and reflective prompts. In addition to the positive influence of using ChatGPT on motivation, critical thinking skills, grammar teaching, and pragmatics instruction, its potential to enhance also language skills such as reading, vocabulary, and writing because of the feedback it provides on students' performance was explored in the previous studies (Bin-Hady et al., 2023). These findings depict the contribution of ChatGPT to diverse aspects of language learning and language skills, including writing.

Assessing Writing and AI Technologies

The spread of English as a global language has made writing ability necessary in language learning programs in education (Warschauer & Ware, 2006). However, writing is a skill that requires a lot of time to teach and evaluate, so automated writing evaluation programs have been meeting this need by eliminating the challenges.

Warschauer and Ware (2006) explain that to improve writing effectively, it is necessary to give learners individual feedback for numerous drafts. This process could be time-consuming for teachers when big class sizes are considered. As a result of the improvements in technology, automated writing evaluation tools have started to be used in education by having an essential impact on writing instruction (Wang, Shang, & Briody, 2012). The use of automated writing software provides a range of benefits, such as learner autonomy and more opportunities for writing practice (Wang et al., 2012).

Some studies have investigated how automated writing software affects learning. They indicated that its use could expand writing performance as learners write different drafts (Stevenson & Phakiti, 2014), improve accuracy (Li et al., 2015), and foster learner autonomy by enabling learners to see their progress (Dikli, 2006). It also has the potential to assist the teacher by decreasing the time of grading and providing feedback (Wang et al., 2022).

Developing writing skills at university is a requisite since it involves a big part of communication in educational activities at university. Assessing writing requires objectivity and reliability to ensure an accurate performance assessment. While AI technologies affect all areas of life, their use is also enhanced in education. As the literature suggests, AI technologies serve different educational purposes and positively affect educational processes. Besides all the positive effects of these technologies, the problems related to privacy, ethics, and reliability issues are raising concerns, and all these worries have been revealed in different studies (Adamopoulou & Moussiades, 2020; Halaweh, 2023). One positive aspect of AI technologies is that their system effectively reduces the teachers' workload and provides an understanding of their students' learning outcomes (Chiu et al., 2022). Writing assessment and scoring writing performance are also functions that educators could benefit from while using AI technology.

The value and effects of ChatGPT as a learning assistant for improving skills, primarily writing, have been investigated in diverse studies. Studies indicate that ChatGPT produces helpful feedback and helps in the planning and idea-generation stages of writing. Guo and Wang (2024), for example, compared teacher feedback to ChatGPT feedback on students' writing in an EFL context and indicated the aspects both feedback focused on. Likewise, Tsai and Brown (2024) investigated how ChatGPT-assisted feedback could improve the quality of

essays. This study showed that ChatGPT feedback helped students improve their writing and get higher scores. The revisions based on ChatGPT might bring about revised texts that sound more natural concerning tone, include more varied vocabulary, and have adequate grammatical accuracy. In a different study, Algaraady and Mahyoob (2023) examined how ChatGPT could detect errors in writing and explored that ChatGPT could spot especially surface-level mistakes so it can be effectively used for error analysis in writing. Similarly, Escalante, Pack, and Barrett (2023) researched whether the students' essays developed more with respect to linguistics when ChatGPT or tutor feedback was received. The results showed the efficacy of the bot, and the researchers suggested incorporating it in the evaluation process in writing. Previous research also indicated that ChatGPT is successful at helping language learners correct accuracy and spelling and generate ideas (Harunasari, 2022). Additionally, Alberth (2023) investigated the potential advantages and disadvantages of ChatGPT for academic writing. The study suggests that the bot assists in writing and generating ideas for finding a research topic.

The attitudes and perceptions of both educators and learners towards utilizing ChatGPT for teaching writing in language learning have been explored in recent studies. Mohamed (2024) investigated what faculty members thought about ChatGPT to improve language learning through interviews. The results indicated positive opinions of the effectiveness of ChatGPT in giving fast and correct answers to the questions. In addition, Guo and Wang (2024) noted teachers' positive and negative perceptions of ChatGPT. Specifically, the teachers found the bot helpful in giving feedback and praising features of the bot. The teachers also stated that it could reduce their workload. Likewise, Evmenova, Broup, and Shin (2024) examined the drawbacks and benefits of ChatGPT that the teachers perceived. Solak (2024) researched both language learners' and teachers' experiences with ChatGPT use and explored the positive effect of ChatGPT on shy students by helping them speak with instant input and translation. The study indicated that both learners and educators held positive attitudes toward ChatGPT use. Also, Ho (2024) investigated language learners' perceptions and attitudes toward ChatGPT and uncovered its effectiveness in assisting learners in translation, checking grammar, paraphrasing, and learning vocabulary. Similarly, Escalante et al. (2023) investigated the students' perceptions of the feedback when received from a human tutor and the ChatGPT. They found that the students receiving feedback from the humans and the ChatGPT stated they would benefit from the feedback.

The aforementioned studies suggest the benefits of using ChatGPT for writing instruction and learning in language learning settings. However, it is also essential to acknowledge the constraints of the bot proposed by the researchers. Alberth (2023), for example, underscores that "plagiarism, the potential for degraded researcher autonomy, and the threatened academic integrity" might be problems relevant to the use of ChatGPT and asserts that these constraints should be considered while implementing ChatGPT (p. 349). Similarly, Mohamed (2024) attracted attention to the "ethical and practical concerns such as the accuracy of AI-generated responses, the ethical implications of content, and the risk of AI replacing human educators" (p. 220). Moreover, the disadvantages, such as a decrease in human interaction and personalization, the possibility of receiving false and unsuitable responses, meaning problems, lack of pronunciation and intonation feedback, and cultural insensitivity, are noted to be essential to consider as potential drawbacks of the bot (Mohamed, 2024).

Although extensive research shows the merits and obstacles of using ChatGPT in writing, these studies focus primarily on writing feedback and perceptions regarding how ChatGPT could be utilized to improve the writing process. Thus, the literature lacks research on the effectiveness of using ChatGPT for scoring writing tasks. In some studies, concerns about the fairness of the scores given by ChatGPT were voiced (Tsai & Brown, 2024); hence, more research is crucial to investigate how ChatGPT could be effective in scoring writing tasks. The present study intends to reveal whether AI technologies could be relied on while scoring learners' writing performance in educational settings and to examine teachers' perceptions from this perspective.

Research Questions

In alignment with the purposes of the present study, the following research questions are addressed in the study:

- To what extent do the scores given by the ChatGPT and the teacher differ?
- What are the EFL teachers' perceptions of ChatGPT as an automated writing evaluation tool (AWET)?

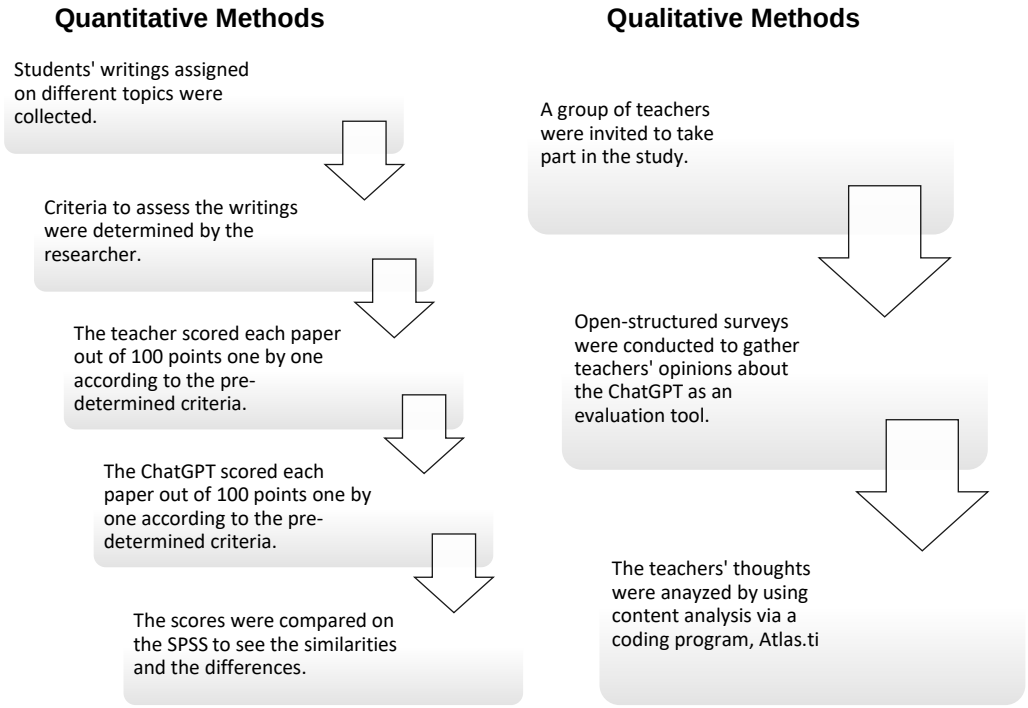
Method

The current study aims to indicate the differences between the scores of ChatGPT and a teacher for writing tasks and to explore teachers' views of using ChatGPT as an automated writing evaluation tool. Below are the details related to the research design, participants, as well as data collection and analysis procedures presented in detail.

Research Design

The present research adopted a mixed-methods approach. Both quantitative and qualitative methods were employed in data collection and analysis procedures. Quantitative methods were used to compare the scores given to EFL students' writings by the ChatGPT and a human teacher, while qualitative methods were employed to collect and analyze language teachers' subjective views of using the ChatGPT as an evaluation tool.

Figure 1
Research Design



As the figure above indicates, the research design involved collecting students' writings, independently scoring each paper by ChatGPT and a teacher, and contrasting the scores using statistical methods on the SPSS. Also, so as to see the reliability and practicality of using the ChatGPT from teachers' perspective, teachers were administered an open survey, the findings of which were analyzed through thematic analysis using Atlas.ti software.

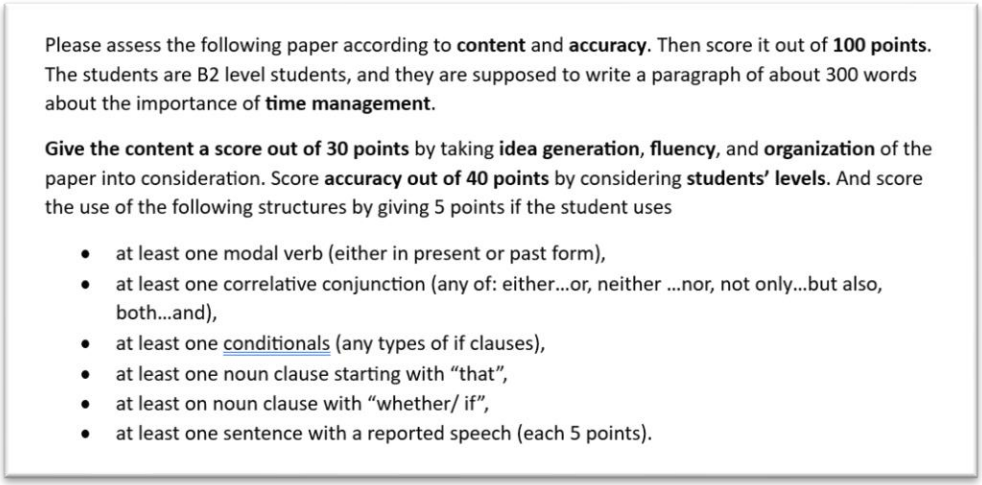
Participants and Procedures

51 English as a foreign language (EFL) learners and 31 EFL instructors from different universities were involved in the study voluntarily by means of convenience sampling. The students were all first-year students at the English Language and Interpretation Department at a private university in Türkiye. Their departments had a language barrier, which was ensured as 80 points out of 100 in the proficiency exam of the institute or with a TOEFL score (85 out of 100 points), which had equivalence with the predetermined score of English at the proficiency exam. The students were assigned five writing tasks as a part of the course they enrolled in. Using specific grammar structures, the students were asked to write an opinion paragraph based on various topics in 300-350 words. Later, the ChatGPT and the teacher scored the students' writings independently to explore the differences in the scores both assessors gave.

Both the teacher and the ChatGPT based their assessment on the same criteria. The criteria involved an analysis of the content and accuracy of the papers. Each assessor evaluated the papers by considering the same criteria and scored them out of 100 points. Accordingly, the content of the papers was scored out of 30 points, considering the idea generation, fluency, and organization of the papers. Due to the course objectives, the accuracy of the papers was given more importance, and more scores were allocated to it. An effective

use of overall vocabulary and grammar in the paper was scored out of 40 points. Additionally, 30 points were given to the students when they accurately used assigned grammar structures in their papers on a weekly basis. The structures were determined depending on the weekly taught grammar structures.

Figure 2
ChatGPT Prompt Sample with the Assessing Criteria


Please assess the following paper according to **content** and **accuracy**. Then score it out of **100 points**. The students are B2 level students, and they are supposed to write a paragraph of about 300 words about the importance of **time management**.

Give the content a score out of 30 points by taking **idea generation**, **fluency**, and **organization** of the paper into consideration. Score **accuracy out of 40 points** by considering **students' levels**. And score the use of the following structures by giving 5 points if the student uses

- 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 on noun clause with "whether/ if",
- at least one sentence with a reported speech (each 5 points).

The teacher scored each paper using the rubric with the criteria. Likewise, to ensure the teacher's and ChatGPT's similarity, the prompt specifying the criteria in detail was entered into the ChatGPT (Figure 2). The prompt was made as detailed as possible. As Figure 2 indicates, the details such as students' levels, assessment criteria, the total score and the scores for each criterion, the word count, the topic assigned to the students, and the grammar structures the students were expected to use in the paper were logged to the bot in detail. Later, the scores given by both raters were compiled and prepared for statistical tests. By employing quantitative research methods, the data were analyzed using SPSS.

The study also intended to analyze the teachers' perceptions of using ChatGPT as a writing evaluation tool. For this purpose, 32 EFL teachers working as EFL instructors at the university level volunteered to take part in the study. All the participant teachers had been working at different universities and had teaching experience ranging from 12 to 20 years. All had a foreign language writing experience. Except for one instructor who stated to have received training for using an automated writing corrective feedback or evaluation tool, all the others explained that they had never had training regarding its use. The participant teachers responded to an online survey with open-ended questions (Appendix A). The survey involved three parts aiming to collect demographic information about the teachers, reveal details regarding the teachers' prior experience with automated corrective feedback or evaluation tools, and learn about the teachers' experiences and thoughts about using ChatGPT as an automated writing evaluation tool. The teachers' responses were collected online and then analyzed qualitatively.

Data Collection and Analysis

The data collection started in the Spring Term of the 2022-2023 Academic Year. The students were assigned five writing tasks as required in the English Grammar in Context course. The study collected 127 papers written in English within five weeks (from May to June 2023). The students submitted their writings online through MOODLE, an educational platform used at the university, and the same teacher scored each paper and gave feedback back to the students. Later, all the papers were downloaded from the platform, and ChatGPT was requested to score each paper. The scores were then entered into the SPSS for data analysis. Paired-sample *t*-tests and descriptive statistics were performed to compare the scores given by the teacher and ChatGPT.

For the second research question, a survey with open-ended questions was addressed to 32 EFL teachers online. The survey gave information about the teachers' demographic information. It aimed to reveal the teachers' general thoughts about the automated writing evaluation and corrective feedback tools, specifically their experiences with ChatGPT as a computerized evaluation tool. The data collected through online surveys were analyzed with a content analysis via a coding program, ATLAS.ti. The program provided a systematic analysis of the frequency of the codes. Firstly, all the qualitative data were prepared for pen-paper coding. After grasping the general categories and the codes, the data were transferred to the ATLAS.ti program for detailed coding. Another researcher with a Ph.D. in the language teaching department assisted with inter-rater reliability. Two researchers handled the whole document separately first; later, the codes and their sub-codes were discussed for inclusion in the research. In case of disagreements, the two researchers discussed and agreed on including or excluding the (sub)codes.

Ethical Issues

In the present study, written and verbal consent, which explained the purpose and length of the study, was taken from all teachers who volunteered to be involved in the study. The participants were informed that the data to be collected would only be used for research purposes, and the participants' names would never be shared in any parts of the research.

Also, an ethics committee approval was obtained from Çankaya University Ethics Committee Board to research the effectiveness of chatbots in assessing students' writings on 11.07.2023 (Number: E-31115241-050.99-131681).

Findings

Based on the research purposes of the present study, the findings regarding the extent to which the scores given by the teacher and ChatGPT differed and the teachers' perceptions of ChatGPT as an automated writing tool were presented below.

To what extent do the scores given by ChatGPT and the teacher differ?

The first research question investigated the extent to which the scores ChatGPT and the teacher gave differed. For this purpose, a Paired-Sample *t*-Test and Descriptive Statistics were performed on the SPSS, and the scores from ChatGPT and the teacher for the tasks

assigned within five weeks were compared. When the overall scores were compared, the results indicated no statistically significant difference between the scores given by the teacher (*M*= 89,4, *SD*= 17.7) and the ChatGPT (*M*= 88,4.6, *SD*= 8.3), *t*(7) = .46, *p*>.00005 (two-tailed).

When the comparisons were made concerning the tasks, the results indicated that the scores of Task 3 (Teacher: *M*= 89.5, *SD*= 5.89; ChatGPT: *M*= 87.3, *SD*= 6.44), Task 4 (Teacher: *M*= 91.5, *SD*= 5.12; ChatGPT: *M*= 84.6, *SD*= 10.8), and Task 5 (Teacher: *M*= 92.6, *SD*= 7.23; ChatGPT: *M*= 88, *SD*= 8.3) did not indicate a statistically significant difference as similarly to the total scores, *p*> 000.5. On the other hand, in the comparison of the scores given to Task 1 (Teacher: *M*=89, *SD*=10.1; ChatGPT: *M*= 84.2, *SD*= 7.48) and Task 2 (Teacher: *M*= 79, *SD*= 18.8; ChatGPT: *M*= 83.5, *SD*= 17.10), a statistically significant difference was found between the scores, *p*< .0005. The eta squared statistics of Task 1 indicated a large size effect (.22). In contrast, Task 2 showed a moderate size effect (.07), suggesting that the scores given to the writings differed substantially, especially for Task 1 when the teacher or ChatGPT evaluated the writings. Considering the mean scores, the scores the teacher gave to Task 1 were higher than the ChatGPT's, whereas in Task 2, the teacher gave lower scores to the students' writings than ChatGPT.

Table 1

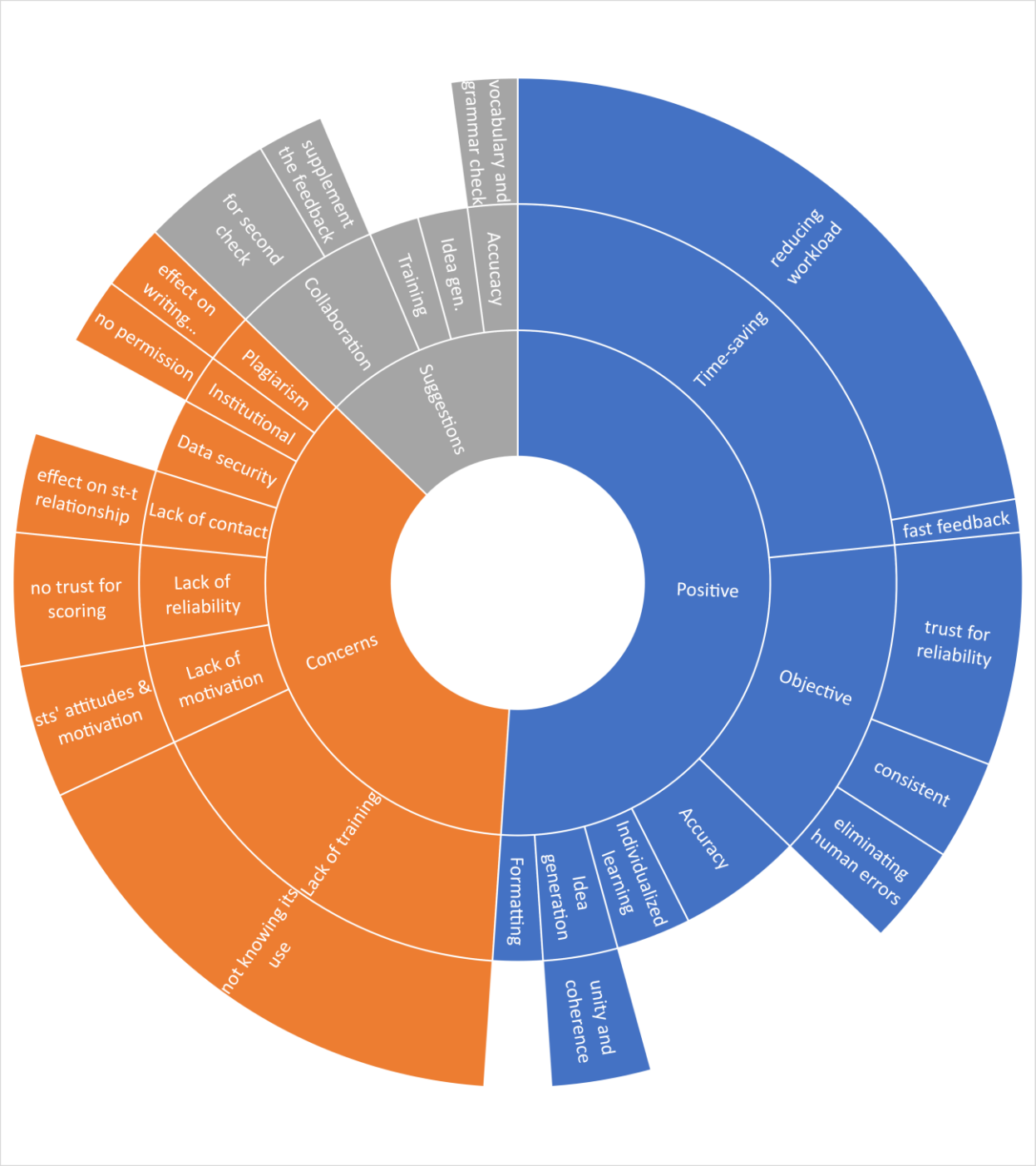
The Differences between ChatGPT scores and the teacher

		N	M	SD	Sig (two-tailed)
Task 1	Teacher Score	29	89	10.1	.02
	ChatGPT Score		84.2	7.48	
Task 2	Teacher Score	45	79	18.8	.00
	ChatGPT Score		83.5	17.10	
Task 3	Teacher Score	21	89.5	5.89	.57
	ChatGPT Score		87.3	6.44	
Task 4	Teacher Score	18	91.5	5.12	.62
	ChatGPT Score		84.6	10.8	
Task 5	Teacher Score	14	92.6	7.23	.69
	ChatGPT Score		88	8.3	
Overall	Teacher Score	127	89,4	17.7	.65
	ChatGPT Score		88,4	8.3	

What are the EFL teachers' perceptions of ChatGPT as an automated writing evaluation tool?

The second research question specifically aimed to indicate the teachers' perceptions of using ChatGPT as an automated writing evaluation tool. All the data collected through the survey were handled first, then the categories were formed, and the codes and sub-codes inside each category were determined. The graphic below indicates the categories and the codes. The inner circle summarizes the broader categories, and the outer circle summarizes the individual codes and their sub-codes by indicating their relationship with each other and the broader codes. Accordingly, the findings indicated positive attitudes towards using the ChatGPT, whereas there were concerns regarding its use. The data also presented the teachers' suggestions about its practical use in education.

Figure 3
Categories and Codes



Positive Attitudes

Through survey questions, teachers' beliefs and thoughts related to ChatGPT use for assessing students' writings were aimed to be revealed. The teachers' responses reflected their positive attitudes towards its use due to its effects on time management, providing students with individualized feedback, and use for idea generation in writing.

The responses to the survey questions indicated that 75 % of the teachers reported not having used ChatGPT for assessing writing. 25 % of the teachers said they had tried to use it briefly just to check how it worked. Although only a few teachers asserted to have employed the chatbot for a short time, their responses indicated that they held positive attitudes toward using ChatGPT based on their assumptions.

The teachers reported that the chatbot could be time-saving. This was the most frequently coded positive aspect of ChatGPT in the analysis. The reports indicated that it could enable teachers to manage their time more effectively since it had the potential to reduce workload. Related to that, some teachers voiced, 'I am for such tools. They enable language teachers to save time and energy.' In another teacher's quote, it was clear that the teacher was for the chatbot as he stated, 'Sometimes it may take too long to check students' writing assignments, so AWET tools may reduce the workload. They would help to check a large number of papers, especially.' It was also evident that the teachers believed ChatGPT could be used to devise test questions. One teacher reported that the bot could be asked to prepare questions provided the questions were revised and edited later, as seen in the quote: 'Sometimes, I used it for devising items for testing, which requires much adaptation on the items.'

The findings indicated that the teachers perceived the potential use of ChatGPT as an automated writing tool since it would enable objective and reliable scoring. One teacher, for example, stated that ChatGPT would be the most reliable method because it would score the papers without any prejudices or difficulties in concentration during the writing evaluation process, as seen in the teacher's comments, 'I believe it will be the most reliable method. AI does not have prejudice, and it does not have a foggy short memory. Given the right rubric, it will outperform us in the near future. However, the level of feedback should also be determined beforehand.' In the same vein, another teacher said: 'To be honest, I find an automated tool to be more reliable than a human being whose mood changes, tiredness, preconceptions, some personal beliefs, possible misjudgments, attitudes towards students or lack of knowledge to understand and implement a grading scale well can hinder the process of grading a written product by students.' Thus, as the teachers' comments make clear, some teachers find the use of ChatGPT reliable, and it can be more trustworthy than human raters since it would not possess humane features such as fallacies or exhaustion.

Parallel to the findings related to the objective scoring, the teachers uttered positive aspects of using ChatGPT for writing as consistent scoring. Some teachers expressed that it would help to diminish human errors, and more consistent scoring would be possible with the ChatGPT scoring. To illustrate, one of the participant teachers said: '….It also eliminates the human errors caused by tiredness or simple carelessness. With too many papers to grade, the possibility of inattentive or imprecise ways of giving feedback or scoring arises. Thus, it would be more objective and fair to grade a writing paper with a tool that focuses on the same things on every paper.' A different teacher also expressed that she believed using ChatGPT was more reliable and consistent concerning scoring; as the quote states, 'It can help with time management and evaluate students' papers more objectively and consistently.'

Another point the teachers were optimistic about using ChatGPT was that it could be suitable for checking the accuracy of the paper as well as its structure and formatting. Regarding the accuracy check, one of the teachers, for example, said: 'ChatGPT may eliminate some redundant work of the teacher, especially regarding grammatical mistakes in writing.' In another teacher's quote, it was clear that she tried and enjoyed using it for an

accuracy check. Her positive attitude towards the tool was evident in her sentence: ‘I have tried it once, and I was fascinated by its accuracy in grading the paper.’ Based on her assumptions, one teacher reported: ‘Also, it could help to structure and format checking.’ Although she was one of the teachers who had not employed the bot before, her comment showed she believed the bot could effectively check the structure and formatting of the writing paper. Some other teachers also said it would be easy to check the structure and format by using ChatGPT.

The survey analysis also showed that some teachers believed the feedback given by ChatGPT would provide learners with individualized feedback on their writing. The teachers also conveyed that tailored student feedback would be given on these terms. One teacher quoted the importance of tailoring feedback based on the students’ needs and the potential usefulness of ChatGPT for tailored feedback: ‘I think the students can receive feedback for their writing papers depending on their needs, and it would be given at the right time and appropriate amount. ChatGPT can be asked to check the papers regarding their content, grammar, and style.’ A different teacher also commented that although she was not sure about whether it was adequate for providing individualized feedback on content, it could be used for individualized feedback, as it was evident in the quote: ‘It could be helpful to support especially the autonomy of student learning and is a time-saver. Also, it could help with structure and format checking. However, I am unsure whether ChatGPT could provide individualized feedback on content.’

Additionally, one common finding was that teachers believed ChatGPT could be most practical at the idea generation step for learners. By cultivating their writing skills, learners can benefit from the bot by asking it to generate ideas for specific topics. Thus, the bot was most likely to serve as an assistant for learners to improve their writing. One teacher voiced: ‘It could provide more time for teachers to focus on the content, idea development, and organization with more detailed feedback about every aspect of a paper.’ Another teacher said he used it in his classes. He explained: ‘While teaching certain paragraph types, I see some students are stuck because they do not have any ideas about the topic. Then I ask them to use ChatGPT to develop relevant ideas.’

Concerns

The analysis showed that the teachers voiced concerns regarding ChatGPT use in general education and evaluation procedures. Among the concerns related to ChatGPT use, the lack of training was the most repeated concern revealed in the data analysis. The teachers said that because they had not experienced its use before and had not received any training, they expressed worry. One teacher expressed his concern regarding its use by stating: ‘I think that it has great potential, yet how to use it is a serious matter. Thus, I am still cautious about it.’ In another quote, the teacher explained: ‘I think I need more training, and there seem to be more features to be added to ChatGPT.,’ and the other one said: ‘I do not have enough information or training on how to use ChatGPT.’ Considering the teachers' comments, it could be thought that they were cautious and lacked information about how to employ ChatGPT.

Also, data security and privacy were concerns because the teachers stated they were unsure how the data would be stored or used when uploaded to the bot. One of the teachers said: ‘As we need to upload students’ work on ChatGPT, data privacy can be a challenge.’ Another teacher expressed: ‘Data privacy is another dimension that must be kept in mind. All

those students' papers and scores will be stored.' As a cause of concern, the teachers noted data security and privacy issues were to be considered. Additionally, regarding the assessment, the teachers believed the bot failed to assess the writing reliably and could not be trusted to score the papers. One teacher said, ' However, I have never used it for scoring. I do not think I can trust it.'

The other concern involved the teachers' apprehensions over how ChatGPT could impact the contact between the teacher and students. The responses showed that some teachers thought its use would decrease communication between teachers and students in education, negatively affecting the relationships. One of the teachers said, 'However, apart from all these things, the biggest disadvantage of automated scoring would be the absence of human contact and motivation.' A teacher reported that the feedback given by a tool cannot equate with the teachers' feedback as explained by the teacher: 'A tool cannot motivate a student when a poor paper is submitted, or praise enough when there is a good work.' Besides, students' attitudes and motivation ought to be considered since using the bot would cause them to have reverse thoughts regarding the education system and decrease their motivation. One teacher commented: 'As for the disadvantages, such tools may demotivate learners since they get feedback from a source rather than their teachers. …however, determining how and when to use them is of utmost importance. Additionally, learners' attitudes towards such tools should be considered.'

In the analysis of the survey data, it was also revealed that institutional factors might cause educators not to include the bot in the assessment process. One teacher said: 'I have not tried it before. I am not sure if I can trust it. Also, it might be problematic in terms of my institution.' The analysis also indicated that some teachers thought conducting a formal assessment using an AI tool would not be welcome on institutional terms. One teacher said: 'I am not sure about its use, but I do not think using an AI technology to score students' writing papers would be appropriate. That is the teachers' responsibility to score students' papers.' Besides the evaluation process, its use negatively influences learners' writing process since some teachers thought it would lead to plagiarism. A teacher said: 'Plagiarism is another big issue regarding AI tools.'

Suggestions

The data analysis also revealed teachers' thoughts regarding the ultimate use of ChatGPT during the evaluation process and for improving writing. The teachers underlined the necessity of balanced collaboration between teachers and the bot to evaluate the students' writing. One of the teachers voiced: 'A balanced collaboration between these tools and teachers is the key to the best grading policy for writing papers.' As can be understood from this quote, the best way to make the most of the bot is to collaborate with it by involving it in the evaluation process with the teacher. The teachers did not seem to approve that the bot would score students' papers. However, it could be used for a second check besides the teacher evaluation, as expressed in the quote: 'It can be used as a secondary source to double check my feedback.' whereas another teacher's suggestion supported the same idea, but with the concern related to its effectiveness as seen in the quote: 'It might be useful but not as effective as an instructor's feedback.' Thus, the teachers suggested it could be used to enhance the feedback the students received from the teacher, so it would be effective when used together with the teacher. The responses indicated that the teachers assumed the bot

effectively checked accuracy problems on a paper. Thus, teachers proposed that it be used effectively to check grammar and vocabulary problems in students' writing. Besides, using it for idea generation rather than evaluation was suggested. Teachers could guide their students to employ the bot to produce ideas on a given topic as a pre-writing activity.

Based on their assumptions, the instructors stated that their experience was limited with respect to using ChatGPT, especially for evaluation. Therefore, training would be needed to learn how to make the most of the bot in the educational system. The teachers recommended that training for ChatGPT use be given to the teachers. One teacher expressed: 'I am very positive about it. They are much better than the traditional technology tools we have been using. However, I would like to mention that proper teacher training and research-based benefits are to be considered before its implementation.' Another one also stated that some training must be provided to teachers and students on using the tool. Thus, teachers' comments indicated they held positive attitudes towards integrating ChatGPT in the evaluation process. However, this was only possible when they were provided with training on how to use it.

Discussion

This study investigated whether or not ChatGPT could be employed to score the students' writings. In alignment with this purpose, EFL students' papers were scored by a teacher and ChatGPT. Then, the scores were compared to reveal the differences, if any. To further understand the current practices and thoughts regarding the use of ChatGPT in educational settings, EFL instructors working at diverse universities were surveyed. In this section, the results were discussed in detail, considering the studies in the literature.

Studies indicate that ChatGPT produces accurate, reliable, and consistent scores while scoring responses (Demir, 2023; Mizumoto & Eguchi, 2023). The findings from those studies propose that it can be beneficial for scoring. In the same vein, the present study showed no statistically significant difference between the scores given by the teacher and ChatGPT, mostly, suggesting that there was not a difference between the two raters' scores at all, and ChatGPT could replace the human rater for scoring students' papers. Such a result may hint that ChatGPT can potentially score students' papers like human raters on the condition that the assessment criteria are logged on the system correctly.

On the other hand, the results also indicated differences between the teacher and ChatGPT in two of the tasks. In one of them, the teachers' scores were higher than the ChatGPT's, while in the other, the ChatGPT's scores were analyzed to be higher. While interpreting this result, meticulous care needs to be given. Ranalli (2018) explains that automated written corrective feedback tools cannot distinguish user differences depending on their language proficiencies, writing abilities, or academic experiences. Wilson, Huang, Palermo, Beard, and MacArthur (2021) explained that teachers and AWET differ in how they provide feedback. Accordingly, AWET's way of giving feedback involves a consistent style without any attention to students' struggles and individualities and without even paying attention to students' level-based stance. Considering this, the differences between ChatGPT and human rater's scores may be an outcome of a human error or the effects of factors such as the teacher's personal experience with the students, or the teacher might have given more emphasis or priority to the content as an effect of the time and effort she/ he spent in the classroom with the students. Another possibility is that because the teacher was one of the

witnesses of the student's progress in writing, her/ his experiences with the students might have reflected upon the teacher's scores. There is no doubt that although giving objective scores was the ultimate aim of assessment and evaluation processes, teachers might bring their personal experiences to the assessment processes. In the present study, the teacher assessed her/his own students; therefore, it might have affected the scores given to the tasks with the statistically significant difference. However, the interpretations cannot go beyond speculations without further research since machines and AI technologies are not error-free. It is also possible that no matter which instructions were logged on the system, ChatGPT could have yielded inconsistent results, and this might have resulted in statistically significant differences in scores given by the teacher and ChatGPT.

When the teachers' thoughts regarding the use of ChatGPT as an AWET were analyzed, the findings showed that the teachers were positive about it since it can potentially reduce the workload and help teachers save time and manage time by assisting them to fix problems such as giving corrective feedback, fixing grammatical, mechanical, and lexical problems in students' written work (Ayan & Erdemir, 2023). In the same vein as the previous research (Wang et al., 2012), the present study revealed that the teachers perceived its use for assessing students' writing production objectively and consistently (Li et al., 2015; Koltovskaia, 2020). Also, the teachers expressed that they found the bot effective for accuracy checks, enabling them to focus more on the content, as Li et al. (2015) mentioned.

Considering the positive attitudes towards ChatGPT integration in the educational systems, one may conclude that teachers perceive the potential of ChatGPT in writing evaluations; however, the findings indicated that the teachers had worries because they lacked knowledge and experience. The emergence of new technologies has also led the education field to experience fast renovations, which demands the people in the field to develop skills such as critical thinking, problem-solving, digital literacy, creative thinking, and cooperative working (Dilekçi & Karatay, 2023; Halaweh, 2023). As explained in Halaweh (2023), since ChatGPT is an evolving tool, students and educators may not have gained sufficient experience. Thus, training is required to employ it suitably and educate students and educators to use its functions effectively. Wilson et al. (2021) discussed that their study did not lack the technology assets, support for integrating AWET in the instruction, and sufficient support for teachers' professional development. However, it still did not indicate an effective use of the AWET. Thus, Wilson et al. (2021) pointed out that the teachers' inadequate content knowledge related to technology use and content might have yielded this result. The conclusion from the present study may also attract attention to the necessity of integrating AI technologies in teacher education programs, and establishing policies in educational institutions to incorporate AI in their teaching and learning systems is inevitable.

Other concerns related to using ChatGPT involved issues such as data privacy and negative impacts on the learner-teacher relationship in the classroom. The integration of AI technologies attracts criticism precisely because of ethical issues related to transparency and fairness (Latif & Zhai, 2024), as can also be understood from the teachers' responses in the present study. The ethics-related problems with the ChatGPT use have been discussed in the literature. The bot has been addressed as possessing inabilities such as transmitting false information and causing ethics-related problems (Zhang & Mao, 2023).

Moussalli and Cardoso (2020) explain the effectiveness of students' interaction with the bot over human-to-human interaction because of the never-ending function of the bot. The present study also indicated teachers' concerns about the undesirable effect of AI technologies

on student and teacher interaction. The finding here was in line with the literature (Zhang & Mao, 2023). There is a negative effect of ChatGPT on student-teacher relations because when students get accustomed to communicating with ChatGPT, they will be less motivated to interact with the teachers and other students. Such concerns related to ChatGPT use may lead to hesitation in benefitting from such technologies.

AI technologies have become inevitable in all aspects of life, so avoiding it in education will be impossible. It seems that ChatGPT will not replace teachers fully yet. Still, it can assist and service them in various ways, such as personalizing teaching resources and decreasing workload for the retrieval and generation of information (Liu et al., 2023). Halaweh (2023) underscores that collaboration between humans and an AI tool would yield a more effective output for research, idea generation, text editing, and writing. As a result, AI can meet human competencies in learning contexts. In addition to its use by teachers, it should be teachers' responsibility to help students reach the knowledge and improve critical thinking skills and vision to develop their capabilities of using AI technology at the age of AI (Liu et al., 2023). AI technologies like ChatGPT are not used instead of human skills yet. They are logical supporters. Therefore, people in the field of education, students and educators, can be helped by these technologies through effective methods (Mizumoto & Eguchi, 2023).

Conclusion

The use and integration of AI technology in school curricula seem to be unavoidable in the future. Educators have already started adopting AI tools for their individual practices and have benefitted from their merits for refining their instruction with technological commodities. This study investigated the use of ChatGPT as a tool to evaluate learners' writing and reveal its effectiveness in this respect. It also indicates how teachers perceive its use for that purpose. The study indicated that whether a human teacher or ChatGPT evaluated writings, the overall scores did not reveal any differences. This might suggest that educators might consider integrating them into the evaluation process. However, the study also draws attention to the fact that they are not entirely reliable since the scores through the tasks indicated differences. Also, the study provided insights regarding the teachers' perceptions of ChatGPT. Accordingly, the teachers were of the opinion that the ChatGPT use has the potential to save time and effort and score the papers consistently and reliably. Also, it could be used effectively to generate ideas for the paragraphs. Besides, the teachers had concerns about data security, the negative impact on teacher-student interaction, and the lack of trust in reliable scoring. The teachers suggested that an ideal use of ChatGPT integration would be a collaboration between the teacher and the AI technology since it is impossible to ignore the value of technology in school curricula. The teachers also emphasized the importance of adequate training to make the most of the technology.

Limitations and Future Research

There are several limitations that the study needs to improve. Firstly, while interpreting the results, the study's small sample size should be considered carefully to reach correct conclusions representing the whole population. Thus, future studies might focus on analyzing more writing and comparing a human teacher and an AI tool concerning the assessment. In addition, the participants in the study were all EFL learners whose English proficiency levels

were at least intermediate. Therefore, the findings from the present study may fail to show the evaluation of lower-level written productions. Thus, further studies might be conducted to explore if there are any differences when writings at different levels were evaluated over time by comparing the scores of human and AI raters. Also, the present study focused on the teachers' thoughts regarding its use. The students' experiences and thoughts could be researched in future studies, and the results from the present one might be compared to their results. Another limitation is based on the preference of the ChatGPT model used in the present study. In the present study, the free standard model of ChatGPT was used to score the papers. However, with the advanced models, such technologies will benefit more when different versions of AI technologies are used, and AI technologies like ChatGPT make quick and reliable assessment possible in education by showing its effectiveness for evaluating the content and accuracy of learners' responses in their study (Latif & Zhai, 2024). All in all, human scoring is individual; it requires time and effort, so automated scoring systems could assist in scoring as they are cost and time-effective and produce consistent and precise results (Hussein, Hassan, & Nassef, 2019). When the results here are considered, it might be assumed that the ChatGPT could benefit in assessment processes; on the other hand, this particular technology needs time to be explored as its primary purpose is not educational, so it should be used cautiously.

Statement of Conflict of Interest

In line with the Committee on Publication Ethics statement, I hereby declare that I had no conflicting interests regarding any parties to this study.

Appendix A. Open-Survey Teacher Questions

1. **Demographic information**

- How long have you been teaching English?
- Do you have L2 teaching writing experience?
- Have you ever received Automated Writing Corrective Feedback (AWCF) / Evaluation (AWCE) Tools training?

2. **Prior experience with AWCF/E tools**

- What do you know about AWE tools and similar tools?
- Have you ever used AWCF tools to evaluate writing papers? If yes, which ones? How long?
- How do you feel about their use? What is your general attitude toward using them in L2 writing classrooms?
- What are the benefits/ disadvantages of using them?

3. **Experiences and Thoughts about using the ChatGPT as an automated writing evaluation tool**

- How do you feel about using ChatGPT to supplement your feedback on your students' writing assignments and to score them?
- Have you ever used the ChatGPT to give feedback on the students' writings? Why? Why not?
- Have you ever used the ChatGPT to score your students' papers? Why? Why not?
- Do you find it reliable for scoring and giving feedback? Why? Why not?
- What are the main benefits/ drawbacks of using the ChatGPT to provide feedback and score the papers?

References

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, *2*, 1-18. https://doi.org/10.1016/j.mlwa.2020.100006

Alberth. (2023). The use of ChatGPT in academic writing: A blessing or a curse in English. *TEFLIN Journal, 34*(2), 337-352. http://dx.doi.org/10.15639/teflinjournal.v34i2/337-352

Algaraady, J., & Mahyoob, M. (2023). ChatGPTs capabilities in spotting and analyzing writing errors experienced by EFL learners. *Arab World English Journal*, *9*, 3-17. https://dx.doi.org/10.24093/awej/call9.1

Avsheniuk, N., Lutsenko, O., Svyrydiuk, T., & Seminikhyna, N. (2024). Empowering language learners' critical thinking: Evaluating ChatGPT 's role in English course implementation. *Arab World English Journal (AWEJ),* 210-224. https://dx.doi.org/10.24093/awej/ChatGPT.14

Ayan, A. D., & Erdemir, N. (2023). EFL teachers' perceptions of automated written corrective feedback and Grammarly. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi (AKEF),* *5*(3), 1183-198. https://doi.org/10.38151/akef.2023.106

Bin-Hady, W. R. A., Al-Kadi, A., Hazaea, A., Ali, J. K. M. (2023). Exploring the dimensions of ChatGPT in English language learning: A global perspective. *Library Hi Tech*. https://dx.doi.org/10.1108/LHT-05-2023-0200

Chiu, M., Hwang, G., Hsia, L., & Shyu, F. (2022). Artificial intelligence-supported art education: A deep learning-based system for promoting university students' artwork appreciation and painting outcomes. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2022.2100426

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. https://doi.org/10.46778/goputeb.1345752

Dilekçi, A., & Karatay, H. (2023). The effects of the 21st century skills curriculum on the development of students' creative thinking skills. *Thinking Skills and Creativity*, *47*, 101229. https://doi.org/10.1016/j.tsc.2022.101229

Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning, and Assessment*, *5*(1),1–35.

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 1-20. https://doi.org/10.1186/s41239-023-00425-2

Evmenova, A. S., Borup, J., & Shin, J. K. (2024). Harnessing the power of generative AI to support ALL learners. *TechTrends*, *68*, 820-831. https://doi.org/10.1007/s11528-024-00966-x

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. https://doi.org/ 10.1007/s10639-023-12146-0

Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, *15*(2), 1-11. https://doi.org/10.30935/cedtech/13036

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.

Ho, P. X. P. (2024). Using ChatGPT in English language learning: A study on I.T. students' attitudes, habits, and perceptions. *International Journal of TESOL & Education*, *4*(1), 55-68. https://doi.org/10.54855/ijte.24414

Hussein, M. A., Hassan, H., & Nassef, M. (2019). Automated language essay scoring systems: A literature review. *Peer Computer Science*, *5*, 1-16. https://doi.org/10.7717/peerj-cs.208

Kucuk, T. (2024). ChatGPT integrated grammar teaching and learning in EFL classes: A study on Tishk international university students in Erbil, Iraq. *Arab World English Journal*, 100-111. https://dx.doi.org/10.24093/awej/ChatGPT.6

Latif, E., & Zhai, X. (2024). Fine-tuning ChatGPT for automatic scoring. *Computers and Education*, *6*, 1-10. https://doi.org/10.1016/j.caeai.2024.100210

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

Liu, X., Zhao, A., Sun, X., Zhang, K., Kou, F., & Gai, J. (2023). The rise of ChatGPT: Unlocking its potential in education. In D. Kumar, P. Loskot, & Q. Chen (Eds.), *The 3rd International Conference on Internet, Education and Information Technology (IETS 2023)* (pp.1230-1236). Springer. https://doi.org/10.2991/978-94-6463-230-9_148

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. https://doi.org/10.1016/j.rmal.2023.100050

Mohamed, A. M. (2024). Exploring the potential of an AI-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

Moussalli, S., & Cardoso, W. (2020). Intelligent personal assistants: Can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning*, *33*(8), 865-890. https://doi.org/10.1080/09588221.2019.1595664

Qu, K., & Wu, X. (2024). ChatGPT as a CALL tool in language education: A study of hedonic motivation adoption models in English learning Environments. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12598-y

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

Solak, E. (2024). Revolutionizing language learning: How ChatGPT and AI are changing the way we learn languages. *International Journal of Technology in Education (IJTE)*, *7*(2), 353-372. https://doi.org/10.46328/ijte.732

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, *19*, 51–65. https://doi.org/10.1016/j.asw.2013.11.007

Susanto, D. A., Priyolistiyanto, A., Pinandhita, F., Prabowo, A. B., & Bimo, D. S. (2024). Utilizing ChatGPT on designing English language teaching (ELT) materials in Indonesia: Opportunities and challenges. *Journal of Culture, English Language, Teaching & Literature, 24*(1), 157-181. https://doi.org/10.24167/celt.v24i1

Tai, T., & Chen, H. H. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, *31*(3), 1485-1502. https://doi.org/10.1080/10494820.2020.1841801

Tsai, C., Lİ, Y., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12722-y

Wang, Y., Shang, H., & Briody, P. (2012). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning*, *26*(3), 1–24. https://doi.org/10.1080/09588221.2012.655300

Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2022). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2022.2056203

Wang, E. L., Matsumura, L. C., Litman, D., Correnti, R., Zhang, H., Rahimi, Z., Kisa, Z., Magooda, A., Howe, E., Quintana, R. (2022). *Contributions to research on automated writing scoring and feedback systems*. RAND Corporation. https://doi.org/10.7249/RBA1062-1

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

Wilson, J., Huang, Y., Palermo, C., Beard, G., & MacArthur, C. A. (2021). Automated feedback and automated scoring in the elementary grades: Usage, Attitudes, and associations with writing outcomes in a districtwide implementation of MI write. *International Journal of Artificial Intelligence in Education*, *31*, 234-276. https://doi.org/10.1007/s40593-020-00236-w

Zhang, B., & Mao, J. (2023). On the teaching and learning in the information age for "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. https://doi.org/10.2991/978-94-6463-192-0_131

Zheng, L., Niu, J., Zhong, L., & Gyasi, J. F. (2021): The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2021.2015693