CHATGPT'S AND HUMANS' WRITTEN CORRECTIVE FEEDBACK IN EFL/ESL WRITING: A SYSTEMATIC REVIEW

Dr. Yustinus Calvin Gai MALI

ORCID: 0000-0001-9075-783X
Master's Program in English Language Education
Universitas Kristen Satya Wacana
Salatiga, INDONESIA

Received: 09/12/2024 **Accepted:** 21/04/2025

ABSTRACT

Written corrective feedback (WCF) is critical in EFL/ESL writing classrooms. This study aims to review existing research studies on the quality of WCF generated by ChatGPT and humans and to examine how WCF generated by ChatGPT can support WCF that humans give to students' written work in higher education EFL/ESL writing contexts. To meet these goals, the researcher examined 23 studies published in 18 peer-reviewed journals from 2023-2024. The analysis results highlighted potential strengths and weaknesses of WCF generated by ChatGPT and humans. WCF generated by ChatGPT could review various aspects of writing and provide extensive descriptive feedback, yet it sometimes might experience system fatigue that affects the WCF quality. ChatGPT could also provide inaccurate WCF, give WCF that lacked human nuances, and were greatly influenced by how prompts were presented to it. Meanwhile, humans could provide aspects of WCF not accounted for ChatGPT, but their WCF might involve subjective judgment and be unclear. With these findings in mind, the researcher identified three phases of how ChatGPT can work harmoniously with humans to provide quality WCF for students in writing classrooms. Those phases were making various efforts to enhance WCF generated by ChatGPT in students' writing in support of the feedback provided by lecturers, enhancing students' in-depth understanding that not all ChatGPT feedback is accurate, and combining ChatGPT-based WCF, specifically on earlier drafts of students' writing. Pedagogical implications, which highlighted the essence of working synergistically with ChatGPT and mastering prompts to generate the most helpful WCF, for EFL/ESL lecturers in writing classrooms who wish to support their feedback with ChatGPT to provide quality WCF to their students' written work were presented. The researcher then proposed recommendations for future systematic review studies, expanding the research contexts, the scope of reviewed materials, and the research method of the reviewed articles, to address gaps and expand on the insights presented in this study.

Keywords: ChatGPT, EFL, ESL, written corrective feedback, writing classrooms.

INTRODUCTION

In higher education contexts, particularly in English language education programs, writing skills are crucial in students' language development and achievement across professional domains (Abas & Abd Aziz, 2018; Hang, 2023; Mali, 2022; Song & Song, 2023). Proficiency in writing can also ensure the success of their studies, as many programs require the students to pass some writing courses, write a research proposal, complete a thesis, and eventually publish the thesis in an academic journal (Harmawan et al., 2023; Mali, 2023; Purwanti & Kastuhandani, 2023; Ramadhanti et al., 2023). Yet, at the same time, writing might be one of the most challenging skills to master when learning English as a Second Language (ESL) or English as a Foreign Language (EFL) given the long, non-linear, and recursive writing processes of planning, drafting, revising, and editing (see Laksmi, 2006; Mali & Salsbury, 2021), and a range of linguistics abilities of grammatical rules, syntactic expressions, and lexical knowledge as well as a range of planning strategies for ideas organization, rhetoric, and styles that students have to master (Aydin & Yildiz, 2014; Chandra & Yuyun, 2018).

To address the challenges of mastering writing skills, regular, constructive, and timely written corrective feedback (WCF) from lecturers to their students' work in EFL/ESL writing classes is critical (Banihashem et al., 2024; Hang, 2023). However, providing such feedback to the students poses a significant challenge for lecturers who often experience fatigue from repetitive tasks of correcting students' written work without seeing the effectiveness of those practices (Lee, 2016; Shadiev & Feng, 2024), especially those who teach a writing class with a high number of students (Baskara, 2023; Bibi & Atta, 2024; Lin & Crosthwaite, 2024; Shadiev & Feng, 2024) and are overloaded with daily administrative tasks besides teaching. These circumstances highlight the need for alternative methods to support lecturers in delivering constructive and quality WCF for their students' written work. In this study, WCF refers to written information, markings, or codes for students' writing to correct errors at local (i.e., sentence-level grammar) or global (i.e., related to content and ideas organization) levels to improve students' writing skills, promoting language acquisition, and enhancing their writing quality (Lin & Crosthwaite, 2024; Luo et al., 2024; Mao et al., 2024; Rahimi et al., 2024).

One of the promising alternatives is to use and maximize the potential of artificial intelligence (AI) technology, namely Chat Generative Pre-Trained Transformer (hereafter called ChatGPT) (https://chat.openai.com/). Briefly, ChatGPT is a chatbot based on a large language model released in November 2022 by OpenAI (a company based in San Fransico, California) (Crcek & Patekar, 2023; Rasul et al., 2018; Rudolph et al., 2023b), capable of generating human-like responses and new content based on a large-scale language database, emphasizing personalized and interactive assistance (Ngo, 2023; Xiao & Zhi, 2023). When prompted by humans, it can address various queries across various knowledge domains (Alkamel & Alwagieh, 2024). More recently, ChatGPT has garnered considerable attention for its ability to grade students' written work (Awidi, 2024; Barrot, 2023; Floden, 2024; Jukiewicz, 2024; Li et al., 2024) and to provide immediate WCF on essay structure, grammatical aspects, clarity and coherence of students' writing, punctuation, and vocabulary choice (Algaraady & Mahyoob, 2023; Alqasham, 2023; Baskara, 2023; Bonner et al., 2023; Cheong & Hong, 2023; Nguyen et al., 2024; Ozcelik & Eksi, 2024; Zadorozhnyy & Lai, 2023). Besides, ChatGPT can provide targeted, adaptive, and helpful feedback; all of which help students enhance their writing (Steiss et al., 2024).

Given ChatGPT's potential, its growing global adoption (Alberth, 2023; Grassini, 2023; Yeo, 2024), and claims that "ChatGPT might be the Google of tomorrow" (Alqasham, 2023, p. 1258), there is a possibility that more EFL/ESL lecturers will increasingly use ChatGPT to help them provide WCF for their students' writing. However, research comparing the quality of WCF generated by ChatGPT and humans remains limited (Solovey, 2024; Teng, 2024b). This limited research exploration highlights the urgency of expanding the results of previous studies, which were limited to a single setting, such as in Macau (see Teng, 2024b) with a limited number of participants, to inform best practices and pedagogical approaches to using ChatGPT to provide WCF, especially in EFL/ESL classroom settings. With all the backgrounds mentioned above, this study aims to address that gap by synthesizing findings from multiple countries to examine ChatGPT-generated WCF compared to human-generated WCF (i.e., lecturers or human evaluators) in EFL/ESL writing classrooms within higher education. The following research question guided the exploration of the current study:

1. How does the quality of WCF provided by ChatGPT compare to that of humans in EFL/ESL writing classrooms?
   a. What are the strengths and weaknesses of WCF generated by ChatGPT?
   b. What are the strengths and weaknesses of WCF provided by humans?
2. How can ChatGPT-generated WCF support human-provided WCF in EFL/ESL writing classrooms?

Answers to the research questions above hope to provide practical insights for language lecturers on how ChatGPT can complement their feedback practices by addressing areas where they may face challenges and vice versa. This study also seeks to demonstrate how ChatGPT and human-generated WCF can work synergistically to enhance students' writing quality in EFL/ESL classrooms.

METHOD

This study is a systematic review aiming to combine and enhance knowledge in specific subjects by collecting and analyzing all pertinent literature on a topic or research question(s) (Callanga et al., 2024; Zain, 2022), mainly related to the WCF-generated by ChatGPT and humans in EFL/ESL writing classrooms. To achieve this goal, the researcher adopted the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) guidelines, which could ensure the selection of high-quality and reliable data for systematic review studies (Imran & Almusharraf, 2023; Teng, 2024a). Previous studies (e.g., Imran & Almusharraf, 2023; Putri et al., 2023; Teng, 2024a; Utami & Santosa, 2023) have demonstrated the effectiveness of using PRISMA in their systematic review studies.

The first stage was *identification.* It involved searching relevant journal articles from *Google Scholar,* the same database previous researchers (e.g., Grassini, 2023; Li, 2012, 2018; Rudolph et al., 2023b) used in their systematic review studies. The search was done in October 2024 with the support of *Publish or Perish* software (8th edition) (https://harzing.com/resources/publish-or-perish), the same software that Putrie et al. (2024) in their systematic review study. Because of its user-friendly design and ability to conduct extensive literature searches, *Publish or Perish* is a helpful tool for academic research (Callanga et al., 2024). In the software, the researcher played with the following keywords: *comparing ChatGPT and human to score essays, comparing ChatGPT and human to students' writing, ChatGPT in EFL/ESL writing class, ChatGPT and human feedback in EFL/ESL writing classrooms,* and *comparing ChatGPT and teacher written corrective feedback.* These keywords were developed from titles of published journal articles relevant to the topic and research questions of the study. The initial search resulted in 200 publications.

The second step was *screening.* The articles retrieved from the search were then further screened based on the following inclusion and exclusion criteria to ensure the quality and novelty of the review. First, the articles should provide ideas related to the research questions (1-2) of this study: (1) the quality of WCF provided by ChatGPT compared to that given by humans in EFL/ESL writing classrooms and (2) how WCF provided by ChatGPT can strengthen WCF humans give to students' written work in EFL/ESL writing classrooms. In this study, the term EFL indicates contexts where people learn English mainly in a formal classroom setting, with limited opportunities to use the language outside their classroom for daily communication or other non-academic activities (Mali, 2017, 2018). Meanwhile, the term ESL shows contexts where people learn English in a place where English plays essential roles in daily communication, education, business, and government (Richards & Schmidt, 2010). Second, the articles are empirical studies that collect evidence based on original research data. Third, the articles were recently published in 2023-2024 to ensure that the discussion on the topic is based on current and up-to-date studies. Fourth, the articles are written in English and are peer-reviewed academic articles. Fifth, the articles are published in peer-reviewed SCOPUS-indexed journals. Sixth, if not indexed by SCOPUS, the articles should be cited at least ten times by other studies. Seventh, the selected articles should be open-access. Lastly, the setting of the study should be related to EFL/ESL writing classrooms or EFL/ESL writing-related activities in a higher education setting. An article that does not meet the inclusion criteria above was excluded from the review.

Table 1. The Inclusion and Exclusion Criteria

Criteria	Inclusion	Exclusion
Relevance	Provide ideas related to the research questions of this study	Not related to the research questions of this study
Article types	Peer-reviewed articles, empirical studies	Non-peer-reviewed articles, non-empirical studies
Year of publication	2023-2024	Before 2023
Language	English	Non-English
Quality	SCOPUS-indexed or has been cited at least ten times	Non-SCOPUS indexed or has been cited less than ten times
Access	Open-access	Non-open access
Setting	Related to EFL/ESL writing classrooms or EFL/ESL writing-related activities in a higher education setting	Not related to EFL/ESL writing classrooms or EFL/ESL writing-related activities; non-higher education setting

The next step was *eligibility.* With the criteria displayed in Table 1, the researcher carefully examined the previous studies' abstracts, methods, findings, and conclusion sections and initially found 45 articles eligible for review in this study. The last step was *inclusion.* The researcher finally selected 17 articles that were most relevant to the research goals and could provide valuable ideas to answer the research questions. Besides, the researcher included five other articles (e.g., Awidi, 2024; Barrett & Pack, 2023; Mizumoto & Eguchi, 2023; Ozcelik & Eksi, 2024; Steiss et al., 2024) from non-EFL/ESL writing contexts, considering the significance of their findings and discussions to help answer the research questions of the present study. The researcher also decided to review a non-empirical study by Tseng and Warschauer (2023) as it has many insights to answer the research questions of this study. Table 2 depicts the distribution of the empirical studies (n=23) reviewed in this article.

Table 2. Distribution of the Empirical Studies

Journal Title	Number of articles	Articles
Annals of Human and Social Sciences	1	Bibi and Atta (2024)
Arab World English Journal	1	Algaraady and Mahyoob (2023)
British Journal of Educational Technology	1	Yavuz et al. (2024)
Computers and Education: Artificial Intelligence	3	Awidi (2024); Meniado et al. (2024); Teng (2024b)
Discover Artificial Intelligence	1	Slamet (2024)
Education and Information Technologies	1	Bui and Barrot (2024)
Electronic Journal of e-Learning	1	Tseng and Lin (2024)
Humanities & Social Sciences Communications	1	Li et al. (2024)
International Journal of Educational Technology in Higher Education	2	Banihashem et al. (2024); Barrett and Pack (2023)
International Journal of Interactive Mobile Technologies	1	Ajlouni et al. (2023)
International Journal of Learning, Teaching and Educational Research	1	Jackaria et al. (2024)
Journal of China Computer-Assisted Language Learning	1	Tseng and Warschauer (2023)
Learning and Instruction	1	Steiss et al. (2024)
Research Methods in Applied Linguistics	2	Mizumoto et al. (2024); Mizumoto and Eguchi (2023)
Smart Learning Environments	1	Ozcelik and Eksi (2024)
Social Sciences & Humanities Open	1	Alkamel and Alwagieh (2024)
System	2	Han and Li (2024); Lin and Crosthwaite (2024)
Technology in Language Teaching & Learning	1	Solovey (2024)

As displayed in Table 2, most articles were published in educational-related journals, while a few were published in educational-technology journals.

FINDINGS

In this section, the researcher presents the results of his review analysis to answer the research questions in this study. This section starts with a matrix of the previous studies presented chronologically (i.e., 2023-2024) and alphabetically based on the authors (i.e., A to Z) to provide a brief overview of the studies reviewed in this paper.

Table 3. Matrix of the Previous Studies

No	Authors	Study Goals	Contexts and Participants	Designs
1	Ajlouni et al. (2023)	To examine students' attitudes toward using ChatGPT as a learning tool	623 undergraduate students at the University of Jordan	A survey study using a questionnaire
2	Algaraady and Mahyoob (2023)	To explore the potential of ChatGPT in detecting writing errors made by EFL students	88 EFL university written tasks in a university in Yemen	A mixed-method study using error analysis performed by ChatGPT and EFL instructors
3	Barrett and Pack (2023)	To compare educators' and university students' perceptions of the appropriate use of ChatGPT	158 university students and 68 lecturers at a public research university in the United States	A cross-sectional questionnaire design
4	Mizumoto and Eguchi (2023)	To perform automated essay scoring by ChatGPT and evaluate its effectiveness	12.100 essays contained in the ETSCorpus of Non-Native Written English	Document and statistical analysis
5	Tseng and Warschauer (2023)	To propose a five-part pedagogical framework to maximize the potential of AI tools	Non-empirical study	
6	Alkamel and Alwagieh (2024)	To explore EFL students' opinion, benefits, and challenges regarding the use of ChatGPT in academic writing	144 students from the Department of English Language at universities in Yemen	A quantitative study using a survey
7	Awidi (2024)	To compare ChatGPT's marking systems with Expert Tutors in assessing written reflective assessments	A sample of 108 written reflection assessment data at a university in Australia; 39 trained tutors	A comparative study using statistical analysis techniques
8	Banihashem et al. (2024)	To compare the quality of ChatGPT-generated feedback with peer feedback	74 graduate students from a Dutch university who completed an online module, namely Argumentative Essay Writing	A comparative study using statistical analysis techniques
9	Bibi and Atta (2024)	To determine how effective ChatGPT is in assisting students in writing English	150 undergraduate students at a university in Pakistan	A mixed-method study using surveys, interviews, and testing
10	Bui and Barrot (2024)	To examine the relationship between the scores assigned by ChatGPT and a human rater, and how consistent ChatGPT-assigned scores are when taken at multiple time points	200 argumentative essays written by college students from four categories of language proficiency in 10 Asian countries (i.e., China, Hong Kong, Indonesia, Japan, Pakistan, the Philippines, Singapore, South Korea, Taiwan, and Thailand)	A cross-sectional quantitative approach
11	Han and Li (2024)	To investigate ChatGPT-supported teacher feedback in an EFL context and explore an innovative AI-aided writing pedagogy	Four instructors and 102 students from two undergraduate classes in the world language education program in China	A quantitative study
12	Jackaria et al. (2024)	To compare ChatGPT version 3.5 and human raters in scoring students' essays	20 essays written by college students at a public university in the Philippines	Comparative-descriptive research
13	Li et al. (2024)	To assess the reliability of the holistic scores assigned to EFL essays by ChatGPT (version 3.5 and 4) compared to college English teachers, and to evaluate the relevance of qualitative feedback provided by both versions of ChatGPT	30 essays written by 30 non-English major undergraduate students and four college English teachers at a university in Beijing, China	A comparative study using statistical analysis techniques

14	Lin and Crosthwaite (2024)	To compare the WCF produced by teachers and ChatGPT, examine their respective capabilities while identifying differences in their feedback practice	25 teachers from China, Australia, and UK to mark three written essays selected from the International Corpus Network of Asian Learners of English	A comparative study using statistical techniques
15	Meniado et al. (2024)	To explore the experiences and perceptions of Thai and Vietnamese EFL learners on the use of ChatGPT for second-language writing	357 survey respondents from three educational institutions in Thailand and Vietnam, and qualitative data from 16 interviewees from the same group of respondents	A mixed-method study using surveys and interviews
16	Mizumoto et al. (2024)	To evaluate the accuracy of ChatGPT against human coding and Grammarly	The Cambridge Learner Corpus First Certificate in English (CLC FCE) dataset	A comparative study
17	Ozcelik and Eksi (2024)	To examine the impact of ChatGPT on the acquisition of register knowledge across various writing tasks	11 students from different departments at a state university in Turkiye	A one-shot case study
18	Slamet (2024)	To gain insight into how the research participants perceive using ChatGPT to improve their language learning	80 EFL teachers and 40 EFL students from different schools and universities in East Java, Indonesia	A survey approach
19	Solovey (2024)	To explore EFL students' perceptions and preferences on the effectiveness of three feedback sources: peers, ChatGPT, and teachers	30 freshmen EFL students at a university in Israel	A qualitative study using a checklist and a feedback survey
20	Steiss et al. (2024)	To examine the ability of ChatGPT to provide formative feedback compared to that given by human evaluators	200 pieces of human-generated formative feedback and 200 pieces of AI-generated formative feedback for the same essays (i.e., written by 200 students from two school districts in Southern California)	A comparative study using statistical techniques
21	Teng (2024b)	To explore the potential impact of ChatGPT on writing and explore students' perceptions and experiences of ChatGPT in generating feedback for writing	45 EFL learners from two classes enrolled in an English Practical Writing course in Macau	A mixed-method study using a questionnaire and interviews
22	Tseng and Lin (2024)	To understand students' subjective experiences and perceptions of the use of ChatGPT in their writing process	15 students in a university-level EFL writing course at a private university in Taiwan	A qualitative methodology
23	Yavuz et al. (2024)	To investigate the validity and reliability of ChatGPT and Google's Bard, in grading student essays in higher education based on an analytical grading rubric	Three students' essays at a university in Turkiye;15 experienced instructors in higher education	A comparative study using statistical analysis techniques

The matrix presented in Table 3 above can provide a general overview of the previous studies, which used various research designs of mixed-methods (e.g., Bibi & Atta, 2024; Meniado et al. 2024; Teng, 2024b), comparative studies using statistical analysis techniques (e.g., Awidi, 2024; Banihashem et al., 2024; Jackaria et al., 2024; Li et al., 2024; Lin & Crosthwaite, 2024; Mizumoto et al., 2024; Steiss et al., 2024; Yavuz et al., 2024), surveys (e.g., Ajlouni et al., 2023; Slamet, 2024), case study (e.g., Ozcelik & Eksi, 2024), document analysis (e.g., Mizumoto & Eguchi, 2023) and qualitative (e.g., see Solovey, 2024; Tseng & Lin, 2024) methods.

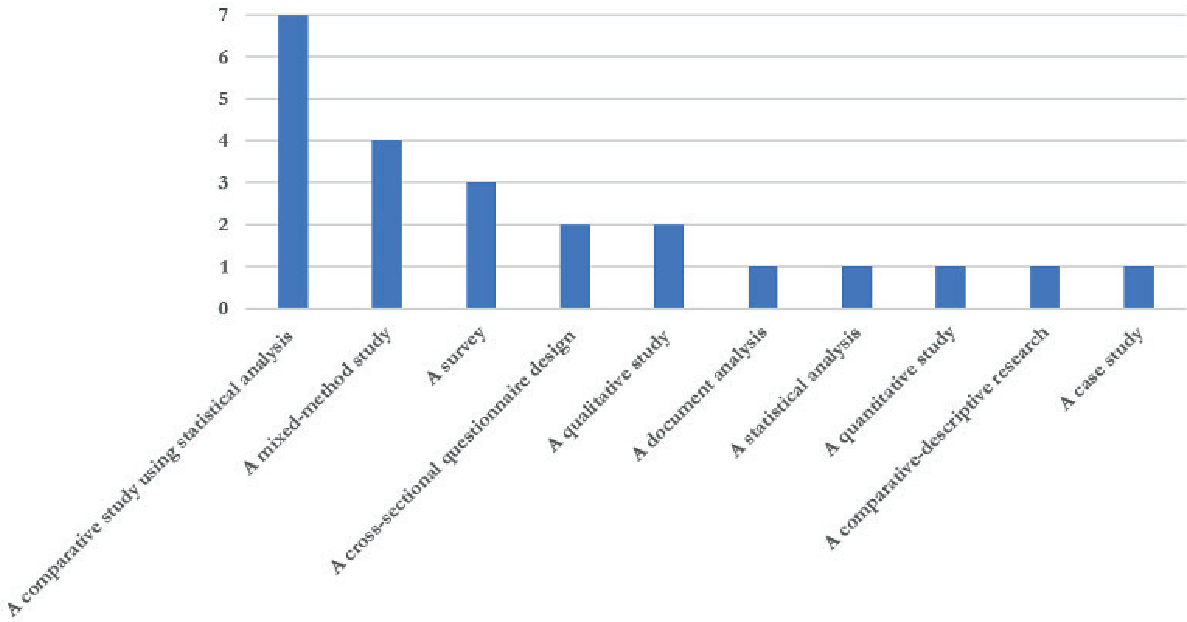

Figure 1. Research Designs of the Previous Studies

Further, the matrix also informs that the WCF by ChatGPT and humans have been discussed in various countries, such as in Asian regions (e.g., Pakistan, Vietnam, Thailand, Macau, Indonesia, Taiwan, China, Saudi Arabia, Jordan, Yemen) and some other countries, which include the United States, Israel, Turkiye, and the Netherlands.

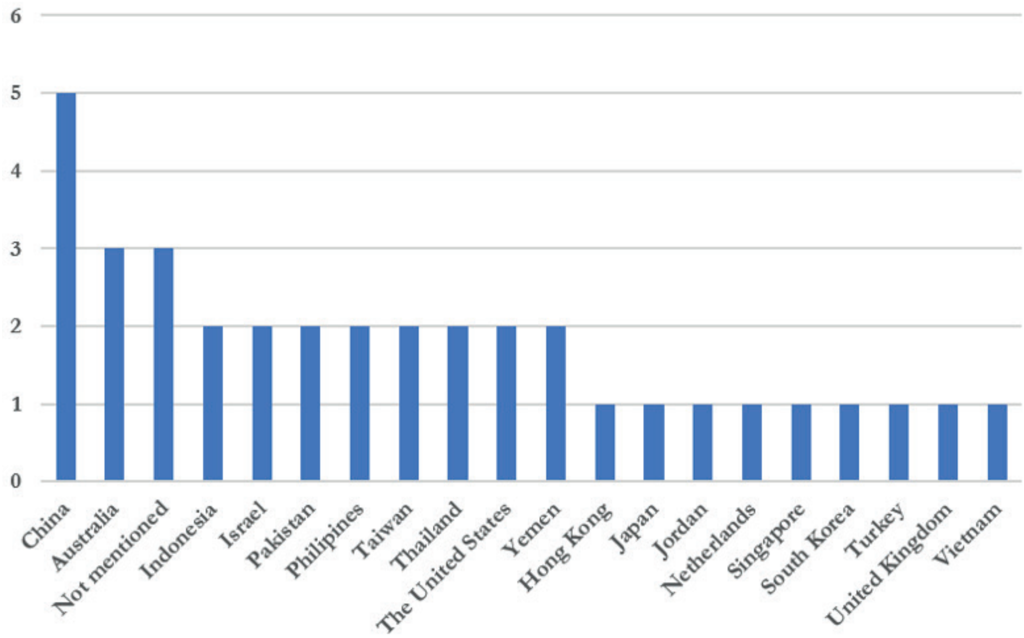

Figure 2. Distributions of the Selected Articles Based on Geographical Contexts

Dealing with Figure 2, the researcher informed that some studies (e.g., Bui & Barrot, 2024; Lin & Crosthwaite, 2024; Meniado et al., 2024) were conducted in more than one context, involving research participants from different geographical locations. In the following sections, the researcher will present the findings and discussions in order of the research questions.

RQ1: The Quality of WCF Provided by ChatGPT Compared to that of Humans in EFL/ESL Writing Classrooms

To define the quality, the researcher looked closely at the strengths and weaknesses of the WCF provided by ChatGPT and humans, as reported by the previous studies. The analysis results will be presented in the subsequent four main sections. Under each section, the researcher presents the related themes as the answer to the first research question.

The Strengths of WCF Generated by ChatGPT

The first theme is *reviewing various aspects of writing*. The previous studies reported that ChatGPT can review and provide feedback for various aspects of students' writing (Bui & Barrot, 2024; Han & Li, 2024), such as sentence structure (Ozcelik & Eksi, 2024), vocabulary, spelling, punctuation, grammar (Meniado et al., 2024; Mizumoto et al., 2024), writing styles, and ideas' organizations (Yavuz et al., 2024). This finding is in harmony with that reported by Alkamel and Alwagieh (2024) that the majority of their respondents strongly agreed (26.6%) and agreed (48.1%) that ChatGPT is a good tool to enhance students' language use and grammar in writing. ChatGPT can also provide recommendations for writers to improve their writing coherence based on contextual information recognized in the writing (Bibi & Atta, 2024). Tseng and Lin (2024) reported a student's voice about ChatGPT that it can correct errors, highlight areas they need to improve, and find ways to improve them; "it can list all the errors for you with one click" (p. 92). Meanwhile, most research participants in Teng's (2024b) study appreciated the accuracy, reliability, and ease of integrating ChatGPT's corrections into their written work; that said, the participants confirmed ChatGPT's effectiveness as a tool for proofreading and editing. In a similar voice, the majority of the respondents (93.3%) in Solovey's (2024) study could understand all the WCF made by ChatGPT (i.e., organization, mechanics, grammar, content, and quality of written expressions) in their writing. ChatGPT can detect errors more comprehensively and focus on many aspects of writing simultaneously (Bui & Barrot, 2024); thus, it might potentially develop students' writing skills (Ajlouni et al., 2023).

The second theme is *providing extensive descriptive feedback*. ChatGPT's capability to provide more extensive descriptive feedback than what humans give might be a strength, which was evidenced in Banihashem et al.'s (2024); Han and Li's (2024) studies. Banihashem et al. (2024) argued that ChatGPT's descriptive feedback might help summarize complex argumentations or provide comprehensive overviews to enhance students' understanding, especially the overall structure and coherence of their written essays. That said, Banihashem et al. (2024), see Figure 1 on p. 7 in their study, provided evidence of the descriptive WCF by ChatGPT that is constructive, shows affection (e.g., the essay nicely argues [...]), can summarize key main ideas (e.g., the essay present multiple arguments against PIF sterilzations [...]), spots weaknesses in the students' writing (e.g., however, some of these arguments lack detailed elaboration [...]), and suggest areas for improvement (e.g., to strengthen this argument, the essay could provide real-life examples [...]). These kinds of descriptive WCF by ChatGPT are also evidenced in Han and Li's (2024) study (see Table 3 on p. 5). For example ChatGPT could summarize the key main ideas of the students' writing (e.g., the article effectively presents the importance of learning a second foreign language, particularly for students majoring in Chinese [...]), spot weaknesses in the students' writing (e.g., while the author mentions experimental results that show the cognitive benefits of learning a second language, they do not provide specific references or sources for these results [...]), and suggest areas for improvement (e.g., adding this information would make the article more credible and convincing [...]).

The Strength of WCF Provided by Humans

The only theme that emerged from the review analysis was *providing aspects of WCF not accounted for ChatGPT*. Previous studies reported that humans can provide some aspects of WCF in students' written work, which ChatGPT might have struggled with. Those aspects include appreciating students' creativity, emotional depth, reasoning, and subtlety in the argumentation of human-produced essays (Bui & Barrot, 2024). Steiss et al. (2024) reported that humans (i.e., well-trained, paid, and relatively time-rich evaluators) could provide higher quality WCF than ChatGPT in these critical areas: the use of supportive tone, clarity of directions for writing improvement, accuracy, and prioritization of essential writing features. Meanwhile, Solovey (2024) presented some students' voices that might indicate the strength of humans in providing WCF compared to ChatGPT: *"[…] the teacher paid attention to things that we and GPT did not see; teachers have the expertise, experience, and real-world context that contribute to the identification of writing errors and the provision of appropriate feedback"* (p. 15).

The Weaknesses of WCF Generated by ChatGPT

The first theme was *providing inaccurate feedback.* Previous researchers considered providing inaccurate WCF to be a potential weakness of ChatGPT. The inaccuracy included giving irrelevant comments, challenges to identify specific writing problems (Teng, 2024b), misinterpreting the writer's intent (Bui & Barrot, 2024), changing meanings or context of the text, adding unnecessary or unrelated items to the text, and making sentences more complex (Ozcelik & Eksi, 2024). Steiss et al. (2024) instanced an inaccurate WCF provided by ChatGPT in their study (see Figure 6 on p.12). Another example was from an undergraduate student in Ozcelik and Eksi's (2024) study, saying that, "*I did not like its neutral correction as it changes the text a lot*" (p.12). One of the possible factors for the WCF inaccuracy was that ChatGPT relied more on its predictive algorithm, which might not be advanced enough, to generate responses (Bui & Barrot, 2024; Steiss et al., 2024). Steiss et al. (2024) warned that giving inaccurate feedback might discourage students from engaging in their revision process.

The second theme was *lacking of human nuances.* ChatGPT lacks human nuances. As Teng (2024b) reported, some EFL students enrolled in an English Practical Writing course were concerned that ChatGPT's feedback features lacked human touching similar to what their teachers offered, such as sharing feelings and experiences in writing. Besides, ChatGPT struggles to capture the nuanced aspects of students' writing, such as creativity, styles, reasoning, and emotional depth, which are all essential in many contexts (Bui & Barrot, 2024).

The third theme was *being sensitive to how prompts are presented to it.* ChatGPT's responses might fluctuate depending on prompts presented to it (Mizumoto et al., 2024), which means that variations in the wording, structure, or contexts of a prompt might influence the WCF quality generated by ChatGPT (Awidi, 2024; Banihashem et al., 2024; Jackaria et al., 2024; Lin & Crosthwaite, 2024). Notably, it might take time and multiple attempts at prompt engineering to obtain the best WCF from ChatGPT (Steiss et al., 2024).

The fourth theme was *experiencing system fatigue.* Awidi (2024) found that ChatGPT could simultaneously process two to three essay scripts to generate more accurate feedback. Yet, ChatGPT might not work well when processing more than four scripts simultaneously, which could lead to providing general feedback lacking specificity to the needs of individual students.

The Weaknesses of WCF Provided by Humans

The first theme was *involving subjective judgement.* Human subjectivity might often influence WCF provided to students' writing (Yavuz et al., 2024). Awidi (2024) explored that humans (i.e., expert tutors) sometimes provide WCF inconsistent with the grade they gave to students' written work. Bui and Barrot (2024) revealed that humans tend to be more permissive about errors in students' essays, particularly those produced by non-native speakers of English; that said, human raters tended to assign higher scores than ChatGPT on students' written work, whereas ChatGPT might have penalized students' technical writing errors more heavily.

The second theme was *providing less clear feedback.* Traditional teachers' feedback alone might have issues with the clarity and helpfulness of the feedback, as voiced by two EFL university students (as reported by Han & Li, 2024, p. 9):

> *Usually, teachers would just mark the errors, for example, by underlining them or circling them with a red pen. Sometimes, I couldn't quite understand what exactly was wrong, whether it was a word choice error or a problem with my sentence structure, so it was difficult to make corrections.* (Student 2)
>
> *Previously, teachers would just grade my essay, and I had no idea what specifically needed to be revised or where exactly I went wrong. However, the feedback from both ChatGPT and the teacher can clearly tell me where the mistakes are and which parts need improvement.* (Student 1)

A language instructor also confessed that *"previously, I would just give a grade, and if time allowed, I would write a few sentences to tell the students where they could improve. I did not have time to praise the students' work or engage in a dialogue with them"* (Han & Li, 2024, p. 9).

RQ2: How WCF Provided by ChatGPT Can Support WCF that Humans Give to Students' Written Work in EFL/ESL Writing Classrooms

As the answer to the second research question, the researcher could synthesize three iterative phases (see Figure 3) in which WCF provided by ChatGPT can support what humans give to students' written work in the writing classes. The first phase is to make various efforts to enhance WCF generated by ChatGPT in students' writing in support of the feedback provided by teachers in their writing classes. In that case, previous studies (i.e., Awidi, 2024; Jackaria et al., 2024; Lin & Crosthwaite, 2024; Mizumoto et al., 2024; Ozcelik & Eksi, 2024) highlighted the importance of providing suitable prompts to ChatGPT so it could generate quality WCF for students' written work.

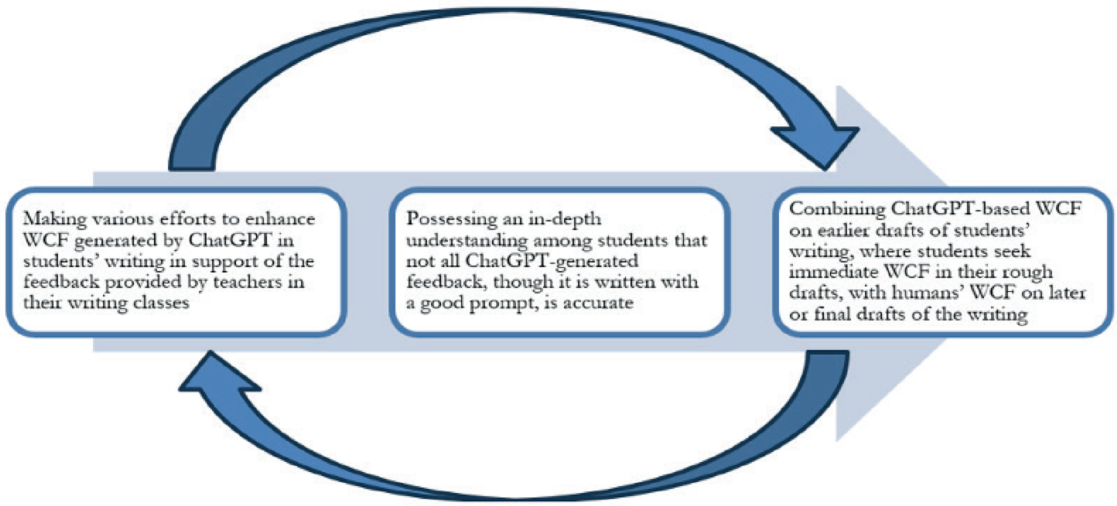

Figure 3. The Three Iterative Phases

Some studies presented sample prompts for ChatGPT that students could use to generate WCF for their writing work. Adopting Teng's (2024b) experimental classroom design, students and their lecturer can discuss and practice these prompt references together. For the prompts' references, see Figure 1 on p. 485 presented in Jackaria et al.'s study (2024); see Table 3 on p. 6 in Lin and Crosthwaite's (2024) study; and see Figure 1 on p. 5 in Mizumoto and Eguchi's (2023) study; the researcher intentionally located the exact pages, figures, or tables of the previous studies to help readers find the sample prompts more easily. In the second phase, there should be an in-depth understanding among students that not all ChatGPT-generated feedback, though it is written with a good prompt, is accurate. Thus, they always need to critically evaluate the feedback accuracy and usefulness before they integrate the feedback into their writing drafts (Jackaria et al., 2024; Ozcelik & Eksi, 2024; Steiss et al., 2024). In the third phase, after discussing and practicing with students how to write good prompts to generate WCF and to have an in-depth understanding of potential weaknesses of ChatGPT when providing WCF, lecturers should start to combine ChatGPT-based WCF on earlier drafts of students' writing, where students seek immediate WCF in their rough drafts, with humans' WCF on later or final drafts of the writing, which was clearly illustrated in Ozcelik and Eksi's (2024) study (i.e., see Figure 1 on p. 7 in their study). By doing so, lecturers could check and (if necessary) correct any inaccurate feedback that ChatGPT generated (Steiss et al., 2024).

DISCUSSIONS

As the answers to the first research question, the quality of WCF provided by ChatGPT compared to that of humans in EFL/ESL writing classrooms, the findings showed that both ChatGPT and humans might provide quality WCF to students' work in EFL/ESL classrooms. For example, ChatGPT can provide immediate descriptive WCF that focuses on various aspects of students' writing, such as sentence structure, vocabulary, spelling, punctuation, grammar, writing styles, ideas' organization, coherence, and quality of written expressions. Meanwhile, humans could provide WCF to students' work that ChatGPT might have struggled with, such as WCF related to creativity, reasoning, and emotional depth. Reflecting on these findings, the researcher needs to cautiously treat any related claims (e.g., by Guo & Wang, 2024; Li et al., 2024) that ChatGPT outperformed humans in providing comprehensive feedback to students' writing. Why? As reported in the literature reviewed in this study, ChatGPT and humans also have weaknesses that could affect the quality of feedback they generate. For example, ChatGPT was reported to give irrelevant comments, misinterpret writers' intent, add unnecessary items to writing text, and even change text (see Bui & Barrot, 2024; Ozcelik & Eksi, 2024; Steiss et al., 2024). At the same time, the researcher agrees with what previous studies (i.e., Awidi, 2024; Bui & Barrot, 2024; Yavuz et al., 2024) reported on humans' subjectivity in providing WCF. That subjectivity made humans provide WCF inconsistent with the grade they gave to students' written work and more permissive about errors in students' essays, especially those written by English language learners.

The inconsistency and accuracy of those WCFs are alarming. When receiving those types of WCF, students might feel discouraged and disengaged from revising their work (as concerned by Steiss et al., 2024). It is also alarming for EFL/ESL students, particularly those with a relatively low English proficiency, who decide to trust the WCF generated by ChatGPT entirely. As revealed by Ozcelik and Eksi (2024), students with that level of language proficiency could not understand why ChatGPT corrected their text as it looked intelligible to them. That said, the researcher asserts that the decision to use ChatGPT in providing WCF for students' writing should always be approached with prudence and continual human oversight. That view signals teachers' essential and irreplaceable role in this Artificial Intelligence era in education. Also, the researcher affirms that "it is essential for teachers to evaluate and supplement machine-generated feedback, even if the machine, like ChatGPT, is highly capable" (Teng, 2024b, p. 9).

Given the capabilities and weaknesses of ChatGPT and humans in providing WCF to students' written work, the researcher continues to discuss the second research question: *how can WCF provided by ChatGPT support WCF that humans give to students' written work in EFL/ESL writing classrooms?* The results of the literature analysis suggested three iterative phases. The first phase highlights the essence of making various efforts to enhance WCF generated by ChatGPT in students' writing in support of the feedback provided by teachers. One of the efforts was to input good prompts to ChatGPT. As explored in the previous studies (i.e., Awidi, 2024; Jackaria et al., 2024; Lin & Crosthwaite, 2024; Mizumoto et al., 2024; Ozcelik & Eksi, 2024), the quality of WCF generated by ChatGPT might be affected by how prompts were written and then assigned to it. Therefore, such claims that students in Teng's (2024b) study "perceived teachers' feedback as superficial, while considering ChatGPT to be more powerful in its feedback capabilities" (p. 9), very likely would not have happened if the students lacked the capabilities to write good prompts. As Tseng and Warschauer (2023) argued that, with a limited understanding of how to ask and talk with ChatGPT, students might ask ChatGPT to generate surface-level or overly general content that is less useful to fulfill their specific goal or task.

The second phase is to enhance students' in-depth understanding that not all ChatGPT feedback, although written with good prompts, is accurate; this highlights the need to evaluate the accuracy and usefulness of ChatGPT-generated feedback critically. Otherwise, completely trusting the WCF-generated ChatGPT is a huge mistake. As discussed previously, making the critical evaluation might be a big task for students with low language proficiency. At this point, the researcher raises a question about the qualities that students in EFL/ESL writing classrooms should have so that they can benefit most from WCF generated by ChatGPT and ways to help students in the classrooms achieve those qualities. Before reaching those qualities, should teachers consider not allowing students to use ChatGPT?

After working with prompts and enhancing students' understanding of potential weaknesses of ChatGPT-generated WCF, the third phase is to suggest that lecturers combine ChatGPT-based WCF, specifically on earlier drafts of students' writing, where students seek immediate WCF in their rough drafts with humans' WCF on later or final drafts of the writing. The early WCF on the drafts could be focused more on the technical aspects of students' writing, such as sentence structure, vocabulary, spelling, punctuation, grammar, writing styles, ideas' organization, coherence, and quality of written expressions, which are areas that ChatGPT might be good at. With this scenario, lecturers could still have chances to review any inaccurate feedback that ChatGPT generated on their students' written work (Ozcelik & Eksi, 2024; Steiss et al., 2024). Then, later in final drafts, lecturers, who were reported to experience fatigue from repetitive tasks of correcting students' written work (Lee, 2016; Shadiev & Feng, 2024) and be overloaded with daily administrative tasks, could spend their limited time focusing their WCF more on the content of the writing, commenting on creativity, the content of the writing, and reasoning made by students in their writing. Then, lecturers could give practical feedback to improve students' writing based on lecturers' expertise and experiences as a writer, strengths that humans WCF have (Bui & Barrot, 2024; Solovey, 2024; Teng, 2024b). With the harmonious WCF generated by ChatGPT and lecturers, students should have a higher-quality WCF beyond just obtaining grades (Han & Li, 2024) and having their writing errors underlined or circled with a red pen (as reported by Han & Li, 2024) without receiving clear guidelines on what the students specifically need to revise.

From the above discussions, the researcher can draw the following pedagogical implications, particularly for EFL/ESL lecturers in writing classrooms who wish to support their WCF with ChatGPT to provide quality WCF to their students' written work. First, the researcher emphasizes that the lecturers need to have a synergistic relationship with ChatGPT to provide quality WCF for their students in EFL/ESL writing classes, where ChatGPT should "serve as a companion, not enemies" (Teng, 2024b, p. 9). The lecturers should positively view ChatGPT as a companion that can reduce their workload when giving feedback on their students' writing (Han & Li, 2024). Second, with this positive mindset, the lecturers should be willing to allocate time to learn and practice using ChatGPT and try various prompts to provide the best WCF on their writing work. As far as Steiss et al. (2024) were concerned, it might take some time and multiple attempts at prompt engineering to obtain the best WCF from ChatGPT. With these practices, the lecturers should have an in-depth understanding of the kinds of WCF generated by ChatGPT and areas in which the lecturers could contribute more. Third, after practicing various prompts that work best with ChatGPT in providing the WCF, the lecturers should allocate time to practice the prompts with their students. Then, the lecturers should allow their students to use the prompts outside the class to ask ChatGPT to generate the best WCF on their writing work. When it worked well, the lecturers could spend more time providing WCF, focusing on their students' writing content. Fourth, there should be open discussions in the writing classrooms between lecturers and students about when ChatGPT would be used to provide WCF to students' written work. The discussions should also cover potential strengths and weaknesses of ChatGPT, as described in this study, and strongly remind the students to always review any WCF in their work generated by ChatGPT. With the discussions, students should no longer assume or feel skeptical that ChatGPT or their teachers generated the WCF provided to them. Last, lecturers who will teach the writing classes should be the ones who have a strong teaching commitment as they need to make time to learn various writing prompts independently and together with their students, as well as be there for their students to review any inaccurate feedback that ChatGPT generated on their students' written work (Ozcelik & Eksi, 2024; Steiss et al., 2024).

CONCLUSIONS AND RECOMMENDATIONS

The researcher has examined 23 journal articles (i.e., from 2023-2024) published in peer-reviewed academic journals on the WCF generated by ChatGPT and humans in EFL/ESL writing classrooms. Two key conclusions emerged as the answer to the research questions. First, each ChatGPT and human has strengths and weaknesses when providing WCF to students' written work in EFL/ESL classrooms. Thus, they should work harmoniously to complement each other's weaknesses to provide quality WCF for students in the writing classrooms. Second, ChatGPT-generated WCF can support human-generated WCF in students' written work in the writing classrooms through three main iterative phases. Those include making various efforts to enhance WCF generated by ChatGPT in students' writing in support of the feedback provided by teachers, enhancing students' in-depth understanding that not all ChatGPT feedback, although written with good prompts, is accurate, and combining ChatGPT-based WCF, specifically on earlier drafts of students' writing.

Although previous studies have provided valuable insights on the WCF generated by humans and ChatGPT in EFL/ESL writing classrooms, several questions remained unanswered in this study, offering directions for future research. Those questions include: What are the best prompting methods for ChatGPT to generate accurate and precise WCF for students' written work in the writing classrooms? What are the best ways for EFL/ESL students to evaluate the accuracy of WCF generated by ChatGPT? What strategies should EFL/ESL lecturers employ to address the weaknesses of ChatGPT and human-generated WCF identified in this study? How do EFL/ESL lecturers' perspectives on the WCF generated by ChatGPT? Do they feel threatened that ChatGPT will replace their roles? Why? Why not? What specific types of writing errors or components would benefit most from WCF generated by ChatGPT and humans?

Last, the researcher would like to acknowledge some limitations of this study. Most of the 23 studies reviewed in this paper were from journal articles in Asia contexts involving undergraduate students. Besides, most of the empirical studies reviewed in this paper relied on statistical analysis and survey methods to examine the WCF generated by ChatGPT and humans. To address these limitations, future researchers are encouraged to review studies on similar topics in African, Australian, American, and European EFL/ESL writing classroom settings that involve students at master's or doctoral levels, such as when they write their thesis or dissertation. Additionally, expanding the scope of reviewed materials to include books, book chapters, conference proceedings, and doctoral dissertations could provide richer perspectives on the topic. Podcast or YouTube videos might also be a source of ideas that should be explored. The researcher also suggests future studies to review similar studies using more diverse qualitative methods (i.e., narrative inquiry, autoethnography) that were not used by previous studies reviewed in this paper to obtain richer insights into the WCF generated by ChatGPT and humans to students' written work in the writing classrooms.

BIODATA and CONTACT ADDRESSES of AUTHOR

Yustinus Calvin Gai MALI is an associate professor in the Master's Program in English Language Education at the Faculty of Language and Arts, Universitas Kristen Satya Wacana (UKSW), Salatiga, Indonesia. He holds a PhD in Language, Literacy, and Technology Education from Washington State University, Pullman, WA, USA. His recent publications include: *exploring the use of ChatGPT in EFL/ESL writing classrooms: a systematic literature review* (Journal of Language & Education, 2025) and *EFL graduate students' voices on their technology-integrated classroom language tasks* (LLT Journal: A Journal on Language and Language Learning, 2024). Calvin has been conducting research in language, literacy, and education technology.

Yustinus Calvin Gai MALI
Master's Program in English Language Education, Faculty of Language and Arts, UKSW
Address: Universitas Kristen Satya Wacana, Jl. Diponegoro No. 52-60, Salatiga, Indonesia 50711
Phone: +62298-321212
E-mail: yustinus.mali@uksw.edu

REFERENCES

Abas, I. H., & Abd Aziz, N. H. (2018). Model of the writing process and strategies of EFL proficient student writers: A case study of Indonesian learners. *Pertanika Journal of Social Sciences and Humanities*, *26*(3), 1815–1842.

Ajlouni, A. O., Abd-Alkareem Fatima, W., & Almahaireh, A. S. (2023). Students' attitudes towards using chatGPT as a learning tool: The case of the University of Jordan. *International Journal of Interactive Mobile Technologies*, *17*(18), 99–117. https://doi.org/https://doi.org/10.3991/ijim.v17i18.41753

Alberth. (2023). The use of chatgpt in academic writing: A blessing or a curse in disguise? *TEFLIN Journal*, *34*(2), 337–352. https://doi.org/10.15639/teflinjournal.v34i2/337-352

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

Alkamel, M. A. A., & Alwagieh, N. A. S. (2024). Utilizing an adaptable artificial intelligence writing tool (ChatGPT) to enhance academic writing skills among Yemeni university EFL students. *Social Sciences and Humanities Open*, *10*, 1–9. https://doi.org/10.1016/j.ssaho.2024.101095

Alqasham, F. H. (2023). ChatGPT in the Saudi EFL classroom: A study of learner usage patterns and possibilities in learning optimization. *Migration Letters*, *20*(S7), 1251–1263. https://doi.org/https://doi.org/10.59670/ml.v20iS7.4828

Awidi, I. T. (2024). Comparing expert tutor evaluation of reflective essays with marking by generative artificial intelligence (AI) tool. *Computers and Education: Artificial Intelligence*, *6*, 1–17. https://doi.org/10.1016/j.caeai.2024.100226

Aydin, Z., & Yildiz, S. (2014). Using wikis to promote collaborative EFL writing. *Language Learning and Technology*, *18*(1), 160–180. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44359/1/18_01_aydinyildiz.pdf

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, *21*(23), 1–15. https://doi.org/10.1186/s41239-024-00455-4

Barrett, A., & Pack, A. (2023). Not quite eye to A.I.: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, *20*(1), 1–24. https://doi.org/10.1186/s41239-023-00427-0

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*, 1–6. https://doi.org/10.1016/j.asw.2023.100745

Baskara, F. R. (2023). Integrating ChatGPT into EFL writing instruction: Benefits and challenges. *International Journal of Education and Learning*, *5*(1), 44–55. https://doi.org/10.31763/ijele.v5i1.858

Bibi, Z., & Atta, A. (2024). The role of ChatGPT as AI English writing assistant: A study of student's perceptions, experiences, and satisfaction. *Annals of Human and Social Sciences*, *5*(I), 433–443. https://doi.org/10.35484/ahss.2024(5-i)39

Bonner, E., Lege, R., & Frazier, E. (2023). Large language model-based artificial intelligence in the language classroom: Practical ideas for teaching. *Teaching English With Technology*, *23*(1), 23–41. https://doi.org/10.56297/bkam1691/wieo1749

Bui, N. M., & Barrot, J. S. (2024). ChatGPT as an automated essay scoring tool in the writing classrooms: How it compares with human scoring. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12891-w

Callanga, C., Ares, J. M., Becbec, J., Elladora, S., Gabucan, J., Gaylan, E., Narca, M., Quibido, J., Quimat, R. M., Taneo, J. K., & Sanchez, J. M. P. (2024). Empowering doctoral students: The role of Publish or Perish software in enhancing systematic reviews in science education. *Internet Reference Services Quarterly*, 1–14. https://doi.org/10.1080/10875301.2024.2384021

Chandra, S. O., & Yuyun, I. (2018). The use of google translate in EFL essay writing. *LLT Journal: A Journal on Language and Language Teaching*, *21*(2), 228–238. https://doi.org/10.24071/llt.2018.210212

Cheong, W., & Hong, H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, *5*(1), 37–45. https://jeti.thewsu.org/index.php/cieti/article/view/103

Crcek, N., & Patekar, J. (2023). Writing with AI: University students' use of ChatGPT. *Journal of Language and Education*, *9*(4), 128–138. https://doi.org/10.17323/jle.2023.17379

Floden, J. (2024). Grading exams using large language models: A comparison between human and AI grading of exams in higher education using ChatGPT. *British Educational Research Journal*, 1–24. https://doi.org/10.1002/berj.4069

Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*, *13*(7), 1–13. https://doi.org/10.3390/educsci13070692

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education & Information Technologies*, *29*(7), 8435–8863. https://doi.org/10.1007/s10639-023-12146-0

Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System*, *126*, 1–11. https://doi.org/10.1016/j.system.2024.103502

Hang, N. T. T. (2023). EFL techers' perspectives toward the use of ChatGPT in writing classes: A case study at Van Lang University. *International Journal of Language Instruction*, *2*(3), 1–47. https://doi.org/https://doi.org/10.54855/ijli.23231

Harmawan, V., Mansyera Pratama, V., & Fadilah, R. (2023). The challenges of writing academic article: Voices of higher education students. *Journal of English Language Teaching*, *12*(2), 402–408. https://doi.org/10.24036/jelt.v12i2.120892

Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, *15*(4), 1–14. https://doi.org/10.30935/cedtech/13605

Jackaria, P. M., Hajan, B. H., & Mastul, A. R. H. (2024). A comparative analysis of the rating of college students' essays by ChatGPT versus human raters. *International Journal of Learning, Teaching and Educational Research*, *23*(2), 478–492. https://doi.org/10.26803/ijlter.23.2.23

Jukiewicz, M. (2024). The future of grading programming assignments in education: The role of ChatGPT in automating the assessment and feedback process. *Thinking Skills and Creativity*, *52*, 1–9. https://doi.org/10.1016/j.tsc.2024.101522

Laksmi, E. D. (2006). "Scaffolding" students' writing in EFL Class: Implementing process approach. *TEFLIN Journal*, *17*(2), 144–156. https://doi.org/10.15639/teflinjournal.v17i2/152-165

Lee, I. (2016). Teacher education on feedback in EFL writing: Issues, challenges, and future directions. *TESOL Quarterly*, *50*(2), 518–527. https://doi.org/10.1002/tesq.303

Li, J., Huang, J., Wu, W., & Whipple, P. B. (2024). Evaluating the role of ChatGPT in enhancing EFL writing assessments in classroom settings: A preliminary investigation. *Humanities and Social Sciences Communications*, *11*(1), 1–9. https://doi.org/10.1057/s41599-024-03755-2

Li, J., Jangamreddy, N., Bhansali, R., Hisamoto, R., Zaphir, L., Dyda, A., & Glencross, M. (2024). AI-assisted marking: Functionality and limitations of ChatGPT in written assessment evaluation. *Australasian Journal of Educational Technology*, *40*(4), 56–72. https://doi.org/https://doi.org/10.14742/ajet.9463

Li, M. (2012). Use of wikis in second/foreign language classes: A literature review. *CALL-EJ*, *13*(1), 17–35. http://callej.org/journal/13-1/Li_2012.pdf

Li, M. (2018). Computer-mediated collaborative writing in L2 contexts: An analysis of empirical research. *Computer Assisted Language Learning*, *31*(8), 1–23. https://doi.org/10.1080/09588221.2018.1465981

Lin, S., & Crosthwaite, P. (2024). The grass is not always greener: Teacher vs. GPT-assisted written corrective feedback. *System*, *127*, 1–19. https://doi.org/10.1016/j.system.2024.103529

Luo, W. A., Murtisari, E. T., & Isharyanti, N. (2024). Lower-proficiency EFL students' use of Grammarly in writing: Behavior, cognition, and affect. *JALT CALL Journal*, *20*(1), 1–25. https://doi.org/10.29140/jaltcall.v20n1.1089

Mali, Y. C. G. (2017). Promoting effort attributions to EFL students. *Accents Asia*, *9*(1), 30–40. http://www.issues.accentsasia.org/issues/9-1/Gai_Mali.pdf#:~:text=Attribution describes an explanation that learners provide for,role to promote the effort attributions to students.

Mali, Y. C. G. (2018). Digital video presentation in an EFL writing wlassroom. *Journal of Creative Practices in Language Learning and Teaching (CLPT)*, *6*(2), 1–21. https://ir.uitm.edu.my/id/eprint/30529/

Mali, Y. C. G. (2022). *Simple technology to support research*. Sanata Dharma University Press.

Mali, Y. C. G. (2023). EFL students' challenges in writing research proposals. *LLT Journal: A Journal on Language and Language Teaching*, *26*(1), 272–289. https://doi.org/10.24071/LLT.V26I1.5296

Mali, Y. C. G., & Salsbury, T. L. (2021). Technology integration in an Indonesian EFL writing classroom. *TEFLIN Journal*, *32*(2), 243–266. https://journal.teflin.org/index.php/journal/article/view/1558/354

Mao, Z., Lee, I., & Li, S. (2024). Written corrective feedback in second language writing: A synthesis of naturalistic classroom studies. *Language Teaching*, 1–29. https://doi.org/10.1017/S0261444823000393

Meniado, J. C., Huyen, D. T. T., Panyadilokpong, N., & Lertkomolwit, P. (2024). Using ChatGPT for second language writing: Experiences and perceptions of EFL learners in Thailand and Vietnam. *Computers and Education: Artificial Intelligence*, *7*, 1–9. https://doi.org/10.1016/j.caeai.2024.100313

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

Mizumoto, A., Shintani, N., Sasaki, M., & Teng, M. F. (2024). Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics*, *3*(2), 1–15. https://doi.org/10.1016/j.rmal.2024.100116

Ngo, T. T. A. (2023). The perception by university students of the use of ChatGPT in education. *International Journal of Emerging Technologies in Learning (IJET)*, *18*(17), 4–19. https://doi.org/https://doi.org/10.3991/ijet.v18i17.39019

Nguyen, H. H. B., Ngoc, H. H. B., & Dan, T. C. (2024). EFL students' perceptions and practices of using ChatGPT for developing English argumentative essay writing skills. *European Journal of Alternative Education Studies*, *9*(1), 168–216. https://doi.org/10.46827/ejae.v9i1.5341

Ozcelik, N. P., & Eksi, G. Y. (2024). Cultivating writing skills: The role of ChatGPT as a learning assistant—a case study. *Smart Learning Environments*, *11*(10), 1–18. https://doi.org/10.1186/s40561-024-00296-8

Purwanti, M. S., & Kastuhandani, F. C. (2023). Student's perceived confidence in writing publishable articles. *LLT Journal: A Journal on Language and Language Learning*, *26*(1), 153–167. https://doi.org/https://doi.org/10.24071/llt.v26i1.5407

Putri, S. A. M. D. U., Ratminingsih, N. M., & Santosa, M. H. (2023). A systematic review on Cake application for students' learning motivation and vocabulary mastery. *Journal of English Development*, *3*(02), 139–153. https://doi.org/10.25217/jed.v3i01.3533

Putrie, D. I., Salam, U., & Riyanti, D. (2024). Unveiling the landscape of ClassDojo in education: A systematic review. *Elsya: Journal of English Language Studies*, *6*(2), 1–24. https://journal.unilak.ac.id/index.php/elsya/article/view/18422/6317

Rahimi, M., Fathi, J., & Zou, D. (2024). Exploring the impact of automated written corrective feedback on the academic writing skills of EFL learners: An activity theory perspective. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12896-5

Ramadhanti, R., Oktaviana, F., Tarihoran, N., & Fitriani, D. (2023). Challenges, strategies, and digital media use in writing research articles: A study of EFL postgraduate students. *Scope: Journal of English Language Teaching*, *8*(1), 167. https://doi.org/10.30998/scope.v8i1.17385

Rasul, T., Nair, S., Kalendra, D., Robin, M., Santini, F. de O., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heathcote, L. (2018). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *The Irish Journal of Psychology*, *1*(1), 25–34. https://doi.org/https://doi.org/10.37074/jalt.2023.6.1.29

Richards, J. C., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics (4th ed.)*. Harlow.

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, *6*(1), 342–363. https://doi.org/10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023b). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, *6*(1), 342–363. https://doi.org/10.37074/jalt.2023.6.1.9

Shadiev, R., & Feng, Y. (2024). Using automated corrective feedback tools in language learning: A review study. *Interactive Learning Environments*, *32*(6), 1–27. https://doi.org/10.1080/10494820.2022.2153145

Slamet, J. (2024). Potential of ChatGPT as a digital language learning assistant: EFL teachers' and students' perceptions. *Discover Artificial Intelligence*, *4*(46). https://doi.org/10.1007/s44163-024-00143-2

Solovey, O. Z. (2024). Comparing peer, ChatGPT, and teacher corrective feedback in EFL writing: Students' perceptions and preferences. *Technology in Language Teaching & Learning*, *6*(3), 1–23. https://doi.org/10.29140/tltl.v6n3.1482

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14*, 1–14. https://doi.org/10.3389/fpsyg.2023.1260843

Steiss, J., Tate, T., Graham, S., Cruz, 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. https://doi.org/10.1016/j.learninstruc.2024.101894

Teng, M. F. (2024a). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies*, *6*(3), 36–57. https://doi.org/10.58304/ijts.20240304

Teng, M. F. (2024b). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, *7*, 1–10. https://doi.org/10.1016/j.caeai.2024.100270

Tseng, W., & Warschauer, M. (2023). AI-writing tools in education: If you can't beat them, join them. *Journal of China Computer-Assisted Language Learning*, *3*(2), 258–262. https://doi.org/10.1515/jccall-2023-0008

Tseng, Y. C., & Lin, Y. H. (2024). Enhancing English as a Foreign Language (EFL) learners' writing with ChatGPT: A university-level course design. *Electronic Journal of E-Learning*, *22*(2), 78–97. https://doi.org/10.34190/ejel.21.5.3329

Utami, N. L. G. F. P., & Santosa, M. H. (2023). Using VoiceThread to develop the speaking skills of EFL learners: A systematic literature review. *Jurnal Inovasi Teknologi Pembelajaran*, *10*(1), 83–91. https://doi.org/10.17977/um031v10i12023p083

Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, *8*(3), 1–12. https://doi.org/10.3390/languages8030212

Yavuz, F., Celik, O., & Yavas Celik, G. (2024). Utilizing large language models for EFL essay grading: An examination of reliability and validity in rubric-based assessments. *British Journal of Educational Technology*, *00*, 1–17. https://doi.org/10.1111/bjet.13494

Yeo, M. A. (2024). ChatGPT and the future of editorial writing. *RELC Journal*, *55*(1), 1–9. https://doi.org/https://doi.org/10.1177/003368822412405

Zadorozhnyy, A., & Lai, W. Y. W. (2023). ChatGPT and L2 written communication: A game-changer or just another tool? *Languages*, *9*(1), 1–11. https://doi.org/https:// doi.org/10.3390/languages9010005

Zain, D. S. M. (2022). Flipped classroom model for EFL/ ESL instruction in higher education: A systematic literature review. *Journal of Language & Education*, *8*(3), 133–149. https://doi.org/10.17323/jle.2022.12855