

ISSUES

CHATGPT-GENERATED CORRECTIVE FEEDBACK: DOES IT DO WHAT IT SAYS on the TIN?

Khaled EIEbyary^{1,2}, Ramy Shabara³

¹ Department of Education, University of York, ² Damanhour University, ³ Academic Services, University of Hertfordshire, UK

Keywords: corrective feedback, teacher feedback, ChatGPT

<https://doi.org/10.56297/vaca6841/BFFO7057/MYEH4562>

Teaching English with Technology

Vol. 24, Issue 3, 2024

Although some published studies have challenged the importance and efficacy of written corrective feedback (WCF), ESL/EFL tutors continue to value its practice and provide WCF to their students in spite of their time and workload constraints. The development of various automated-feedback applications over the last few years contributed to our understanding of what machine-operated feedback is capable of in comparison to teachers. The release of OpenAI's ChatGPT in 2022 has taken the world by surprise. This AI-operated tool has attracted the attention of students, teachers and researchers in the field of L2 teaching and learning. There are now claims that it has the capabilities for providing WCF on written output. The overarching aim of this study is to compare the quantity and quality of analytical corrective feedback provided and moderated by trained classroom teachers with ChatGPT-generated feedback on the written output of a group of students (English Level CEFR B1) attending an EAP course at an international branch campus of a UK university. Both teachers and ChatGPT used the same analytical marking criteria to mark students' essays and provide feedback on language use and use of academic vocabulary and style. Analysis of results provided valuable insights into the adequacy of the feedback strategy, guidance and polarity, offering a foundation and guidelines for future endeavours to optimize the integration of AI-generated feedback within L2 contexts. Implications and recommendations regarding the potentials/future of using ChatGPT as a feedback tool in L2 are discussed.

1. Introduction

ChatGPT has been defined as “a state-of-the-art language model ... designed to generate text that can be indistinguishable from text written by humans... [and it] can engage in conversation with users in a seemingly natural and intuitive way” (Rudolph et al., 2023). ChatGPT has also been described as the world's most advanced chatbot (G. Liu & Ma, 2023; Rudolph et al., 2023), which has the potential of revolutionizing EFL education (Koraishi, 2023; Ray, 2023) by driving change in learning goals, activities, and assessment practices (Zhai, 2022). In fact, different labels have been coined to mark the advent of ChatGPT and similar tools, such as the coming tsunami (Tate et al., 2023), the more knowledgeable other (Stojanov, 2023) and a new era of AI (Escalante et al., 2023).

Thus, various ChatGPT-related claims have been at the centre of heated debates in the L2 arena. This is justifiable because unanimous agreement is simply unexpected, especially that, up to now, little is known about the consequences of adapting/adopting ChatGPT as a language-learning tool over time. According to Escalante et al. (2023, p. 20), the release of ChatGPT marks a “new era” of AI use. These authors also state:

Given the critical role that writing plays in learning and assessment..., it is of growing importance for educators to make thoughtful and informed decisions as to how and in what capacity generative AI tools should be leveraged to assist in the development of students’ writing skills.

Because providing written corrective feedback (WCF) is an everyday practice for EFL teachers, one debated topic is the extent to which ChatGPT can provide adequate feedback on L2 writing. Guo and Wang (2023) state that few empirical studies have focused on the role of ChatGPT in helping writing tutors with feedback provision. Some recent studies, however, report on the positive impact of ChatGPT-generated feedback (Shaikh et al., 2023; L. Wang et al., 2024). However, Kushmar et al. (2022) argue that human feedback represents the social side of learning, but AI-operated tools would probably diminish these social and humanistic aspects. Given that more research is needed on whether ChatGPT can provide effective feedback on L2 writing, the aim of this study is to compare the quantity and quality of WCF provided by ChatGPT with teacher feedback on specific language areas using the same marking criteria.

2. Literature review

2.1. Background on WCF

Earlier views in EFL/ESL considered errors as flaws or the defective elements of the learner’s speaking and/or writing abilities (Dulay et al., 1982). However, the advent of the Communicative Approach led to the explicit view that errors are inevitable behaviours of L2 learners simply because they, as Lightbound (2005) explained, are triggered due to the nature of the language processing required. Teacher feedback, therefore, has generally been viewed by various educationists as a powerful tool (e.g., Ellis, 2009b, 2009a; Ferris, 2003; Hyland & Hyland, 2006; McMartin-Miller, 2014). It is defined as the responses and comments provided in reaction to the imperfections in L2 learners’ written output whereby they can recognize their errors and make appropriate corrections accordingly (Bitchener & Knoch, 2010; Li & Vuono, 2019). It represents a crucial component in the context of L2 pedagogy and has been the subject of extensive research in both SLA and L2 writing (Mao & Lee, 2020; T. Zhang et al., 2022). In fact, feedback is viewed as an intervention strategy and an effective pedagogical tool which can (a) support L2 students’ writing process (Bitchener & Storch, 2016; L. J. Zhang &

Cheng, 2021) and (b) positively influence writing quality (Bitchener, 2008; Bitchener & Ferris, 2012; Ellis et al., 2008; Fan, 2023), learners' interlanguage (Bonilla et al., 2018), and L2 acquisition (Bonilla, 2021).

Three main forms of WCF delivery have been generally acknowledged in the literature as direct, indirect, and metalinguistic (Ellis, 2009a). While direct WCF involves locating errors for learners and providing explicit corrections (Li & Roshan, 2019), indirect WCF identifies the errors without providing amendments (Luo & Liu, 2017) leaving space for learners to self-correct (Ellis et al., 2008; Q. Liu & Brown, 2015). On the other hand, metalinguistic WCF provides learners with clues (e.g., a comment, a brief explanation or a code) to explain the cause/nature of errors (Li & Roshan, 2019; Li & Vuono, 2019).

However, previous L2 research seems to have provided inconclusive evidence about the most effective WCF strategy (Bitchener, 2012; Cheng & Zhang, 2022; Ferris et al., 2013; Hanaoka & Izumi, 2012; Z. Wang & Han, 2022; Westmacott, 2017; Wondimm et al., 2023) although various studies (e.g., Bitchener, 2008; Bitchener & Knoch, 2008; Chandler, 2003; Sheen, 2007; Storch & Wigglesworth, 2010) have generally suggested the effectiveness of *direct* feedback and a preference for such a strategy among lower proficiency students (Amrhein & Nassaji, 2010; Lee, 2017). According to van Beuningen (2011), students' language proficiency as well as meta-linguistic awareness affect the extent to which these students could make use of WCF. However, other research evidence present either inconsistent results (La Russa, 2017) or no significant differences among WCF strategies (e.g., La Russa, 2017; Robb et al., 1986; Semke, 1984).

Another important issue that would enable, or possibly hinder, the beneficial side of any feedback strategy is what Bitchener (2012) referred to as conditions determining the effectiveness of written CF. In this context, Bitchener (2012) assembled theoretical insights derived from L2 acquisition studies (e.g., Pienemann, 1998; Schmidt, 1994; Skehan, 1998) which suggest that for feedback to yield benefits, students should (1) actively engage with the received WCF, (2) consciously identify disparities between their output and the feedback provided, (3) access the necessary linguistic information from their long-term memory, and (4) be developmentally prepared to assimilate the targeted forms and structures.

Equally important is the support, and sometimes the frustration, that can result from receiving comments (e.g., Truscott, 1996). It is what Ellis (2009b) described as the polarity of feedback in terms of negativity and positivity. In this context, positive feedback confirms the correctness of learners' output and consequently, establishes "affective support...and motivation to continue learning" (Ellis, 2009b, p. 3). Conversely, negative feedback is corrective in nature. Yet, various researchers emphasized the need for a supportive tone in WCF in order for it to be effective and beneficial (Al-Olimat &

AbuSeileek, 2015; Graham et al., 2015; Hattie & Timperley, 2007). The current researchers claim that such issues need to be considered, especially when AI-operated feedback is under scrutiny.

2.2. *AI-operated WCF and ChatGPT*

In the introduction of her book *Artificial Intelligence in Second Language Learning*, Dodigovic (2005, p. 1) stated:

Second language learners all over the world seem doomed to making errors, which clearly label them as non-native speakers of English, Chinese, Arabic, French, German, Indonesian or indeed any other language they are trying to acquire in addition to their mother tongue.

While this justifies the constant need for feedback, many language learners around the world could probably provide anecdotal evidence about teachers' intolerance of language errors or probably incidents of students taking the feedback personally. In addition, it is not unusual to hear that teacher feedback on multiple drafts is not viable in numerous EFL contexts. The advancement in technological tools paved the way for various automated writing evaluation (AWE) systems which provide instant feedback on students' written output. Technology-based feedback was viewed to have medium high to high effects on learning (Hattie & Timperley, 2007; Stevenson & Phakiti, 2024). Some of these include Criterion (<https://www.ets.org/criterion/about/>), MyAccess (<http://www.vantagelearning.com/>), WriteToLearn (<https://www.pearsonassessments.com/professional-assessments/products/programs/write-to-learn.html>), Write & Improve (<https://writeandimprove.com/>), Pigai (<http://en.pigai.org/>), Grammarly (<https://www.grammarly.com/>), PaperRater (<https://www.paperrater.com/>), and Online Correction (<https://www.onlinecorrection.com/>).

Writing research suggests that most AWE tools are capable of detecting surface-level errors such as grammar and mechanics, but they are less effective in dealing with content and organization issues (deep-level errors) (Fu et al., 2022; Gardner et al., 2020; Yoon et al., 2023). Furthermore, the language proficiency of students can play a detrimental role in how well they understand/comprehend the automated feedback (Fan, 2023). Therefore, research suggests that AWE tools would best be employed to supplement teacher feedback (Jiang et al., 2020).

However, the advent of ChatGPT (<https://chat.openai.com/>) caused a stir among teachers, learners and researchers. There were reports indicating that it had more than 100 million users within just two months (Guo & Wang, 2023; Yeo, 2023) causing a “sudden spike” (Koraishi, 2023, p. 55), which “has taken the world by storm” (Hong, 2023) and is currently drawing unprecedented worldwide attention (Ma & Huo, 2023) to the extent that

it is estimated to have a huge consequence on education (Yoon et al., 2023) and on every aspect of society (Zhai, 2022). In other words, it is yet another tool that is described to cause a paradigm shift (Meyer et al., 2023) which necessitates “teaching our students how to effectively partner with AI” (Tseng & Warschauer, 2023).

Several publications support the premise that ChatGPT has the potential of enhancing language learning outcomes and learning experiences (Barrios-Beltran, 2024; W. Liu, 2024; Mohamed, 2023). The release of ChatGPT has led to a surge of research on a wide range of topics including ChatGPT acceptance (e.g., G. Liu & Ma, 2023), L2 course/material design (e.g., Kim et al., 2023), independent language learning (e.g., Jiamiao & Li, 2023) and, most importantly to this paper, WCF on writing (e.g., Barrot, 2023; Warschauer et al., 2023).

2.3. Previous studies on use of ChatGPT in WCF

Some studies explored teachers’ and students’ views/acceptance of ChatGPT use in L2 in general and in writing classes in specific. Lui and Ma (2023) used Davis’s (1989) technology adoption model (TAM) to explore the perspectives and behaviour of 405 EFL learners. Findings support views of the potential capabilities of ChatGPT as a language learning tool. Similarly, Xiao and Zhi’s (2023) study suggested that learners considered ChatGPT “a valuable learning partner” which is capable of helping in language task completion. Analogous positive students’ attitudes towards ChatGPT usability in language learning were reported by other researchers (Shaikh et al., 2023).

As for tutors, Mohamed (2023) examined 10 Northern Border University faculty members’ perceptions of ChatGPT’s efficiency in enhancing L2 learning. Interview data revealed discrepancies in participants’ responses as some acknowledged its usefulness in facilitating L2 learning while others showed concerns about hampering language development and possibly reinforcing biases or misinformation. Likewise, based on the challenge of limited individual teacher–student interaction in large classes (40 plus), Nguyen’s (2023) study used a questionnaire and a structured interview to examine how EFL teachers evaluate their experiences with ChatGPT use in writing classes in Vietnam. In addition to teachers’ use of ChatGPT to create learning resources and lesson plans, results indicated perceptions of a positive impact on students’ writing processes. Similar findings about teachers’ and students’ positive perceptions of ChatGPT use in education at large (Chan & Hu, 2023) and in L2 writing were reported by Barret and Pack (2023) who also emphasized the need for teacher CPD programmes to provide unambiguous strategies on the use of GenAI.

In relation to the ChatGPT use in WCF, Guo and Wang (2023) explained that the emergence of ChatGPT, as well as other AI-assisted tools, has identified a new big gap in our understanding of what machines can do.

To these authors, research on AWE revealed that such applications are less competent than teachers, but the advent of cutting-edge AI technologies such as ChatGPT makes it essential to re-evaluate our understanding and to weigh the feedback of teachers against that provided by ChatGPT.

In their study, Guo and Wang (2023) examined the ability of ChatGPT to provide feedback on L2 students' writing and to recognize whether or not it has the potential to support teacher feedback. The study initially compared ChatGPT feedback with that of five Chinese EFL teachers on 50 argumentative essays (300 words each), and then it explored teachers' views about the feedback in an attempt to envisage the future integration of ChatGPT in WCF. A three-staged prompt was used with ChatGPT including "please provide comments and suggestions on (a) the content, (b) organisation and (c) language of the argumentative essay". Results from ChatGPT and teacher feedback were categorized into language (grammar/word-choice/mechanics), organization (paragraph topic sentences/summarizing/concluding) and content (argument development/supporting evidence/logical reasoning). A teacher questionnaire was used to judge the quality of ChatGPT feedback and to examine whether this tool can support teacher feedback. Their findings suggested that ChatGPT produced significantly more feedback comments compared to the five teachers. Teachers focused more on content, while ChatGPT distributed comments on content, organization, and language.

However, the use of marking criteria was not particularly clear in Guo and Wang's (2023) study, and their prompts seemed simplistic, suggesting that the AI-system drew on its own marking scale. In fact, the authors of the current paper believe that ChatGPT's strength lies in asking the tool to assess students' writing in light of teachers' prompts and the possibility of asking the AI system to adopt, adapt, explain or even modify its responses accordingly. In this context, Barrot (2023) states:

Quality feedback is not always available to the students. ChatGPT addresses this need as it offers useful features linked to writing assessment. This AI tool can automatically grade students' written work. The system assigns a score to the paper based on predefined criteria and provides specific comments to support the assigned score... By prompting the tool to provide feedback (e.g., Provide feedback on this text), it highlights the strengths and weaknesses of the text and offers recommendations on how the paper can be improved.

In another study, Yoon et al. (2023) emphasized the limitations of AWE systems that preceded ChatGPT in providing personalized feedback on content and coherence. They put forward a claim that writing teachers are enthusiastically using ChatGPT as a tool for grading and generating feedback on students' essays. They examined ChatGPT-generated feedback

on the coherence and cohesion of 50 argumentative essays written by grade 12 L2 learners. Eliciting data from the English Language Learner Insight, Proficiency, and Skills Evaluation (ELLIPSE) Corpus, the researchers used a prompt including a student's essay as input for ChatGPT. The study findings suggest that while a strong correlation with human scores was demonstrated, feedback from ChatGPT (a) detected surface-level use of cohesive devices, (b) was not particularly effective in terms of coherence and cohesion, and (c) was short of actionable steps for development.

Steiss et al. (2023) compared the quality of human and ChatGPT feedback on 200 secondary students' compositions. The study focused on comparing the extent to which feedback was accurate and criteria-based, delivered advice for development, highlighted necessary features, and used a supportive tone. Although human raters provided better feedback in all areas apart from criteria-based, the study suggested insignificant differences between human and ChatGPT feedback. However, the practical side of ChatGPT's ease of use put it in favour.

In summary, there has been a plethora of personal websites, blogs, Facebook groups and other online sources that provide anecdotal accounts about ChatGPT's capabilities of providing effective and timely WCF, but research evidence has been limited thus far. Available research reported positive attitudes, acceptance and capabilities of providing accurate feedback (e.g., Guo & Wang, 2023; Shaikh et al., 2023; Steiss et al., 2023).

However, a gap in research seems to suggest that, compared to teachers, we still need to examine whether ChatGPT is capable of providing effective WCF to suit L2 learners with specified language abilities. To fill this gap, this exploratory study still compares teacher feedback with ChatGPT-generated feedback, but it also examines whether the AI system is capable of (a) using a specific feedback strategy, (b) generating comprehensible and constructive input to learners at the B1 level, and (c) providing enough guidance on feedback.

3. Methodology

3.1. The aim of the study

The study had one main research question: To what extent do ChatGPT-generated analytical feedback comments on written output differ from/agree with teacher-moderated feedback comments in terms of the quality of feedback which included (a) adequacy of feedback strategy and guidance, and (b) comprehensibility and polarity (positive/negative) of feedback comments?

3.2. Participants

The dataset included 32 cause-effect essays, randomly selected from 599, written by EFL undergraduates in an international foundation program at an Egyptian branch university. Participants, aged 18-19, were native Arabic speakers with CEFR B1 English proficiency and diverse high school backgrounds (IGCSE, American Diploma, public schools).

The written essays were part of students' routine formative assessment activities with topic selection and word count (i.e., 350-400 words) and based on a source-based prompt. Students wrote at home, citing sources, and submitted digitally within a week. All personal identifiers were removed from the data prior to ChatGPT use. The researchers also attained ethics approval from students and teachers to participate in the study.

3.3. Design and procedure

The departure point for the researchers in this study was to have a conversation with ChatGPT about what it already knows of established literature on feedback and its capabilities in relation to WCF. ChatGPT assured the researchers about its capability of providing corrective feedback. It not only showed awareness of the need for marking criteria, but offered its own rubric. ChatGPT was also aware of CEFR, the CEFR levels and the appropriate feedback strategies for each level.

For the writing corpus, this study used the WCF provided and moderated by five experienced classroom teachers on participants' written output. Teachers' consent was obtained after WCF was provided. In other words, because WCF was intended to be compared to ChatGPT-generated WCF, teachers provided feedback to participants and then they were invited to participate in the study through an email containing the information sheets and consent forms.

Teachers' WCF was guided by an in-house set of standardized marking-criteria, which was also included with a student's essay as input in a three-stage prompt provided for ChatGPT. The criteria focused on three areas: (1) organization, content and relevance; (2) language use; and (3) communicative quality, use of academic vocabulary and style. Individual essays were given to ChatGPT following the conversation referred to earlier. However, the prompt focused only on the explanation of the marking criteria and no limits/further instructions were provided in relation to feedback length. [Figure 1](#) below provides the first stage of the prompt.

Imagine you are an English for Academic Purposes teacher at university level. Your student is B1 on the Common European Framework of Reference. The student was given a topic to write about. The topic is "Some students prefer to have new skills and a useful experience. Therefore, many students try to have a job in their leisure time. Discuss the causes and the effects of getting a job during study". You should use the following marking criteria to provide corrective feedback on Language use. A score of 95 or 85 refers to "An impressive/excellent range of simple and complex sentences used; no or few errors." A score of 77 or 72 refers to "A very wide range of simple and complex sentences both used with minor errors". A score of 68, 65 or 62 refers to "A wide range of simple and complex sentences used; simple sentences mainly error-free; complex sentences contain a few inaccuracies". A score of 58, 55 or 52 refers to "An appropriate range of simple sentences with few grammatical and/or punctuation errors. Complex sentences attempted but with some inaccuracies". A score of 48, 45 or 42 refers to "An adequate range of simple sentences with a few grammatical and/or punctuation errors but they rarely impede communication. Complex sentences attempted but with many inaccuracies". A score of 38, 35 or 32 refers to "A limited range of sentences, many grammatical and/or punctuation errors which impede communication". A score of 25, 10 or 0 refers to "Pre-Foundation Essay". Mind that the higher the score the better the writing should be within each band. Use this marking criteria only to provide individualized feedback and a score. The feedback strategy (i.e., direct feedback) must be geared to suit B1 level.

Shall I give you the essay now?

<INSERT ESSAY>.

Figure 1. ChatGPT prompt (language use)

4. Findings

The research question posed in the study aimed to qualitatively compare ChatGPT-generated feedback comments on students' written outputs with teachers' feedback with both parties using the same marking criteria. The comparison is presented here in terms of (a) adequacy of feedback strategy and guidance, and (b) accuracy and polarity (positive/negative feedback).

4.1. *Adequacy of the feedback strategy and guidance*

Data analysis aimed to examine the adequacy of the feedback strategy as well as the guidance provided to B1 participants. As mentioned in the literature review section, there are no best feedback strategies due to many factors that relate to teachers' workload, context and others. However, there seems to be a preference for direct feedback for limited proficiency students such as those involved in the current study. Direct feedback was the strategy adopted by teachers involved in this study for B1 students. The conversation the current researchers had with ChatGPT about suitability of feedback strategy to B1 learners showed the system's awareness of the issue. ChatGPT was, therefore, prompted to provide direct feedback to participants.

Data analysis suggests that teachers highlighted and struck through texts; and in most cases they were more capable of providing direct feedback on the written output. However, some of their comments were also metalinguistic. In contrast, ChatGPT tended to either rewrite correct forms of whole sentences or provide guidance on errors which were viewed as more guided metalinguistic in nature. Some examples are provided in [Table 1](#).

Table 1. Discrepancy of in teachers & ChatGPT feedback on language use

Student's written output	Teacher input & guidance	ChatGPT input & guidance
Example#1		
Many students prefer to work while studying to cover their needs. Working with studying it may be because need for money or maybe to gain new communication skills learn how to be independent person, part-time jobs may have many positive or good things but also have negative or problem like how to control time between work and studying and it can cause physical and metal issues, but there are people cannot afford their studying materials, so they have to work and be more confident.	Many students prefer to work while studying to cover their needs. Working with studying it [This] maybe because students [who needs money] need for money or maybe to gain new communication skills [and] learn how to be independent person , part-time jobs may have many positive or good things but [also] have negative or problem [problematic effects] like [Non-academic] how to control time between work and studying and it can cause physical and metal issues, but there are [some] people [who] cannot afford their studying materials, so they have to work and [to] be more confident.	The sentence is overly complex and contains multiple errors. Consider breaking it down into simpler sentences for better clarity. Errors include grammatical issues (e.g., "Working with studying it maybe") and confusion between "metal" and "mental."
Example#2		
This essay will discuss the causes and effects of part-time jobs with studying.	This essay will discuss the causes and effects of part-time jobs with studying [during study].	Minor improvement: Consider using "while studying" instead of "during studying" for smoother phrasing.
Example#3		
One of the most causes of part-time jobs is money need to cover their needs, like there are students cannot afford their college materials and there are students raised in poor families or immigrants, so they must be able to earn money and know how to control this money.	One of the most causes of part-time jobs is money need [the need for money] to cover their needs , like [For example] there are students cannot afford their college materials and there are students raised in poor families or immigrants, so they must be able to earn money and know how to control this money [manage it].	Simplify sentences for better flow and clarity. For example, "One of the main causes of part-time jobs is the need for money to cover expenses. Some students cannot afford college materials, while others are raised in poor families or are immigrants. As a result, they must earn and manage their money."
Example#4		
Some students prefer to have new skills and a useful experence ,therefore many of students try to have a job in there freetime furthermore have part time job such as shops,hospitals,restaurants ,perhaps in this essay will discuss the causes and the effects of get a job during there studies that's make many effects will asked in the essay.	Some students prefer to have new skills and a useful experence [spelling mistake], therefore [add a comma] many of students try to have a job in there freetime [leisure time] [Stop here and start a new sentence] furthermore have part time job such as shops,hospitals,restaurants [,perhaps in] this essay will discuss the causes and the effects of get [getting] a job during [there] studies [that's make many effects will asked in the essay].	Revise for grammar and coherence.
Example#5		
Working in a part time job during study its hard and important for many reasons,first of all the most important reason is helping there parents,to use there freetime in something useful,buying what they ever want without obstruction,its improve there confidence,some students have some	Working in [Having] a part time job during study its hard and important [is challenging] for many reasons [start a new sentence],first of all [add a comma] the most important reason is helping [there] parents,[space]to use there freetime in something [think of a better academic word] useful, buying what they ever want [such as] without obstruction,its improve [It improves] there confidence,[start a new sentence] some students have some problems [find it	Revise for grammar and coherence. Consider using "it's" instead of "its" for "it is improve," and "their" instead of

Student's written output	Teacher input & guidance	ChatGPT input & guidance
proplems to balance there life in another word they cant manage there time ,for that reason part time job helps them ,according to (mihail2005) the part time job something important to these students to reduce the amount of stress and to balance there life.	challenging] to balance there life in another word [In other words] they cant [cannot] manage there time , for that reason [Therefore] part time job helps them , according to (mihail2005) [According to Mihail (2005)] the part time job something important to these students to reduce the amount of stress and to balance there life.	"there" for possessive pronouns. Additionally, the phrase "buying what they ever want without obstruction" could be clarified for smoother expression. Great effort!
Example#6		
Part time job refers to working position in few hours are put in per week, can offer opportunities to earn income which can help students to cover their needs of living. The work during study for him positive effects and negative effects of the living and the importance of making this decision and this essay will discuss the causes and effects of part time job.	Part time job refers to working position in [for] few hours are put in per week, [Please join the two sentences correctly or separate them] can offer opportunities to earn income which can help students to cover their needs of living [living needs/life expenses]. The work during study for him positive effects and negative effects of the living and the importance of making this decision [elaborate this idea in a separate sentence] and this essay will discuss the causes and effects of part time job [on university students].	Some sentence structure issues; revise for smoother expression.

There were also various cases in which ChatGPT's comments on surface-level errors agreed exactly with teachers' direct feedback, but the AI system often provided lengthier comments with some guidance, which sometimes showed more precise input on the sentence level. Some examples are provided in [Table 2](#).

Table 2. Similarity between teachers and ChatGPT feedback on language use

Students' written output	Teacher input & guidance	ChatGPT input & guidance
Example#7		
Nowadays parents face difficult problems to pay for there young adults the university fees.	Add comma after Nowadays	Nowadays, Consider adding a comma after "Nowadays" for better punctuation.
Example#8		
55% of student work as a part time job all over the world.	students	Student students
Example#9		
The primary cause of this phenomena is that students who work on campus have more chances to socialise with teachers and classmates outside of the classroom.	phenomenon	Minor errors: Consider using "phenomenon" instead of "phenomena" for correct singular form. Add a comma after "classroom" for better punctuation.

Similarly, ChatGPT was more beneficial in relation to feedback on vocabulary, style and academic tone.

Therefore, teachers adhered to providing direct feedback in most cases despite deviating sometimes in which their comments lacked guidance and were posed as questions, presumably in an attempt to get their students to think more about their output (see Example 14). However, ChatGPT, compared to teachers, provided sufficient explanation and guidance on most comments.

Table 3. Teachers and ChatGPT feedback on vocabulary

Students' written output	Teacher input & guidance	ChatGPT input & guidance
Example#10		
Part-time jobs motivate students who are lazy to study.	No feedback	Change "lazy" to "reluctant" for a more formal tone.
Example#11		
If you do not balance between work life and study, you will fail in both.	If you [Excessive use of non-academic language] do not balance between work life and study, you will fail in both.	Change "between" to "your."
Example#12		
"Changes in energy, lack of sleep, hunger, loss of interest or pleasure in activities, difficulties focusing, and feelings of sadness, emptiness, hopelessness, worthlessness, or guilt are all symptoms."	No feedback	The essay communicates ideas but lacks a more formal and academic style. Enhance the formal tone and sentence complexity for academic writing. Corrected form: "Symptoms include variations in energy levels, sleep deprivation, appetite changes, loss of interest in activities, concentration difficulties, and emotional manifestations like sadness, emptiness, hopelessness, worthlessness, or guilt."
Example#13		
"To conclude, this essay discusses two causes and two effects of student employment."	No feedback	Comment: Formality Corrected form: "In conclusion, this essay has highlighted two causes and effects of student employment, emphasizing the impact on students' lives. The causes include familial financial pressures and the impact of jobs on high GPA achievers, highlighting the importance of time management skills."

In examining the length of comments, a simple descriptive analysis revealed that ChatGPT provided varied lengths of comments across the language areas involved. The lengths of the mean and median were 83 and 85 words for organization, 53 and 55 for language use, and 47 and 48 for vocabulary. These values suggest a consistent and moderate length in feedback, with some variability. [Table 4](#) illustrates some of these examples.

Unlike results from previous AWE studies in relation to feedback on content and organization, the data analysis revealed that ChatGPT generally provided useful comments mainly on coherence and occasionally on cohesion. A typical error of the participants in this study was long sentences. Examples of ChatGPT guidance were as follows:

- The sentence is lengthy and lacks clarity due to multiple ideas packed together. It needs restructuring for coherence. Consider breaking it down into smaller, more focused sentences to enhance readability and coherence.
- The essay has a clear introduction, presenting the causes and effects of student employment. However, the body paragraphs lack depth and coherent subdivisions to discuss each cause and effect with

Table 4. Teachers and ChatGPT feedback guidance

Students' written output	Teacher input & guidance	ChatGPT input & guidance
Example#14		
"Part- time jobs can affect students' life negatively if it causes sleep deprivation."	Comment: the first paragraph should discuss the causes?	Comment: The sentence structure and vocabulary used are generally clear, but there are some grammatical errors and imprecise expressions that could be revised for better coherence. Corrected Form: Use "Part-time" consistently. Change "A part-time jobs" to "Having a part-time job." "Part-time jobs can have a negative impact on students' lives if they cause sleep deprivation."
Example#15		
"Recognizing the importance of gaining and develop essential life skills and experiences..."	None	Comment: there are grammatical errors and awkward expressions that affect the clarity of ideas. Corrected Form: "Recognizing the importance of gaining and developing essential life skills and experiences..."
Example#16		
Around the age of twelve, a lot of young people being their informal employment,	"being" is highlighted by the teacher, but no comment.	Comment: Minor grammar issue: 'Being their informal employment' should be 'begin their informal employment.'
Example#17		
Part-time jobs typically do not get as much attention, but the experience is often just as valuable and useful being in a part-time job can affect your life positively or negatively Part Part-time jobs motivate students who are lazy to study.	Pay more attention to sentence and grammar structure. Be attentive to the spelling, sentence structure, and punctuation. Keep a variety of structure. Avoid using personal pronouns.	None.

clarity. Subdivide the body paragraphs into clear sections to discuss each cause and effect separately. This will enhance the organization and readability of ideas.

- The sentence lacks precise academic vocabulary and clarity. It would benefit from more academic phrasing and specific examples to support the argument.
- The sentence lacks a clear transition or link to the subsequent idea. Consider a more explicit link between balancing studies and jobs as a distinctive aspect.

Table 5. ChatGPT comment comprehensibility

	A1	A2	B1	B2	C1	C2
Essays	64%	9%	14%	8%	2%	0%
ChatGPT	45%	11%	19%	12%	3%	5%

As seen in the examples above, ChatGPT provided good advice on breaking the long sentences into shorter ones to enhance readability. However, ChatGPT also rewrote whole sentences for most students despite being trained to just provide feedback. While this might be a good model to use, it might inhibit students' responses to feedback and could initiate some ethical concerns too. In other words, the concept behind providing feedback is to increase students' abilities to notice their errors and produce correct language in subsequent drafts.

4.2. Feedback comprehensibility

Bitchener's (2012) accounts, referred to earlier in this paper, provided a foundational framework for conditions under which feedback can be beneficial, i.e., L2 learners need to comprehend comments to maximize opportunities for uptake. The language used in feedback should enhance the student's understanding of why an element of the written output is incorrect and how the student can rectify the error using pertinent linguistic information. This is particularly crucial when explicit correction/explanation is provided for lower proficiency students who need support in understanding the nature of their errors and how to improve them. Using VocabularyKitchen (<https://www.vocabkitchen.com/>), a corpus was created from the students' essays, and another was created from the comments provided by ChatGPT. VocabularyKitchen was used to examine the level of the comprehensibility of the language used in the corpora (see Table 5). There was no need to do the same with teachers' comments as they were generally very simple in nature.

The results obtained were informative. Although 8% (3% at C1 and 5% at C2) of the comments included lexical items far beyond the language level of the students involved in the study, 75% of the comments were provided within a comprehensible range and 12% pushed the boundaries of that range towards B2. This means that the language used by ChatGPT to comment on errors was generally easy to understand and could consequently lead to uptake. However, there was no control on the frequency of difficulty level within every single comment. For example, a 75-word comment included nine lexical items at B2, one at C1 and two at C2. However, this was beyond the focus of the study.

Table 6. Feedback polarity

Students' written output	Teacher input & guidance	ChatGPT input & guidance
Example#18		
There are many causes for the students who work a part of time.	A part of the time Part-time	Good topic sentence. Clear expression. Minor improvement Use 'part-time' consistently. Change 'of time' to 'of their time.'
Example#19		
Many students want to be more confident and become independent and learn how to be responsible for themselves and become more confident their interactions with others. learn how to work with people and gain great experience with people which can improve their language that consequently help in the academic studies.1	None	Positive intentions and ideas expressed. Minor improvement Repetition and structural issues. Consider revising for clarity and coherence. For example, "Many students aim to gain confidence, independence, and responsibility. Developing interpersonal skills and gaining valuable experience can enhance their language abilities, consequently benefiting their academic studies."
Example#20		
Firstly, a lot of students work during their college years these days for a variety of reasons, such as the growing expense of attendance and the pressure to gain experiences and skills that will be useful in their future employment before they graduate.	None	Well-structured paragraph with clear points. Minor improvement Consider using "first" instead of "firstly" for consistency.
Essay#21		
Part-time jobs are essential for students to help with financial problems and to gain experience.	None	Clear introduction presenting the main topic.
Essay#22		
Other cause that many students have a poor family, so he has to work part of time to help them.	Other Another cause is... a poor family poor families No concluding sentences	Varied sentence structures used effectively. Minor improvement Consider revising "Another cause that many students have poor Families, so they have to work part time to help them."

4.3. Polarity of feedback

The use of positive comments and positive tone (Graham et al., 2015) is often preferred, as they could be motivating for students. Results obtained in this study suggest that both teachers and ChatGPT generally provided accurate comments on students' written work although teachers provided shorter comments in almost all their feedback. Additionally, ChatGPT was good at spotting surface level errors (e.g., spelling), which were sometimes overlooked by teachers. The feedback generated by ChatGPT was both positive and negative. Interestingly, ChatGPT was able to provide more supportive annotations to students whereas teachers seemed to use less sympathetic comments. [Table 6](#) provides some examples from ChatGPT and teachers.

Thus, in relation to negative feedback, ChatGPT was highly constructive in providing the comments, and generally it added some guidance that could benefit correction in subsequent drafts. When such a dimension is compared

to teacher feedback, the concise nature of teachers' comments made such constructivist responses less noticeable or even non-existent. To sum up, the findings suggest that teachers' feedback was generally accurate and concise, but less sympathetic. However, ChatGPT offered immediate accurate and supportive comments, while also acknowledging erroneous output in the students' written work. It also aligned itself to a great extent with established principles of effective WCF.

5. Discussion and conclusion

While teacher feedback has been acknowledged as a beneficial pedagogical practice in L2, the findings of this study present intriguing insights into the effectiveness of the WCF provided by ChatGPT. These findings highlighted both strengths and areas for improvement in using this tool. The study examined the quality of feedback provided by ChatGPT and teachers and compared them in terms of adequacy of feedback strategy and guidance, and polarity of feedback. The comparison revealed that teachers, mandated by the department to adopt a direct strategy in their comments, demonstrated a greater capability to offer direct feedback. In fact, teachers involved in this study generally outperformed ChatGPT in providing direct feedback, with a specific focus on correcting errors. Conversely, ChatGPT tended to provide more metalinguistic feedback, often rewriting correct sentences and offering guidance that was perceived as less direct. The authors' speculation here is ChatGPT's inclination towards metalinguistic guidance may be attributed to its training data, which may prioritize overall sentence structure and language flow rather than specific error identification.

Nevertheless, ChatGPT's comments occasionally aligned with teacher feedback, demonstrating a capacity for precise input at the sentence level. These findings correspond with studies on automated feedback (Fan, 2023), highlighting the capacity of AI systems to offer comprehensive and informative commentary. ChatGPT could provide lengthier comments with useful guidance. Its comments exhibited depth in language analysis, offering valuable insights into coherence and, occasionally, cohesion. This suggests a new dimension of the potential of AWE and AI-generated feedback as it is less consistent with the notion that AI systems can contribute to deep-level errors (Jiang et al., 2020).

Regarding comprehensibility, ChatGPT's comments were predominantly within a reasonable range, making them generally easy to understand. Despite instances where the language used pushed towards a higher proficiency level, the majority of comments were deemed accessible to B1-level students, potentially facilitating uptake. The lack of control over difficulty levels within individual comments suggests an area for improvement in the system's design. Interestingly, ChatGPT demonstrated a notable strength in providing supportive annotations, indicating a more sympathetic tone compared to

teachers. The AI system's constructive negative feedback, coupled with additional guidance, holds promise for enhancing student learning and subsequent draft corrections.

Although the current research only focused on a relatively small number of essays, findings offer foundations for future endeavours to optimize the integration of AI-generated feedback within language learning contexts. While ChatGPT can provide valuable insights, its tendency towards a metalinguistic approach may not fully meet the directness often required for learners at lower proficiency levels. Nonetheless, the conversational nature of ChatGPT and its ability to adapt prompts would benefit integration in WCF practice provided that L2 teachers receive adequate training on prompt design.

The authors suggest that a blended approach, combining the immediacy and efficiency of ChatGPT-generated feedback with the precision of teacher-moderated feedback, emerges as a potential solution to enhance WCF. Furthermore, future research can examine the impact of AI-generated feedback on students' motivation and engagement in the learning process. Yet, two major elements are needed and these are the focus on ethical considerations in AI-generated feedback (e.g., privacy, data security, consent) and examining the effect of constant ChatGPT-generated feedback use over a length of time and its impact on students' writing.

Submitted: December 09, 2024 EET



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

REFERENCES

- Al-Olimat, S., & AbuSeileek, A. (2015). Using computer-mediated corrective feedback modes in developing students' writing performance. *Teaching English with Technology*, 15(3), 3–30. <https://doi.org/10.33985/0531-021-002-015>
- Amrhein, H. R., & Nassaji, H. (2010). Written corrective feedback: What do students and teachers think is right and why? *Canadian Journal of Applied Linguistics*, 13(2), 95–127.
- Barrios-Beltran, D. (2024). Enriching the teaching-learning experience by using AI tools in the L2 classroom. In F. Pan (Ed.), *AI in Language Teaching, Learning, and Assessment* (pp. 61–77). Springer. <https://doi.org/10.4018/979-8-3693-0872-1.ch004>
- Barrot, J. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17(2), 102–118. <https://doi.org/10.1016/j.jslw.2007.11.004>
- Bitchener, J. (2012). A reflection on the 'language learning potential' of written CF. *Journal of Second Language Writing*, 21, 348–363. <https://doi.org/10.1016/j.jslw.2012.09.006>
- Bitchener, J., & Ferris, D. R. (2012). *Written Corrective Feedback in Second Language Acquisition and Writing*. Routledge. <https://doi.org/10.4324/9780203832400>
- Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research*, 12(3), 409–431. <https://doi.org/10.1177/1362168808089924>
- Bitchener, J., & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, 19(4), 207–217. <https://doi.org/10.1016/j.jslw.2010.10.002>
- Bitchener, J., & Storch, N. (2016). *Written Corrective Feedback for L2 Development*. Multilingual Matters. <https://doi.org/10.21832/9781783095056>
- Bonilla, M. (2021). An updated typology of written corrective feedback: Resolving terminology issues. *Revista Educación*, 45(2), 1–25. <https://doi.org/10.15517/revedu.v45i1.43289>
- Bonilla, M., Van Steendam, E., Speelman, D., & Buyse, K. (2018). The differential effects of comprehensive feedback forms in the second language writing class: Comprehensive feedback in the L2 writing class. *Language Learning*, 68(3), 813–850. <https://doi.org/10.1111/lang.12295>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(43). <https://doi.org/10.1186/s41239-023-00411-8>
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12, 267–296. [https://doi.org/10.1016/S1060-3743\(03\)00038-9](https://doi.org/10.1016/S1060-3743(03)00038-9)
- Cheng, X., & Zhang, L. J. (2022). Teachers helping EFL students improve their writing through written feedback: The case of native and non-native English-speaking teachers' beliefs. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.804313>
- Dodigovic, M. (2005). *Artificial Intelligence in Second Language Learning: Raising Error Awareness*. Multilingual Matters. <https://doi.org/10.21832/9781853598319>
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language Two*. Oxford University Press.
- Ellis, R. (2009a). A typology of written corrective feedback types. *ELT Journal*, 63(2), 97–107. <https://doi.org/10.1093/elt/ccn023>

- Ellis, R. (2009b). Corrective feedback and teacher development. *L2 Journal*, 1, 3–18. <https://doi.org/10.5070/l2.v1i1.9054>
- Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, 36, 353–371. <https://doi.org/10.1016/j.system.2008.02.001>
- 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*, 20(57). <https://doi.org/10.1186/s41239-023-00425-2>
- Fan, N. (2023). Exploring the effects of automated written corrective feedback on EFL students' writing quality: A mixed-methods study. *SAGE Open*, 13(2). <https://doi.org/10.1177/21582440231181296>
- Ferris, D. R. (2003). *Response to Student Writing: Implications for Second Language Students*. Routledge. <https://doi.org/10.4324/9781410607201>
- Ferris, D. R., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Assessing Writing*, 22, 307–329.
- Fu, Q., Zou, D., Xie, H., & Cheng, G. (2022). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning*, 35(7), 1–43. <https://doi.org/10.1080/09588221.2022.2033787>
- Gardner, J., O'Leary, M., & Yuan, L. (2020). Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo?' *Journal of Computer Assisted Learning*, 37, 1207–1216. <https://doi.org/10.1111/jcal.12577>
- Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal*, 115(4), 523–547. <https://doi.org/10.1086/681947>
- Guo, K., & Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-023-12146-0>
- Hanaoka, O., & Izumi, S. (2012). Noticing and uptake: Addressing pre-articulated covert problems in L2 writing. *Journal of Second Language Writing*, 21, 332–347. <https://doi.org/10.1016/j.jslw.2012.09.008>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112. <https://doi.org/10.3102/003465430298>
- Hong, W. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation (JETI)*, 5(1). <https://doi.org/10.61414/jeti.v5i1.103>
- Hyland, K., & Hyland, F. (2006). Interpersonal aspects of response: Constructing and interpreting teacher written feedback. In K. H. F. Hyland (Ed.), *Feedback in ESL Writing: Contexts and Issues* (pp. 206–224). Cambridge University Press.
- Jiamiao, H., & Li, S. (2023). Opportunities and challenges in the application of ChatGPT in foreign language teaching. *International Journal of Education and Social Science Research (IJESSR)*, 6(4), 75–89. https://ijessr.com/uploads2023/ijessr_06_805.pdf
- Jiang, L., Yu, S., & Wang, C. (2020). Second language writing instructors' feedback practice in response to automated writing evaluation: A sociocultural perspective. *System*, 93, 102302. <https://doi.org/10.1016/j.system.2020.102302>
- Kim, S., Shim, J., & Shim, J. (2023). A study on the utilization of OpenAI ChatGPT as a second language learning tool. *Journal of Multimedia Information System*, 10(1), 79–88. <https://doi.org/10.33851/JMIS.2023.10.1.79>

- Koraishi, O. (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *Language Education and Technology*, 3(1), 55–72.
- Kushmar, V., Vornachev, O., Korobova, O., & Kaida, O. (2022). Artificial Intelligence in language learning: What are we afraid of? *Arab World English Journal (AWEJ) Special Issue on CALL*, 8, 262–273. <https://doi.org/10.24093/awej/call8.18>
- La Russa, F. (2017). Treating errors in learners' writing: Techniques and processing of corrective feedback. In E. Piechurska-Kuciel, E. Szymańska-Czaplak, & M. Szyska (Eds.), *At the Crossroads: Challenges of Foreign Language Learning* (pp. 3–17). Springer. https://doi.org/10.1007/978-3-319-55155-5_1
- Lee, I. (2017). *Classroom Assessment and Feedback in L2 School Contexts*. Springer. <https://doi.org/10.1007/978-981-10-3924-9>
- Li, S., & Roshan, S. (2019). The associations between working memory and the effects of four different types of written corrective feedback. *Journal of Second Language Writing*, 45, 1–15. <https://doi.org/10.1016/j.jslw.2019.03.003>
- Li, S., & Vuono, A. (2019). Twenty-five years of research on oral and written corrective feedback in System. *System*, 84, 93–109. <https://doi.org/10.1016/j.system.2019.05.006>
- Lightbound, P. (2005). *An Analysis of Interlanguage Errors in Synchronous / Asynchronous Intercultural Communication Exchanges* [Universitat de València]. <http://roderic.uv.es/bitstream/handle/10550/15306/penny.pdf?sequence=1&isAllowed=y>
- Liu, G., & Ma, C. (2023). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*. <https://doi.org/10.1080/17501229.2023.2240316>
- Liu, Q., & Brown, D. (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing*, 30, 66–81. <https://doi.org/10.1016/j.jslw.2015.08.011>
- Liu, W. (2024). The integration of ChatGPT in language education: Exploring contexts, assessing impacts, and unpacking implications. In F. Pan (Ed.), *AI in Language Teaching, Learning, and Assessment* (pp. 1–21). IGI Global. <https://doi.org/10.4018/979-8-3693-0872-1.ch001>
- Luo, Y., & Liu, Y. (2017). Comparison between peer feedback and automated feedback in college English writing: A case study. *Open Journal of Modern Linguistics*, 7, 197–215. <https://doi.org/10.4236/ojml.2017.74015>
- Ma, X., & Huo, Y. (2023). Are users willing to embrace ChatGPT? Exploring the factors on the acceptance of chatbots from the perspective of AIDUA framework. *Technology in Society*, 75, 102362. <https://doi.org/10.1016/j.techsoc.2023.102362>
- Mao, Z., & Lee, I. (2020). Feedback scope in written corrective feedback: Analysis of empirical research in L2 contexts. *Assessing Writing*, 45, 1–14. <https://doi.org/10.1016/j.asw.2020.100469>
- McMartin-Miller, C. (2014). How much feedback is enough? Instructor practices and student attitudes toward error treatment in second language writing. *Assessing Writing*, 19, 24–35. <https://doi.org/10.1016/j.asw.2013.11.003>
- Meyer, J., Urbanowicz, R., Martin, P., O'Connor, K., Li, R., Peng, P., Bright, T., Tatonetti, N., Jae Won, K., Gonzalez-Hernandez, G., & Moore, J. (2023). ChatGPT and large language models in academia: Opportunities and challenges. *BioData Mining*, 16(20). <https://doi.org/10.1186/s13040-023-00339-9>

- Mohamed, A. M. (2023). Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: Perceptions of EFL faculty members. *Education and Information Technologies*, 29, 3195–3217. <https://doi.org/10.1007/s10639-023-11917-z>
- Nguyen, T. (2023). EFL teachers' perspectives toward the use of ChatGPT in writing classes: A case study at Van Lang University. *International Journal of Language Instruction*, 2(3), 1–47. <https://doi.org/10.54855/ijli.23231>
- Pienemann, M. (1998). *Language Processing and Second Language Development: Processability Theory*. John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.15>
- Ray, P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121–154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, 20, 83–93. <https://doi.org/10.2307/3586390>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 1–22. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11, 11–26.
- Semke, H. (1984). The effects of the red pen. *Foreign Language Annals*, 17, 195–202. <https://doi.org/10.1111/j.1944-9720.1984.tb01727.x>
- Shaikh, S., Yayilgan, S. Y., Klimova, B., & Pikhart, M. (2023). Assessing the usability of ChatGPT for formal English Language Learning. *European Journal of Investigation in Health, Psychology and Education*, 13(9), 1937–1960. <https://doi.org/10.3390/ejihpe13090140>
- Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41, 255–283. <https://doi.org/10.1002/j.1545-7249.2007.tb00059.x>
- Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford University Press.
- Steiss, J., Tate, T. P., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., & Warschauer, M. (2023). Comparing the quality of human and ChatGPT feedback on students' writing. <https://doi.org/10.35542/osf.io/ty3em>
- Stevenson, M., & Phakiti, A. (2024). 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>
- Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study. *International Journal of Educational Technology in Higher Education*, 20(35). <https://doi.org/10.1186/s41239-023-00404-7>
- Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback of writing. *Studies in Second Language Acquisition*, 32, 303–334. <https://doi.org/10.1017/S0272263109990532>
- Tate, T. P., Doroudi, S., Ritchie, D., Xu, Y., & uci, m. w. (2023, January 10). *Educational research and AI-generated writing: Confronting the coming tsunami*. <https://doi.org/10.35542/osf.io/4mec3>
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46, 327–369. <https://doi.org/10.1111/j.1467-1770.1996.tb01238.x>

- 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>
- Van Beuningen, C. (2011). *The Effectiveness of Comprehensive Corrective Feedback in Second Language Writing* [Unpublished doctoral dissertation]. Universiteit van Amsterdam.
- Wang, L., Chen, X., Wang, C., Xu, L., Shadiev, R., & Li, Y. (2024). ChatGPT's capabilities in providing feedback on undergraduate students' argumentation: A case study. *Thinking Skills and Creativity*, 51, 101440. <https://doi.org/10.1016/j.tsc.2023.101440>
- Wang, Z., & Han, F. (2022). The effects of teacher feedback and automated feedback on cognitive and psychological aspects of foreign language writing: A mixed-methods research. *Frontiers in Psychology*, 13, 1–12. <https://doi.org/10.3389/fpsyg.2022.909802>
- Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a Second or Foreign Language. *Journal of Second Language Writing*, 62, 101071. <https://doi.org/10.1016/j.jslw.2023.101071>
- Westmacott, A. (2017). Direct vs. indirect written corrective feedback: Student perceptions. *Íkala Revista de Lenguaje y Cultura*, 22(2), 17–32. <https://doi.org/10.17533/udea.ikala.v22n01a02>
- Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, 8(3). <https://doi.org/10.3390/languages8030212>
- Yeo, M. A. (2023). Academic integrity in the age of artificial intelligence (AI) authoring apps. *TESOL Journal*, 14(3). <https://doi.org/10.1002/tesj.716>
- Yoon, S.-Y., Miszoglád, E., & Pierce, L. (2023). *Evaluation of ChatGPT Feedback on ELL Writers' Coherence and Cohesion*. <https://arxiv.org/pdf/2310.06505.pdf>
- Zhai, X. (2022). *ChatGPT User Experience: Implications for Education*. <https://doi.org/10.2139/ssrn.4312418>
- Zhang, L. J., & Cheng, X. (2021). Examining the effects of comprehensive written corrective feedback on L2 EAP students' linguistic performance: A mixed-methods study. *Journal of English for Academic Purposes*, 54, 101043. <https://doi.org/10.1016/j.jeap.2021.101043>
- Zhang, T., Chen, X., Hu, J., & Ketwan, P. (2022). EFL students' preferences for written corrective feedback: Do error types, language proficiency, and foreign language enjoyment matter? *Frontiers in Psychology*, 13, 1–25. <https://doi.org/10.3389/fpsyg.2021.660564>