

Students' Perceptions and Usage Patterns of ChatGPT as an Automated Writing Evaluation (AWE) Tool

Linqian Ding^{1*}, Di Zou², Lucas Kohnke³

^{1,3} The Education University of Hong Kong

² The Hong Kong Polytechnic University

* 15080791126@163.com

Abstract: The emergence of Generative AI (GAI) has significantly influenced writing evaluation, with recent literature primarily focusing on comparing GAI feedback or scores with those provided by humans. However, fewer studies have explored students' perspectives, which play a crucial role in the effectiveness of feedback. To address this gap, this study investigates how EFL students interact with ChatGPT as an Automated Writing Evaluation (AWE) tool and their perceptions of its usefulness in improving academic writing. The study involved 47 graduate students at a university in Hong Kong, and data were collected through a questionnaire survey to examine their usage patterns and perceptions of ChatGPT feedback. The findings reveal that students primarily used ChatGPT during the revision stage for grammar and vocabulary enhancement, while its use for higher-order writing tasks was limited. Moreover, students generally viewed ChatGPT as trustworthy, useful, and easy to use, though there were concerns about information security and ethical implications. Despite these challenges, students expressed a willingness to continue using ChatGPT, suggesting its potential as a supplementary tool for academic writing instruction. This study highlights the need to refine GAI tools for better writing support and ethical use.

Keywords: Generative AI, ChatGPT, Automated Writing Evaluation (AWE)

1. Introduction

The advent of technology has transformed language education significantly, particularly in the field of writing instruction (Wen & Walters, 2022). One notable advancement is the implementation of automated writing evaluation (AWE) tools, which have evolved greatly since their inception (Ding & Zou, 2024). Originally, AWE tools assessed and scored texts, but technological advancements, such as natural language processing and artificial intelligence (AI) have enhanced their ability to provide more accurate, customized feedback (Bai & Hu, 2017). With AWE systems gaining popularity in writing instruction, research has shown that they enhance writing performance across various dimensions and users generally maintaining a positive attitude toward these technologies (Barrot, 2021). However, AWE systems have received criticism. Some argue that AWE feedback can be inaccurate, generic, and confusing, and it lacks deep analysis of content and organization (Han & Sari, 2022). With the emergence of generative AI technologies, such as ChatGPT, it is believed they have the potential overcome these limitations (Ding & Zou, 2024).

As an AI-powered chatbot, ChatGPT can mimic human conversation and may revolutionize AWE technologies by providing more effective evaluations and feedback (Mizumoto & Eguchi, 2023). There have been explorations of using ChatGPT as an AWE tool, but there is a lack of comprehensive studies on students' usage patterns and perceptions. This study aims to fill this gap by investigating the student usage patterns and perceptions towards ChatGPT feedback. The study is guided by the following research questions:

RQ1: What usage patterns emerge when students use ChatGPT for writing evaluation?

RQ2: What are students' perceptions of ChatGPT feedback on writing evaluation?

2. Methodology

The study involved 47 graduate students majoring in English Education at a university in Hong Kong. The majority of participants were female ($n=43$), aged between 18-34 years ($n=45$). All participants were native Chinese speakers with more than 10 years of English learning experience in China from primary to undergraduate education. They were enrolled in the same English Writing course which ran for 16 weeks with one 120-minute session per week.

To address the two research questions, a questionnaire survey was designed to collect data on students' usage patterns and perceptions. The questionnaire included two sections: (1) Usage of ChatGPT as an AWE tool and (2) Perceptions of using ChatGPT as an AWE tool. The first section comprised six multiple-choice questions and the second section consisted of 27 closed-ended items measured on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree) and three multiple-choice questions. The items were adapted from established scales and prior research on AWE feedback (e.g., Davis, 1989; Huang & Renandya, 2020) to probe students' (a) perceived trust in ChatGPT's feedback, (b) perceived usefulness of the feedback, (c) perceived ease of use, (d) perceived value of the feedback for revision and English writing performance, (e) AWE system anxiety, (f) behavioral intention to use ChatGPT as an AWE tool, and (g) beliefs about the ethicality of such usage. The Cronbach's α coefficients for each were 0.87, 0.94, 0.89, 0.90, 0.82, and 0.84, respectively, indicating good reliability.

This quasi-experimental study was integrated into the course in week 5 with a 60-minute training session on accessing ChatGPT and using it for writing feedback. The session covered uploading files, inputting instructions, and refining feedback through follow-up questions. In week 6, students completed a 60-minute in-class writing task based on an IELTS writing topic. After submitting their first draft, students were instructed to use ChatGPT to receive feedback on their assignment. To facilitate this process, they were provided with sample prompts designed to elicit specific and constructive feedback. Examples were presented below:

"Assume the role of a language teacher. I am writing an IELTS essay. Please provide feedback on my vocabulary/ grammar/ structure"; "Explain any mistakes in my writing"; "Provide revised sentence examples/ synonyms".

The entire interaction process was conducted over 60 minutes during class. Following this, they completed the questionnaire to report how they used the tool and share their perceptions and experiences with it.

3.Results

3.1. Usage pattern of ChatGPT as AWE tool

The results indicate that a significant portion of the students in the class frequently used ChatGPT to provide writing feedback in their coursework. A majority of the students ($n=34$) reported using ChatGPT "sometimes" for writing feedback. Conversely, 19.1% of the participants seldom used these tools, with only two students reporting that they "never" used them. The reasons cited by the two students were concerns about potential penalties for using such technology. As for the timing of ChatGPT usage, most students ($n=40$) used it after completing draft but before submission. Ten students used it during the planning stage, while six used it during drafting. Notably, none used ChatGPT after receiving grades and instructor feedback.

Moreover, students predominantly used ChatGPT to identify and correct grammatical errors or vocabulary mistakes (77.8%), improve vocabulary or sentence structure (75.6%), and provide feedback on the structure, organization, and logic of their writing (64.4%). Over half of the respondents also used ChatGPT to refine entire paragraphs or complete texts. Less frequently, students used ChatGPT to score their writing, assess originality, prevent plagiarism, or receive suggestions on writing style. One student mentioned using ChatGPT to suggest an outline for their assignment.

Regarding the impact of ChatGPT's feedback on the revision process, most students reported making changes based on the feedback received. Specifically, 35 students made significant or moderate changes to their writing, while eight students made minor changes. In contrast, two students reported that the feedback did not influence their revision process, either because they acknowledged the feedback without implementing changes or they found it unhelpful. In terms of the language used to interact with ChatGPT, the majority of participants ($n=29$) communicated in English,

while 35.6% used both Chinese and English. Notably, despite their mother tongue being Chinese, none of the participants exclusively used Chinese to interact with ChatGPT.

3.2. Perceptions of ChatGPT as AWE tool

Our findings indicate that students generally perceive ChatGPT as a reliable and useful AWE tool, though concerns remain regarding information security and ethical considerations (see Table 1). Regarding perceived trust, students expressed moderate confidence in ChatGPT's reliability ($M = 3.51$, $SD = 0.67$). However, concerns about information security were evident, with some students remaining cautious about data privacy and potential security risks. Both perceived usefulness ($M = 3.81$, $SD = 0.79$) and perceived ease of use ($M = 3.70$, $SD = 0.72$) were rated positively, which indicates that students found ChatGPT both beneficial for writing improvement and easy to use.

Regarding students' perceived value of the feedback, they found it useful not only for revising their current essays ($M = 3.76$, $SD = 0.71$) but also for improving their long-term English writing skills ($M = 3.89$, $SD = 0.73$). Specifically, most students found ChatGPT feedback particularly helpful for grammar improvement and vocabulary expansion. They also acknowledged its effectiveness in identifying writing problems in future writing practice and enhancing overall writing proficiency. However, their perceived value of the feedback on deeper-level writing aspects, such as organization and content, was relatively lower. These findings align with prior research on AWE tools, which suggests that students value AI-generated feedback primarily for surface-level corrections rather than higher-order writing improvements (Huang & Renandya, 2020). When asked about their intentions to continue using ChatGPT, students showed a moderately positive response ($M = 3.50$, $SD = 0.72$). This indicates that while students acknowledge and value the benefits of ChatGPT, they still have concerns about its limitations.

Table 1. Students' Perceptions of ChatGPT as AWE tool

	Mean	Standard Deviation
Perceived trust	3.51	0.67
Perceived usefulness	3.81	0.79
Perceived ease of use	3.70	0.72
Perceived value for revision	3.76	0.71
Perceived value for English writing performance	3.89	0.73
Anxiety	3.01	0.63
Behavioral intention	3.50	0.72

Regarding students concerns, they exhibited moderate levels of anxiety when using ChatGPT for writing tasks ($M = 3.01$, $SD = 0.63$). Many students expressed concerns about the possibility of making irreversible mistakes, fearing that errors in their writing might not be accurately identified or appropriately corrected by the tool. Furthermore, some students felt intimidated by the evaluation process itself, as they were uncertain about how ChatGPT generates feedback and whether its suggestions truly reflect their writing proficiency. This underscores the need for explicit training on how to effectively interact with ChatGPT to minimize uncertainty and enhance user confidence.

As for ethical perceptions, students expressed diverse opinions. While only one student considered its use inherently unethical, nearly half of the students regarded it as ethically acceptable under any circumstances ($n = 20$). Beyond this, many students perceived it as cheating in certain situations, 19 students believed it was unethical when explicitly prohibited, while 14 students felt it was unethical when students relied too heavily on its output without making any revisions. Therefore, it is essential for institutions to provide clear guidelines to help them understand what is considered appropriate and responsible use.

4. Discussion and conclusion

This study examined how EFL students use ChatGPT for writing evaluation and their perceptions towards the tool. The findings suggest that students primarily use ChatGPT as a revision tool rather than for initial drafting or brainstorming. Most students rely on it for grammar and vocabulary enhancement, using it after completing their first

drafts but before final submission. However, fewer students engage with ChatGPT for higher-order writing tasks, such as improving argumentation or content organization. This indicates that while students find ChatGPT useful for surface-level corrections, they may perceive its feedback as less effective for deeper revisions, limiting its role in refining critical thinking and coherence in writing.

Students' perceptions of ChatGPT were generally positive, particularly regarding its usefulness and ease of use. However, concerns were raised about information security and ethical considerations. While many students viewed ChatGPT as a helpful writing aid, others expressed uncertainty about its appropriateness in academic settings, especially when its use was restricted or when AI-generated content was integrated without proper modification. Additionally, some students experienced moderate anxiety of using ChatGPT for writing evaluation, particularly regarding the reliability of ChatGPT's feedback and the potential risks of over-reliance on AI-generated suggestions.

In conclusion, this study underscores the need for both pedagogical and technological improvements in the integration of GAI tools in writing evaluation. Instructors should provide explicit training to help students critically engage with ChatGPT feedback, ensuring they use it as a support tool rather than a substitute for independent writing. Institutions must establish clear ethical guidelines to address concerns about academic integrity and responsible GAI usage. From a technological perspective, developers should enhance ChatGPT ability to provide more nuanced feedback on high-order aspects, such as argumentation, structure, and coherence.

Reference

- Bai, L., & Hu, G. (2017). In the face of fallible AWE feedback: How do students respond? *Educational Psychology*, 37(1), 67–81. <https://doi.org/10.1080/01443410.2016.1223275>
- Barrot, J. S. (2021). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning*, 1–24. <https://doi.org/10.1080/09588221.2021.1936071>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.
- Ding, L., & Zou, D. (2024). Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence. *Education and Information Technologies*, 1-53. <https://doi-org/10.1007/s10639-023-12402-3>
- Han, T., & Sari, E. (2022). An investigation on the use of automated feedback in Turkish EFL students' writing classes. *Computer Assisted Language Learning*, 1–24. <https://doi-org/10.1080/09588221.2022.2067179>
- Huang, S., & Renandya, W. A. (2020). Exploring the integration of automated feedback among lower-proficiency EFL learners. *Innovation in Language Learning and Teaching*, 14(1), 15–26. <https://doi.org/10.1080/17501229.2018.1471083>
- 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), 100050. <https://doi-org/10.1016/j.rmal.2023.100050>