



Artificial Intelligence in Language Education, 103356 (2026)
<https://doi.org/10.29140/aile.2026.103356>



Enhancing L2 writing with AI: Effects of ChatGPT feedback and the role of student engagement

LINQIAN DING^a 

DI ZOU^b 

LUCAS KOHNKE^a 

^aThe Education University of Hong Kong, Hong Kong

^bThe Hong Kong Polytechnic University, Hong Kong

Abstract

The rapid development of Generative AI (GAI) tools offers new opportunities for providing immediate and personalized feedback in writing instruction. While previous studies have explored its potential as an automated written corrective feedback (AWCF) tool, little is known about how effective its feedback is across different aspects of L2 writing, and how such engagement influences writing outcomes. The present study examined the effects of ChatGPT feedback on undergraduate students' L2 writing performance, including lexis, grammar, content, and organization, and investigated the role of learner engagement, including behavioral, cognitive, and emotional engagement, in influencing these effects. A total of 273 students with comparable English proficiency were divided into two groups: one received teacher feedback while the other obtained ChatGPT feedback. Results showed that ChatGPT feedback led to significant improvements in grammar and lexis but had limited impact on content and organization. In contrast, the teacher group did not show significant improvement in any writing dimensions. Analysis of student engagement revealed that all three dimensions were positively associated with writing improvement. Among them, behavioral engagement was the strongest and most direct predictor of the effectiveness of ChatGPT feedback. Practical implications for using ChatGPT in writing instruction and directions for future research are discussed.

Keywords: ChatGPT; written corrective feedback; learner engagement

Introduction

Written corrective feedback has been widely considered as an important role in writing instruction (Hyland & Hyland, 2006). With the continuing development of technology, using computers to provide feedback on writing has been a growing area of research (Yeung, 2025).



Derived from advances in natural language processing and educational technology, Automated Written Corrective Feedback (AWCF) has been increasingly used in second language (L2) writing instruction in recent years (Rahimi et al., 2025). This type of feedback is generally delivered through Automated Writing Evaluation (AWE) tools, such as Grammarly and Criterion. These tools allow users to upload their writing online, where it is compared with authentic datasets and predefined linguistic criteria. The tools then apply algorithms to generate immediate scores and AWCF that are consistent with the submitted text (Barrot, 2023; Hockly, 2019). AWCF has been widely integrated in writing instruction (Guo et al., 2022). It not only reduces the workload of human raters but also creates opportunities for students to make timely revisions and consolidate their learning through repeated practice (Wu et al., 2025).

Many studies have highlighted the benefits of AWCF, such as its positive effects on learners' attention to language form, overall writing quality, motivation to write, and writing self-efficacy (e.g., Liu, 2025; Ranalli, 2021; Wilson & Roscoe, 2020). However, AWCF has also faced criticism. It relies on rigid scoring criteria that tend to promote formulaic writing (Stevenson & Phakiti, 2014). It also assesses writing based on predefined skills on a narrow range of essay types, which limits students' development of more complex written expression (Zhang, 2020). In addition, many have expressed concern that AWCF tends to focus too much on surface-level features, such as grammar, spelling, and punctuation, while paying less attention to higher-order aspects of writing, such as idea development, organization, and argumentation (Barrot, 2023; Gao, 2021). As a result, students may improve in mechanical accuracy, but struggle to make progress in their overall writing skills.

The advancements in Generative Artificial Intelligence (GAI) have opened up new possibilities for developing innovative learning approaches (Allison et al., 2025). With tools such as ChatGPT by OpenAI, AWCF may be entering a new phase of evolution (Escalante et al., 2023; Yeung, 2025). ChatGPT can provide feedback or suggestions tailored to individual learner inputs, which offers strong potential to enhance academic performance (Chang & Hwang, 2025). Moreover, unlike previous AWE tools, these GAI tools don't rely on training human input datasets for a specific writing genre (Alsofyani & Barzanji, 2025). It is expected that these tools can deliver timely and personalized feedback that is relevant to students' needs, which allows learners to improve their writing skills (Li et al., 2024). Some research has utilized GAI tools to provide AWCF and explored the effectiveness of GAI feedback, but the findings have been mixed. While some findings suggest that GAI tools can significantly improve writing quality, reduce writing anxiety, and boost learner motivation (Kartika, 2025; Zhang & Wang, 2025), other research points out that GAI feedback might limit students' creativity and lead to lower scores in originality and content accuracy (Chowdhury et al., 2024). This mixed evidence underscores the need for further investigation into the effectiveness of GAI feedback, particularly in the context of L2 writing (Zhang et al., 2025).

To understand the effectiveness of GAI feedback, it is necessary to understand students' engagement with it. Learner engagement is a multidimensional construct, including behavioral, cognitive, and emotional engagement (Zhang & Hyland, 2018). All three components interact with one another and collectively play a crucial role in writing development (Hiver et al., 2021). Many studies have shown that the same feedback can have very different outcomes depending on how learners interact with it (Jiang & Yu, 2025; Koltovskaia & Mahapatra, 2022). Specifically, students who actively process and apply AWCF tend to make greater improvements than those who passively receive or ignore the feedback (Cheng & Zhang, 2024). Therefore, to understand the effectiveness of AI-generated AWCF, it is necessary to understand how learners engage with it. Nevertheless, despite increasing interest in using GAI tools for writing support, limited research has explored how learners engage with AI-generated AWCF and how this

engagement influences writing performance. Most existing studies focus on either teacher or peer feedback, and few examine the role of learner engagement in the context of GAI-supported learning. This represents a key gap in the literature. To address this gap, the present study sets out to achieve two main objectives: (1) To examine the effectiveness of GAI feedback on writing performance, including different aspects of writing such as content, organization, use of grammar, and use of lexis; and (2) To explore the relationship between student engagement and the effectiveness of GAI feedback. By addressing these objectives, we can understand how GAI feedback can be effectively integrated into L2 writing instruction to enhance writing development.

Literature Review

Automated Written Corrective Feedback and Writing Performance

Many recent review studies have reported that AWCF is effective in improving L2 learners' writing performance (Ding & Zou, 2024; Liu, 2025). Specifically, Guo et al. (2022) found that learners who received AWCF correctly revised about 85% of the writing errors pointed out by Grammarly, thus significantly reduced their error scores after revision. Similarly, Barrot (2023) and Ware (2023) both reported that learners who engaged with AWCF outperformed their counterparts. They attributed this improvement to the tool's ability to promote noticing, offer detailed explanation, and foster self-directed learning. In addition, the immediate feedback allows students to revise their compositions promptly which potentially led to the more significant improvements in writing performance as well (Tan et al., 2023). On the other hand, some studies claimed that AWCF did not lead to significant improvements in writing performance (e.g., Fan, 2023; Huang & Renandya, 2020). They attributed the unsatisfied results to learners' lower English proficiency as well as lower feedback literacy, which limited their ability to understand and apply feedback effectively (Fan, 2023). In a more detailed analysis, Rahimi et al. (2025) found that AWCF had a positive impact on several dimensions of writing, including overall performance, task achievement, and grammatical range and accuracy. However, they reported no significant differences between the AWCF and control groups in terms of coherence, cohesion, and lexical resource. This result was in line with Barrot (2023) and Gao's (2021) findings, which also reported that AWCF focuses more on surface-level features such as grammar, punctuation, and spelling rather than higher-order concerns. This is because traditional AWE tools relied on fixed rubrics and predefined datasets, which limited their ability to deeply understand students' context, interpret complex emotions, or generate original ideas (Teng, 2024).

With the emergence and development of GAI tools, research has begun to explore their potential to provide AWCF and even replace traditional AWE systems in writing evaluation. Unlike traditional AWE tools, GAI is based on large language models (LLMs), which are able to comprehend complex syntactic and semantic structures (Wu et al., 2025). This allows them to generate more accurate and contextually relevant feedback. Furthermore, GAI can tailor its output to match learners' different needs and proficiency level, and then generate customized feedback to support personalized learning (Jeon, 2023). Yancey et al. (2023) evaluated the scoring performance of GPT-3.5 and GPT-4 against human raters and found that GPT-4's scores closely aligned with those of human evaluators, especially when appropriate examples were provided. Özdere (2025) compared two GAI tools, ChatGPT and You.com, to examine their differences in grading consistency between students' first and final drafts. The study found that You.com generally assigned higher scores to final drafts than ChatGPT. In addition, both GAI models demonstrated reliable scoring patterns. Nevertheless, Steiss et al. (2024) observed contrasting results. They found that human feedback outperformed ChatGPT in

terms of clarity, accuracy, relevance, and tone, and they attributed it to the model's inability to truly comprehend text in a human-like manner.

Several studies examined the effectiveness of AI-generated AWCF in enhancing writing performance. The findings have also been mixed. Chan et al. (2024), as well as Pariyanto and Tungka (2025), found that students who received ChatGPT feedback on their essays significantly improved their writing quality compared to those who did not. Nevertheless, both studies focused on one single writing task and lacked any examination of long-term effects on writing performance. Alsofyani and Barzanji (2025) conducted a six-week study to examine the effects of ChatGPT feedback on EFL learners' writing skills. They also found that ChatGPT feedback led to improvements in writing, and learners reported positive perceptions of ChatGPT's usefulness for academic writing. However, Zhang et al. (2025) found that while GAI feedback alone improved grammar and sentence variety, its impact on higher-order writing skills was limited. Therefore, in order to better understand the benefits of GAI on writing performance, it is necessary to investigate its impact on specific components of writing, and to conduct long-term examinations to assess its sustained effectiveness over time.

Learner Engagement with Automated Written Corrective Feedback and Writing Performance

Student engagement with AWCF is a critical but overlooked area in L2 writing (Mao & Lee, 2024). This concept originated from the broader notion of learner engagement, which is defined as the extent of learners' active involvement in the learning process (Fredricks et al., 2004). It is commonly known as a multidimensional framework that includes behavioral, cognitive, and emotional dimensions (Ellis, 2010; Han & Hyland, 2015). Based on this, Zhang and Hyland (2018) extended its application to AWCF and built a model of students' engagement with AWCF. In this model, cognitive engagement refers to the mental effort learners invest in processing feedback, such as noticing errors, understanding feedback, and the use of cognitive and metacognitive strategies (Koltovskaia, 2020). Emotional engagement refers to learners' feelings, attitudes, and motivation toward the feedback (Zhang & Hyland, 2018), while behavioral engagement refers to students' behavioral reaction to feedback, such as making revisions or the time invested in revising (Zhang & Hyland, 2023).

This framework is now widely used in recent studies on learner engagement with AWCF. For example, Koltovskaia (2020) conducted a case study to examine how two students engaged with AWCF provided by Grammarly. The two students showed different levels of engagement. One student showed higher cognitive engagement by questioning the feedback, while the other showed lower cognitive engagement and depended heavily on the feedback without thinking critically. Afifi (2023) conducted a case study to investigate students' three dimensions of engagement with teacher feedback and AWCF. The results showed that learners who received teacher feedback exhibited stronger engagement across all dimensions compared to those using the AWCF. Some recent studies have investigated how students engage with GAI-provided AWCF. Koltovskaia et al. (2024) investigated graduate students' engagement with AWCF generated by ChatGPT. They found that learners demonstrated strong cognitive engagement and emotional engagement. Similarly, Yeung's (2025) case study also reported students generally engaged positively with GAI-provided AWCF. Students with higher proficiency showed deeper cognitive engagement, whereas others tended to focus on surface-level revisions.

Although these studies acknowledge the importance of engagement in writing development, there has been limited research on the relationship between engagement with AWCF and writing performance. Jin et al. (2022) explored the impact of student engagement in peer feedback on writing performance and reported that cognitive, emotional, and behavioral engagement were all significantly associated with improved writing outcomes among learners. Li and Yuan (2024)

examined how different dimensions of engagement related to English writing achievement. They claimed that emotional and behavioral engagement had a significant positive impact on writing scores, while cognitive engagement did not show a significant influence. Fleckenstein et al. (2024) explored behavioral engagement as a mediating factor in the effectiveness of AWCF. Their findings indicated that students who spent more time engaging with feedback improved more in their writing performance. Huang and Mizumoto (2025) investigated the mediating role of GAI in the relationship between learners' motivation and engagement in EFL writing contexts. They found that GAI usage significantly influenced all three dimensions of engagement and confirmed that GAI mediated the relationship between motivation and engagement both directly and indirectly. However, little is known about how engagement with GAI feedback influences writing outcomes. It is necessary to conduct more comprehensive studies that adopt various quantitative and qualitative methods to examine learners' engagement of GAI tools in writing (Koltovskaia et al., 2024; Xue, 2024). To address these research gaps, the present study not only aims to investigate the long-term effects of ChatGPT-generated feedback on writing performance but also explores the role of learner engagement in mediating the relationship between feedback and writing outcomes. We used ChatGPT as a representative GAI tool to support writing instruction for undergraduate L2 learners. To better understand its effectiveness, we compared the impact of ChatGPT feedback with that of traditional teacher feedback. The study is guided by the following research questions:

- RQ1. To what extent does ChatGPT feedback improve L2 students' writing performance across different dimensions (i.e., content, organization, use of grammar, and use of lexis)?
- RQ2. How does ChatGPT feedback compare with traditional teacher feedback in influencing students' writing performance?
- RQ3. What is the relationship between students' engagement with ChatGPT feedback—across behavioral, cognitive, and emotional dimensions—and their improvement in writing performance?

Methodology

Participants and Context

The study involved 273 undergraduate students from a university in Hong Kong. Most of them were first-year students ($n = 271$) and two were second-year students. The participants ranged in age from 18 to 23, which is typical for undergraduate students. The gender distribution was relatively balanced, with 131 male and 140 female students. Among them, 147 were local students from Hong Kong, 122 were students from mainland China, and 4 were international students from other countries. They majored in different subjects, but were all required to enroll in a compulsory University English Course. The course was designed to develop students' academic writing skills and academic oral presentation skills. There were 12 classes arranged to deliver the course at the university, and all first-year students were randomly assigned to one class. The classes used the same syllabus and teaching materials, and met twice a week for 2-hour sessions over 16 weeks. To ensure consistency and reliability of the study, five classes taught by the same teacher were selected for the study. Two classes ($n = 116$) were randomly assigned as the experimental group (EG) and the other three classes ($n = 157$) formed the control group (CG). The gender distribution in the EG was 56 males and 60 females, while the CG consisted of 76 male and 81 female students. All participants were fully informed about the purpose of the study. Their participation was voluntary, and students were informed that

they could withdraw at any time without any academic penalty. Informed consent was obtained from all participants prior to data collection.

During the course, students completed a writing assignment every two weeks. In addition, they also completed two in-class writing tests: the first in Week 4 (pre-test), and the second in Week 15 (post-test). These two tests were designed to assess students’ academic writing progress and served as the primary data for the study. Both were identical across each group in terms of genre, topic, word count, and assessment criteria: in the pre-test, students wrote a 250–300 word argumentative essay in 45 minutes on “Should university education be free for all students?” and in the post-test, they wrote a 350–400 word essay in 60 minutes on “Has artificial intelligence done more harm than good to society?” Moreover, both the pre-test and post-test were handwritten and completed in class without the assistance of digital tools. This ensured that the writing produced during the tests reflected the students’ genuine writing ability.

Procedure

The overview of the study procedure can be seen in Figure 1. In the first week, all participants completed the first writing assessment in class. The completed papers were then collected by the instructor. In Week 2, students continued with regular class activities. The EG also received a 60-minute training session on how to use ChatGPT to obtain feedback on their writing. The

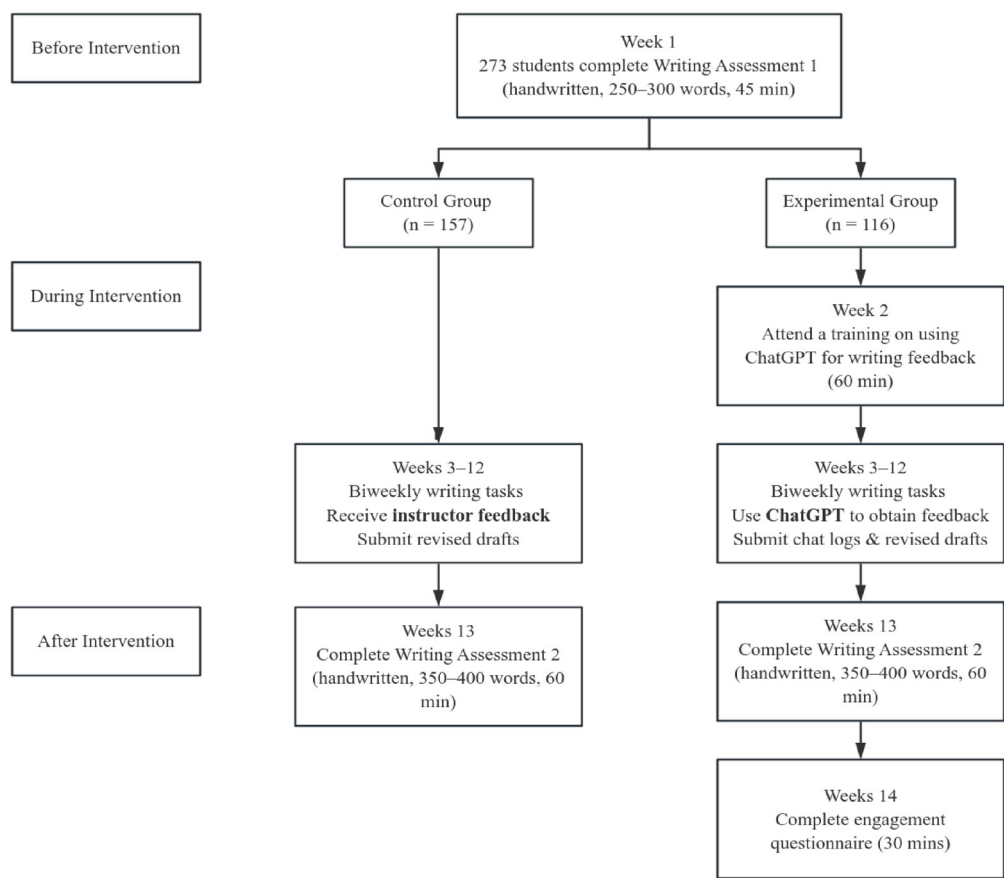


Figure 1. Outline of the experimental procedure

training introduced the process of uploading their drafts to ChatGPT, using effective prompts to request feedback, and interpreting the feedback provided by the tool. Example prompts included: “Can you give me feedback on the coherence of this passage?” and “Are there any grammar mistakes in my writing?” Students were also shown how to apply the suggested revisions critically rather than accepting them without reflection.

Starting from Week 3, all students in both the EG and CG were assigned with a writing task every two weeks. These assignments were completed independently and submitted through an online platform accessible to the instructor. After their submission, CG students received instructor’s feedback on areas such as content development, organization, grammar, and vocabulary use. The teacher also provided general suggestions for improvement and occasionally discussed common issues during class. Feedback was returned within one week, after which students had another week to revise and upload their final drafts. In contrast, EG students were guided to use the prompts introduced during training to obtain feedback from ChatGPT on the same areas of their writing. A sample of student interaction with ChatGPT is provided in Figure 2.

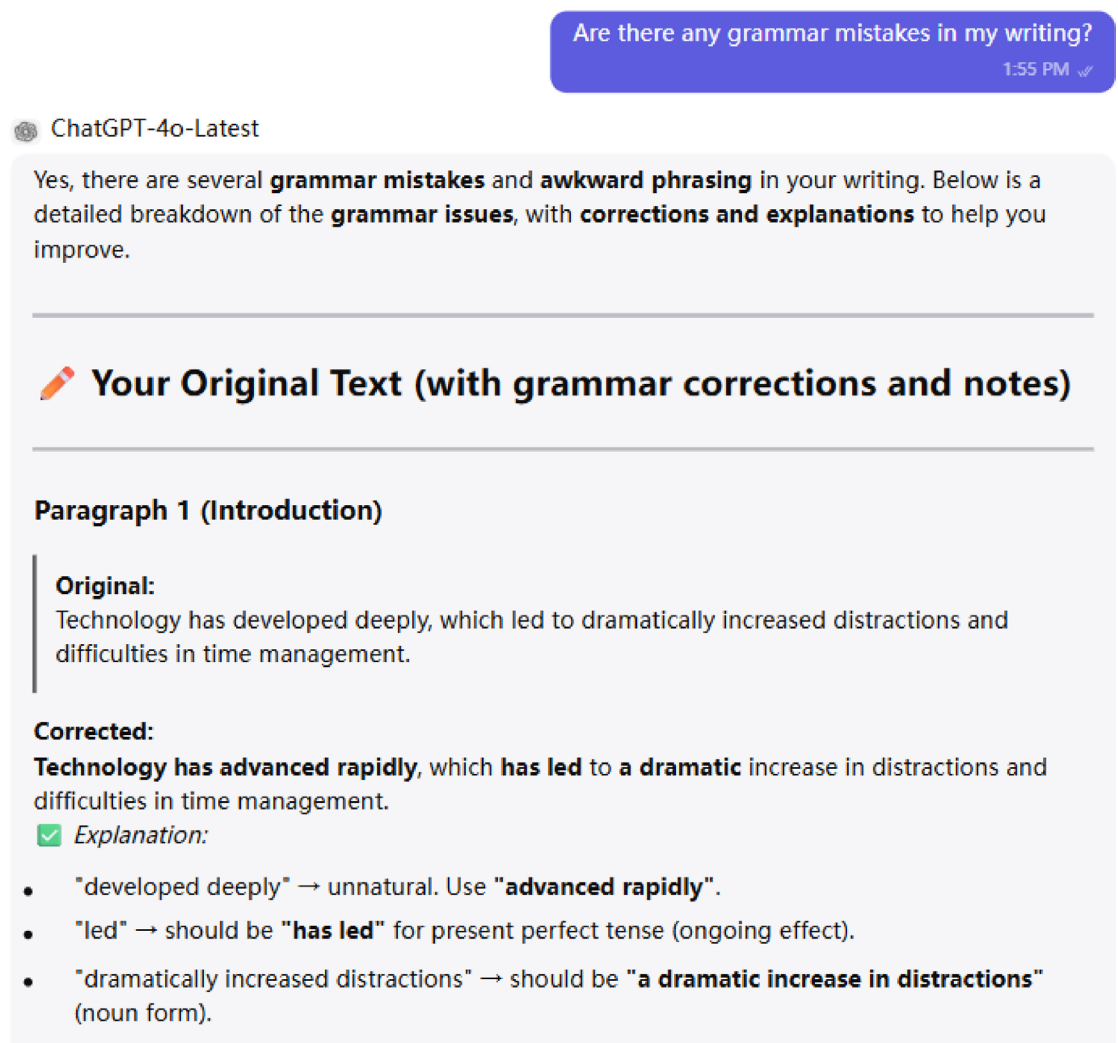


Figure 2. Sample of student interaction with ChatGPT

In addition to submitting their revised drafts online, EG students were also required to upload their chat history with ChatGPT as part of the submission process.

In Week 13, all students took the second writing assessment in class, and the assessment was graded by the same instructor using the same rubrics. In Week 14, students in the EG completed a questionnaire to gather information about their engagement with ChatGPT feedback during the revising process. They were also invited to volunteer for follow-up interviews. A total of 34 students participated in the interviews. Each of them joined an online interview with the researchers for approximately 30 minutes.

Data Collection

Our study employed a mixed-method approach to collect both quantitative data on students' English assessment grades and engagement, and qualitative data on more detailed information about their experiences with ChatGPT feedback. For the quantitative data, we selected students' grades from the Writing Assessment 1 as pre-test scores, and grades from Writing Assessment 2 as post-test scores. Both assessments were evaluated by the same human rater with standardized rubrics. The rubrics covered four aspects of writing: content, organization, use of lexis, and use of grammar. Each aspect was scored out of 100, with the total score being the sum of the four aspects. All assessments were graded by the second researcher, who had more than ten years of English teaching experience.

As for the EG students' engagement, our study employed a questionnaire adapted from Zou et al.'s (2023) study. They used this questionnaire to investigate students' engagement with AWCF. For the purposes of our study, we modified the questionnaire by replacing all mentions of AWCF with ChatGPT feedback. The questionnaire was divided into four sections: background information, behavioral, cognitive, and emotional engagement. The background section asked students to provide basic demographic details, such as their age, gender, major, and level of English proficiency. The behavioral engagement section included ten items, such as "I carefully read all the feedback provided by ChatGPT on my writing." The cognitive engagement section included seven items, such as "I reflect on the feedback I receive from ChatGPT and think about how I can apply it to future writing tasks." The emotional engagement section included eight items, such as "I feel optimistic about revising my writing when I receive feedback from ChatGPT." The questionnaire showed high reliability, with strong internal consistency across behavioral ($\alpha = .84$), cognitive ($\alpha = .82$), and emotional ($\alpha = .92$) engagement. In addition, to have a deeper insight of students' attitudes towards ChatGPT feedback, we adapted interview questions from Koltovskaia's (2020) study on student engagement with Grammarly feedback. The interview questions focused on students' experiences and perceptions of the effects of feedback on their writing performance, such as "Were you satisfied with the feedback provided? Why or why not?" All interviews were audio-recorded for further data analysis.

Data Analysis

A mixed-methods approach was employed to analyze both quantitative and qualitative data collected in this study. First, to evaluate the effectiveness of ChatGPT feedback on students' writing performance, we conducted a paired sample *t*-test to compare the pre-test and post-test scores of the EG. Specifically, we examined the scores across the four rubric categories, as well as the total writing score, to see whether there was a statistically significant improvement after using ChatGPT. Second, an independent sample *t*-test was used to compare the score improvements between the EG and CG to determine whether

the ChatGPT intervention led to significantly greater writing gains than teacher feedback. Regarding the questionnaire data, a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) was employed to measure students’ engagement with ChatGPT feedback across behavioral, cognitive, and emotional dimensions. We then ran a correlation analysis to explore whether there was a significant relationship between students’ engagement levels and their improvement in writing scores.

Moreover, we used thematic analysis to analyze the interview results. The recordings were transcribed verbatim and coded using both inductive and deductive approaches. Deductive coding was guided by our research objective to explore how ChatGPT feedback contributed to students’ writing improvement, while inductive coding allowed us to identify unexpected themes that emerged directly from the data. After coding, the data were organized and grouped into three main themes: perceived usefulness of feedback, increased motivation to revise, and uncertainty about the clarity and relevance of some feedback. To ensure reliability, one researcher conducted the initial coding, and a second researcher independently reviewed the codes and themes. Any discrepancies were discussed and resolved collaboratively to ensure consistency and accuracy in the findings.

Results

Effects of ChatGPT Feedback on Students’ Writing Performance

The pre-test and post-test writing scores of students in the EG are presented in Table 1. Students in this group showed significant improvement in grammar ($t(115) = -5.09, p < .001, d = 0.47$) and lexis ($t(115) = -4.99, p < .001, d = 0.46$) in the post-test. According to Cohen (1988), d -values of approximately 0.20 are considered small, 0.50 medium, and 0.80 or above large. Therefore, these findings suggest that ChatGPT feedback had a moderate impact on students’ grammar and lexical use. In contrast, no significant improvement was found in content ($t(115) = -0.08, p = .94, d = 0.01$) or organization ($t(115) = -0.97, p = .17, d = 0.09$). This suggests that while ChatGPT feedback was effective in enhancing surface-level writing skills, it did not improve students’ higher-order writing skills. For the total writing score, there was a statistically significant improvement as well ($t(115) = -3.07, p = .003$). However, the effect size was smaller ($d = -0.29$). This smaller effect size may be due to the limited improvement in higher-order writing skills, which contributed less to the overall gain.

The collected qualitative data further confirmed ChatGPT’s effectiveness in improving writing performance. During the interview, most participants claimed that its feedback was helpful and

Table 1. Pre-test and post-test writing scores in the EG

Aspects	Pre-test		Post-test		<i>t</i>	<i>p</i>	Cohen’s <i>d</i>
	Mean	<i>SD</i>	Mean	<i>SD</i>			
Content	69.77	7.46	69.87	13.01	−0.08	.938	−0.01
Organization	70.27	9.16	71.45	11.42	−0.97	.166	−0.09
Use of grammar	66.74	8.68	71.72	7.48	−5.09	<.001	−0.47
Use of lexis	67.39	9.02	72.16	6.94	−4.99	<.001	−0.46
Total score	274.16	27.46	285.20	33.27	−3.07	.003	−0.29

useful, especially in “providing synonyms,” “correcting grammar mistakes,” and “spotting spelling errors.” They were willing to correct the errors or use synonyms based on the suggestions generated by ChatGPT. This reinforces the findings that ChatGPT was particularly effective at addressing surface-level issues. When it came to higher-order writing skills, students had more mixed opinions. Some students reported that ChatGPT’s suggestions in these areas was useful. They claimed that suggestions such as “provide authentic data” and “offer relevant references” helped them to enrich the content of their writing and support their arguments. Others mentioned that ChatGPT gave them additional examples of transitional and linking phrases, which made it easier for them to improve the coherence and flow of their essays. On the other hand, other students struggled to understand or trust the higher-order feedback. A few students felt that the suggestions were “generic” or “robotic,” and “not really tailored to the writing content.” For instance, four students pointed out that the feedback on organization was nearly identical across different assignments, which led them to believe that the comments were not tailored to the specific content of their texts. As a result, they were less motivated to revise their writing based on the feedback.

Teacher Feedback vs. ChatGPT Feedback: Effects on Students’ Writing Improvement

To investigate whether teacher and ChatGPT feedback differ in their impact on students’ writing improvement, we first conducted a paired-sample *t*-test within the CG to compare students’ pre-test and post-test scores. The results showed that students in the CG did not exhibit significant improvement in any of the writing aspects. This may be due to two possible reasons. First, the instructor was unable to provide detailed and personalized feedback in large-scale classes. The course met twice a week, and the instructor was responsible for teaching three separate classes, each with a considerable number of students. Given the heavy teaching load and limited time, it was extremely challenging for the instructor to carefully review each student’s writing and offer individualized comments or suggestions. As a result, many students received only general feedback. For example, the instructor might have given feedback such as “Try to improve the coherence of your paragraphs” or “Check your use of connecting words.” Without specific corrections, students might struggle to recognize their mistakes, which could lead to the fossilization of such mistakes and hinder overall writing improvement (Bitchener, 2008). Second, unlike ChatGPT that can provide immediate feedback, teachers need much more time to read and comment on students’ writing. When students received teacher feedback, they might have already moved on to new tasks and forgotten what they had written. This could discourage them from revising their drafts. Therefore, the delayed feedback could have reduced its effectiveness in improving writing performance (Hattie & Timperley, 2007; Shute, 2008).

Table 2 Comparison of writing improvement between the CG and EG

Improvement	CG		EG		<i>t</i>	<i>p</i>	Cohen’s <i>d</i>
	Mean	<i>SD</i>	Mean	<i>SD</i>			
Content	0.17	9.50	0.10	14.28	0.04	.970	0.01
Organization	−1.58	11.27	1.18	13.07	−1.83	.070	−0.23
Use of grammar	−0.51	8.41	4.98	10.54	−4.63	<.001	−0.59
Use of lexis	0.28	7.97	4.77	10.28	−3.91	<.001	−0.50
Total score	−0.30	38.98	11.03	38.77	−3.06	.002	−0.38

To further investigate whether students who received ChatGPT feedback improved more than those who received teacher feedback, we conducted an independent samples *t*-test to compare differences in writing improvement between the EG and CG. The results are shown in Table 2. Students in the EG demonstrated significantly greater improvement than the CG in the total writing score ($t(271) = -3.06, p = .002, d = -0.38$). Specifically, students in the EG outperformed those in the CG in both use of grammar ($t(271) = -4.63, p < .001, d = -0.59$) and use of lexis ($t(271) = -3.91, p < .001, d = -0.50$). In contrast, no significant differences were found between the groups in content ($t(271) = .041, p = .97, d = 0.01$) or organization ($t(271) = -1.83, p = .07, d = -0.23$). Overall, these results further suggest ChatGPT feedback’s effectiveness in improving surface-level issues. However, for higher-order aspects, neither the ChatGPT intervention nor traditional teacher feedback led to significant improvement, and no significant differences were observed between the two groups in these areas.

The Relationship Between Students’ Engagement and Writing Improvement

Table 3 presents the Pearson correlation coefficients among writing improvement dimensions and the three dimensions of student engagement. The results revealed significant positive relationships between all three dimensions of student engagement and improvements in writing performance. Among the three dimensions, behavioral engagement showed the strongest correlations across all writing dimensions, including content ($r = 0.33, p < .001$), organization ($r = 0.36, p < .001$), grammar ($r = 0.36, p < .001$), lexis ($r = 0.40, p < .001$), and overall improvement ($r = 0.45, p < .001$). Cognitive and emotional engagement were also significantly correlated with writing improvement. These positive correlations suggest that students who were more actively engaged with ChatGPT feedback tended to improve more in their writing performance.

Table 3 Associations between EG students’ engagement and writing improvement

Writing improvement	Behavioral engagement		Cognitive engagement		Emotional engagement	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Content	0.33	<.001	0.30	.001	0.28	.003
Organization	0.36	<.001	0.30	<.001	0.25	.007
Use of grammar	0.36	<.001	0.25	.006	0.27	.004
Use of lexis	0.40	<.001	0.27	.003	0.29	.001
Total score	0.45	<.001	0.35	<.001	0.34	<.001

To further explore the predictive power of different dimensions of student engagement on writing improvement, a multiple regression analysis was conducted (see Table 4). Behavioral engagement, cognitive engagement, and emotional engagement were entered together as independent variables, and overall writing improvement as the dependent variable. The overall regression model reached statistical significance ($F(3, 112) = 10.58, p < .001$) and accounted for 22.1% of the variance in writing improvement ($R^2 = .22, \text{Adjusted } R^2 = .20$). This further supports that student engagement plays a meaningful role in affecting writing outcomes.

Table 4 Regression analysis of the EG students’ engagement and writing improvement

Predictor	B	SE B	β	<i>t</i>	<i>p</i>
Constant	−52.47	13.30	—	−3.95	< .001
Behavioral engagement	32.34	8.73	.78	3.71	< .001
Cognitive engagement	−6.04	10.13	−.149	−0.60	.552
Emotional engagement	−8.73	9.34	−.219	−0.94	.352

Moreover, among the three predictors, behavioral engagement was the only significant contributor to the model ($\beta = .777, t = 3.71, p < .001$), while cognitive engagement ($\beta = -.149, t = -.60, p = .552$) and emotional engagement ($\beta = -.219, t = -.94, p = .352$) did not significantly predict writing improvement. Therefore, while all three dimensions of engagement were positively correlated with writing improvement, only behavioral engagement significantly predicted improvement when controlling the other predictors. In other words, students who had more active behavioral engagement with ChatGPT feedback, such as revising drafts or interacting more rounds with ChatGPT, were more likely to make more progress in their writing. The interview results echoed the survey findings. They showed that students who actively engaged with ChatGPT’s feedback made the most progress in their writing. These students didn’t just read the feedback but used it to make concrete changes to their writing. For example, one student said, “I usually revised my draft three times after getting feedback from ChatGPT. Every time I fixed something, I asked it again to check my changes.” Some students also paid close attention to the quality of the suggestions. One student said, “ChatGPT gave me a suggestion that didn’t really fit my argument, so I thought about it and decided not to use it.” Although a few students showed this kind of critical thinking, they didn’t always go back to revise their work. Emotionally, students responded in different ways. One student said, “I felt more confident because I knew I could check my writing anytime.” Another felt unsure, saying, “I liked getting quick help, but I wasn’t always sure if the feedback was correct.” These insights suggest that although cognitive and emotional engagement contribute to the overall learning experience, it is behavioral engagement that most strongly supports measurable writing improvement.

Discussion

Effectiveness of ChatGPT Feedback on Writing Performance

While recent literature on the effectiveness of AI-generated AWCF remains mixed, our study used ChatGPT as an AWE tool to provide feedback on students’ writing and examined its long-term effects across different aspects of writing, including content, organization, use of lexis, and use of grammar. To assess its impact, we compared students’ pre- and post-test writing performance after ChatGPT intervention. The results show that ChatGPT feedback significantly improved students’ grammatical accuracy and lexical use in their writing. These results align with earlier studies suggesting that GAI feedback is particularly effective in addressing lower-level writing issues, such as grammar and word choice (e.g., Koltovskaia et al., 2024; Zhan & Yan, 2025). Unlike prior studies that examined the effectiveness of ChatGPT feedback after a single draft revision (e.g., Chan et al., 2024; Pariyanto & Tungka, 2025), our study allowed students to use ChatGPT multiple times over the course of a semester. The findings revealed that the long-term effects of ChatGPT feedback on students’ writing performance remained significant across time. One likely reason is that ChatGPT provides immediate, consistent, and

straightforward explanations, along with alternative phrasing and revision suggestions that help students refine their writing more efficiently (Zhang et al., 2025). Moreover, students' positive attitudes toward ChatGPT feedback may also enhance their willingness to engage with the feedback, which in turn improved its effectiveness. Our qualitative data showed that students generally hold a favorable attitude towards ChatGPT-generated AWCF. This was in line with Yan and Zhang (2024) and Tossell et al.'s (2024) findings, who also reported that students viewed AI-generated feedback as supportive and motivating in the writing process. Many students described the feedback they obtained from ChatGPT as helpful, clear, and easy to understand, which made it easier to recognize and correct the errors independently. As a result, they were more motivated to revise their writing, which in turn led to greater improvements in their overall writing performance.

However, this ChatGPT intervention was less effective in enhancing higher-order writing skills such as content development and organization. Several factors may explain this. First, our interviews showed that some students did not fully understand or trust the feedback ChatGPT provided on these aspects. Even though ChatGPT can generate human-like responses, it lacks true understanding of authorial intent, which limits its ability to offer feedback on the writing content (Roumeliotis & Tselikas, 2023; Steiss et al., 2024). As a result, the feedback ChatGPT generated might not always be relevant or aligned with the students' intended meaning or writing goals, which reduced its usefulness in helping them improve these areas. Second, students' engagement with ChatGPT feedback can also influence its effectiveness. Feedback on surface-level issues is less cognitively demanding. It often involves simple corrections that students can apply directly without much effort (Zhang et al., 2025). In contrast, when feedback requires deep thinking, such as revising content or reorganizing ideas, learners may experience cognitive overload, which can reduce their willingness or ability to revise effectively (Cho & MacArthur, 2010; Lee, 2017). This lower engagement with the feedback may lead to less improvement in higher-order writing skills.

Our study also conducted a comparison between the EG and the CG. The findings showed that the EG made significantly more progress in grammar, vocabulary, and overall writing performance. In contrast, the CG showed no statistically significant improvement across any writing domain. This finding differs from earlier research, which emphasized the importance of teacher feedback in helping students revise complex aspects of their writing (Bitchener & Ferris, 2012; Hyland & Hyland, 2019). One possible reason was that the large class size and heavy teaching workload may have limited the instructor's capacity to provide detailed and individualized comments. As a result, the feedback students received was likely too general to support meaningful revision. However, this finding does not necessarily suggest that ChatGPT feedback is superior to teacher feedback. Recent studies have highlighted the complementary roles of AI and human instructors in writing instruction. For example, Zhang et al. (2025) proposed a hybrid approach, in which GAI focuses on surface-level revisions and teachers focus on content, structure, and rhetorical development. Chen (2025) further proposed the Student-Teacher-AI Collaboration Model, which highlighted teachers' responsibility in the hybrid model. They use AI-generated data to track students' writing development and give tailored support based on their strengths and weaknesses. This is in line with the roles and framework of GAI in mobile learning environments proposed by Lai and Tu (2025). GAI should act as a supportive tool in teaching by providing resources such as knowledge, explanations, and performance evaluation. Importantly, learners should not merely use GAI passively. Instead, they are expected to engage actively, such as asking critical questions, verifying the feasibility of information, and generating original ideas. This approach not only reinforces students' academic integrity but also strengthens critical thinking skills.

To sum up, while ChatGPT was effective in improving students' use of lexis and grammar, its limitations in supporting higher-order writing skills suggest that it cannot be a replacement of human feedback. ChatGPT, as well as other GAI tools, is best viewed not as replacements for human instruction, but as assistive scaffolds that can enhance the writing process (Javier & Moorhouse, 2024). Ultimately, the most effective writing instruction comes from integration of GAI support and human guidance, where each complements the other to foster writing improvement.

Role of Student Engagement in Writing Improvement

Based on Zhang and Hyland's (2018) model of learner engagement with AWCF, our study also explored the relationship between students' writing improvement and three dimensions of engagement, including behavioral, cognitive, and emotional engagement. The findings show that all three dimensions of engagement were positively correlated with the EG students' writing improvement after ChatGPT intervention. This is in line with earlier findings, which suggest students' engagement plays a central role in supporting learning outcomes (Lipnevich et al., 2021; Zhang & Hyland, 2018). However, our study found that behavioral engagement was the only significant predictor in the regression model. This suggests that students who actively interacted with ChatGPT feedback behaviorally were more likely to improve in their writing. These results align with Fleckenstein et al. (2024) and Sun and Zhang (2024), who also found that actual behavioral engagement was more directly linked to student progress. Behavioral engagement directly reflects the effort students put into the revision process (Ellis, 2010). When students act on ChatGPT's suggestions and revise their work, they are not only modifying their drafts but also engaging in a process of implicit learning. For instance, by adopting a suggested synonym, a student may subconsciously retain the new lexical item and later apply it in future writing. Over time, these revisions can accumulate into long-term linguistic development; hence, the writing performance beyond the immediate task can be improved.

It is also important to notice that the three dimensions of engagement are interconnected. While behavioral engagement was the strongest predictor, cognitive and emotional engagement might also influence the degree and quality of behavioral engagement. Emotional engagement fosters a sense of motivation and self-efficacy, which in turn can increase the invested behavioral effort (D'Mello & Graesser, 2012). When students perceived ChatGPT feedback as useful and relevant to their writing, they tended to feel more confident, positive, and willing to revise. This emotional response also increased their trust in the feedback. As a result, they were more likely to take up the feedback and apply it effectively in revision, which contributed to greater writing improvement. Moreover, emotionally engaged students are usually more motivated to use cognitive strategies in processing feedback. These include evaluating the validity of suggestions, connecting them with existing linguistic knowledge, or making decisions about whether to accept, adapt, or reject the feedback (Koltovskaia, 2020). These strategies help students internalize new language patterns and improve their writing skills over time.

Therefore, in order to improve the effectiveness of ChatGPT feedback, it is important to enhance student engagement. One way is to provide students with clear guidance on how to obtain the feedback (Chen, 2025). When students understand how the tool works, they are more likely to trust and use its suggestions. Teachers should also train students to critically refine AI-generated suggestions. For example, students can write a short reflection after using ChatGPT to explain which suggestions they accepted, which they did not, and why. Teachers can organize group discussions where students explain how they interpreted and applied the feedback in their own context. This helps them adapt the feedback to fit their own voice and ideas, rather than accepting it passively (Yang et al., 2025). Moreover, the limited effective-

ness of ChatGPT feedback in higher-order aspects further emphasizes the importance of teachers. Teachers' expertise is crucial in guiding students through the more complex aspects of the revision process, which supports deeper and more meaningful learning (Bitchener & Ferris, 2012; Hyland & Hyland, 2019). For example, teachers can guide students to give more detailed prompts to ChatGPT based on their specific writing needs. When students do not understand higher-order feedback from ChatGPT, teachers can step in to explain the meaning, clarify the purpose of the suggestion, and offer targeted advice on how to revise. Teachers may also provide mini lessons on key writing skills that ChatGPT cannot teach effectively, such as building logical arguments, using sources properly, or adjusting tone for different audiences.

Conclusion

This study investigated the impact of ChatGPT feedback on students' writing performance in a higher education context, comparing its effects with traditional teacher feedback and exploring the role of student engagement in predicting writing improvement. The findings provide empirical support for the growing role of GAI in writing instruction, particularly in enhancing surface-level language features such as grammar and lexis. In contrast, its limitations in addressing higher-order writing skills were also evident. Moreover, we found that students who engaged more actively with ChatGPT feedback tended to gain more improvement in writing. To be more specific, behavioral engagement emerged as the most significant and direct predictor of the effectiveness of ChatGPT feedback.

These findings suggest that while GAI can serve as a valuable supplementary tool, it should be integrated with other sources of feedback, especially human feedback (Ding & Zou, 2024). Teacher instruction and professional guidance is still essential to ensure the individualization of writing instruction (Evmenova et al., 2024). Moreover, simply incorporating GAI into the classroom is not enough. Meaningful interactions with such tools should be actively encouraged to maximize their pedagogical potential (Graham et al., 2024). Future research should further explore this process by examining how different types of student engagement influence the effectiveness of GAI feedback. Such investigations can help identify best practices for integrating GAI feedback into writing instruction.

Our study has several limitations. First, the sample was limited to students from a similar background. This may reduce the applicability of the findings to broader populations, as learners with different levels of English proficiency, educational experiences, or cultural perspectives may engage with GAI feedback in different ways. Future research should include more diverse student groups to examine whether similar results can be observed in other educational contexts, such as secondary schools, vocational institutions, or universities in non-English-speaking regions. Second, the study did not examine the potential of hybrid feedback model that integrates GAI feedback with human feedback. Given the growing interest in such approaches, future research could explore how combining GAI feedback with teacher guidance may enhance writing outcomes more effectively than either source alone (Zhang et al., 2025).

References

- Allison, J., Hwang, G.-J., Mayer, R. E., Pellas, N., Karnalim, O., de Freitas, S., Ng, O.-L., Huang, Y.-M., Hooshyar, D., Seidman, R. H., AlEmran, M., Mikropoulos, T. A., Schroeder, N. L., Roscoe, R. D., & Sanusi, I. (2025). From generative AI to extended reality: Multidisciplinary perspectives on the challenges, opportunities, and future of educational computing. *Journal of Educational Computing Research*, 63(6), 1327–1363. <https://doi.org/10.1177/07356331251359964>
- Alsofyani, A. H., & Barzanji, A. M. (2025). The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study. *Journal of Educational Computing Research*, 63(2), 431–463. <https://doi.org/10.1177/07356331241307297>

- Barrot, J. S. (2023). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning*, 36(4), 584–607. <https://doi.org/10.1080/09588221.2021.1936071>
- 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., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.
- Chan, S., Lo, N., & Wong, A. (2024). Generative AI and essay writing: Impacts of automated feedback on revision performance and engagement. *rEFlections*, 31(3), 1249–1284. <https://doi.org/10.61508/refl.v31i3.277514>
- Chang, C. C., & Hwang, G. J. (2025). ChatGPT-facilitated professional development: Evidence from professional trainers' learning achievements, self-worth, and self-confidence. *Interactive Learning Environments*, 33(1), 883–900. <https://doi.org/10.1080/10494820.2024.2362798>
- Chen, Q. (2025). Students' perceptions of AI-powered feedback in English writing: Benefits and challenges in higher education. *International Journal of Changes in Education*. <https://doi.org/10.47852/bonviewIJCE52025580>
- Cheng, X., & Zhang, L. J. (2024). Examining Second Language (L2) Learners' Engagement with AWE-Teacher Integrated Feedback in a Technology-Empowered Context. *The Asia-Pacific Education Researcher*, 1–13. <https://doi.org/10.1007/s40299-024-00877-8>
- Chowdhury, A. G., Islam, M. M., Kumar, V., Shezan, F. H., Jain, V., & Chadha, A. (2024). Breaking down the defenses: A comparative survey of attacks on large language models. *arXiv*. <https://arxiv.org/abs/2403.04786>
- 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*, 29(11), 14151–14203. <https://doi.org/10.1007/s10639-023-12402-3>
- D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145–157. <https://doi.org/10.1016/j.learninstruc.2011.10.001>
- Ellis, R. (2010). A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*, 32(2), 335–349. <https://doi.org/10.1017/S0272263109990544>
- 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(1), Article 57. <https://doi.org/10.1186/s41239-023-00425-2>
- Evmenova, A. S., Regan, K., Mergen, R., & Hrisseh, R. (2024). Improving writing feedback for struggling writers: Generative AI to the rescue? *TechTrends*, 68(4), 790–802. <https://doi.org/10.1007/s11528-024-00965-y>
- Fan, N. (2023). Exploring the effects of automated written corrective feedback on EFL students' writing quality: A mixed-methods study. *SAGE Open*, 13(2), Article 21582440231181296. <https://doi.org/10.1177/21582440231181296>
- Fleckenstein, J., Jansen, T., Meyer, J., Trueb, R., Raubach, E. E., & Keller, S. D. (2024). How am I going? Behavioral engagement mediates the effect of individual feedback on writing performance. *Learning and Instruction*, 93, Article 101977. <https://doi.org/10.1016/j.learninstruc.2024.101977>
- Graham, S., Cao, Y., Kim, Y. G., Lee, J., Tate, T., Collins, P., Cho, M., Moon, Y., Chung, H. Q., & Olson, C. B. (2024). Effective writing instruction for students in grades 6 to 12: A best evidence meta-analysis. *Reading and Writing*, 38, 1–46. <https://doi.org/10.1007/s11145-024-10539-2>
- Guo, Q., Feng, R., & Hua, Y. (2022). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning*, 35(9), 2312–2331. <https://doi.org/10.1080/09588221.2021.1879161>
- Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, 30, 31–44. <https://doi.org/10.1016/j.jslw.2015.08.002>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/0034654302984>
- Hockly, N. (2019). Automated writing evaluation. *ELT Journal*, 73(1), 82–88. <https://doi.org/10.1093/elt/ccy044>

- Huang, J., & Mizumoto, A. (2025). The role of generative AI in mediating L2MSS and engagement with written feedback in EFL learning: A structural equation modeling approach. *Annual Review of Applied Linguistics*, 1–17. <https://doi.org/10.1017/S0267190525000029>
- 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>
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83–101. <https://doi.org/10.1017/S0261444806003399>
- Javier, D. R. C., & Moorhouse, B. L. (2024). Developing secondary school English language learners' productive and critical use of ChatGPT. *TESOL Journal*, 15(2), Article e755. <https://doi.org/10.1002/tesj.755>
- Jeon, J. (2023). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning*, 36(7), 1338–1364. <https://doi.org/10.1080/09588221.2021.1987272>
- Jiang, H., & Yu, X. (2025). L2 grit and its association with online learning engagement and interaction strategies: A structural equation model. *The Asia-Pacific Education Researcher*, 34(3), 1145–1156. <https://doi.org/10.1007/s40299-024-00929-z>
- Kartika, S. (2025). The role of generative AI in automating English writing assessment: Benefits and challenges. *LinguaEducare: Journal of English and Linguistic Studies*, 2(1), 1–10. <https://doi.org/10.63324/lec.2v.1i.63>
- Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, 44, Article 100450. <https://doi.org/10.1016/j.asw.2020.100450>
- Koltovskaia, S., & Mahapatra, S. (2022). Student engagement with computer-mediated teacher written corrective feedback: A case study. *The JALT CALL Journal*, 18(2), 286–315. <https://doi.org/10.29140/jaltcall.v18n2.519>
- Koltovskaia, S., Rahmati, P., & Saeli, H. (2024). Graduate students' use of ChatGPT for academic text revision: Behavioral, cognitive, and affective engagement. *Journal of Second Language Writing*, 65, Article 101130. <https://doi.org/10.1016/j.jslw.2024.101130>
- Lai, C. L., & Tu, Y. F. (2024). Roles, strategies, and research issues of generative AI in the mobile learning era. *International Journal of Mobile Learning and Organisation*, 18(4), 516–537. <https://doi.org/10.1504/IJMLO.2025.10059769>
- Li, B., Lowell, V. L., Wang, C., & Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. *Computers and Education: Artificial Intelligence*, 7, Article 100266. <https://doi.org/10.1016/j.caeai.2024.100266>
- Li, J., & Yuan, F. (2024). Does engagement with feedback matter? Unveiling the impact of learner engagement and grit on EFL learners' English writing achievements. *Language Teaching Research*. <https://doi.org/10.1177/13621688241257865>
- Liu, W. (2025). A bibliometric analysis of written corrective feedback in second language writing. *Language Teaching Research Quarterly*, 49, 133–150. <https://doi.org/10.32038/ltrq.2025.49.07>
- Lipnevich, A. A., Murano, D., Krannich, M., & Goetz, T. (2021). Should I grade or should I comment: Links among feedback, emotions, and performance. *Learning and Individual Differences*, 89, Article 102020. <https://doi.org/10.1016/j.lindif.2021.102020>
- Mao, Z., & Lee, I. (2024). Student engagement with written feedback: Critical issues and way forward. *RELJ Journal*, 55(3), 810–818. <https://doi.org/10.1177/00336882221150811>
- Özdere, M. (2025). AI in academic writing: Assessing the effectiveness, grading consistency, and student perspectives of ChatGPT and You.com for EFL students. *International Journal of Technology in Education (IJTE)*, 8(1), 123–154. <https://doi.org/10.46328/ijte.1001>
- Pariyanto, P., & Tungka, N. F. (2025). ChatGPT as a formative feedback tool: Improving narrative essay writing among EFL students. *Edulitics (Education, Literature, and Linguistics) Journal*, 10(1), 71–84. <http://e-journal.unisda.ac.id/>
- Rahimi, M., Fathi, J., & Zou, D. (2025). Exploring the impact of automated written corrective feedback on the academic writing skills of EFL learners: An activity theory perspective. *Education and Information Technologies*, 30(3), 2691–2735. <https://doi.org/10.1007/s10639-024-12896-5>

- Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing*, 52, Article 100816. <https://doi.org/10.1016/j.jslw.2021.100816>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, Article 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19, 51–65. <https://doi.org/10.1016/j.asw.2013.11.007>
- Sun, P. P., & Zhang, L. J. (2024). Investigating the effects of Chinese university students' online engagement on their EFL learning outcomes. *The Asia-Pacific Education Researcher*, 33(4), 747–757. <https://doi.org/10.1007/s40299-023-00800-7>
- Tan, S., Cho, Y. W., & Xu, W. (2023). Exploring the effects of automated written corrective feedback, computer-mediated peer feedback and their combination mode on EFL learner's writing performance. *Interactive Learning Environments*, 31(10), 72761–7286. <https://doi.org/10.1080/10494820.2022.2066137>
- Teng, M. F. (2024). 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>
- Tossell, C. C., Tenhundfeld, N. L., Momen, A., Cooley, K., & De Visser, E. J. (2024). Student perceptions of ChatGPT use in a college essay assignment: Implications for learning, grading, and trust in artificial intelligence. *IEEE Transactions on Learning Technologies*, 17, 1069–1081. <https://doi.org/10.1109/TLT.2024.3355015>
- Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, 58(1), 87–125. <https://doi.org/10.1177/0735633119830764>
- Wu, J., Li, J., Ge, Z., Xu, M., Lin, L., & Zhang, R. (2025). Effectiveness of generative AI in automated written corrective feedback with prompting. *Journal of Educational Computing Research*, 63(6), 1493–1527. <https://doi.org/10.1177/07356331251359430>
- Xue, Y. (2024). Towards automated writing evaluation: A comprehensive review with bibliometric, scientometric, and meta-analytic approaches. *Education and Information Technologies*, 29(15), 19553–19594. <https://doi.org/10.1007/s10639-024-12596-0>
- Yan, D., & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. *Humanities and Social Sciences Communications*, 11(1), 1–14. <https://doi.org/10.1057/s41599-024-03543-y>
- Yancey, K. P., Laflair, G., Verardi, A., & Burstein, J. (2023, July). Rating short L2 essays on the CEFR scale with GPT-4. In *Proceedings of the 18th Workshop on Innovative use of NLP for Building Educational Applications (BEA 2023)* (pp. 576–584).
- Yang, K., Raković, M., Liang, Z., Yan, L., Zeng, Z., Fan, Y., & Chen, G. (2025, March). Modifying AI, enhancing essays: How active engagement with generative AI boosts writing quality. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference* (pp. 568–578).
- Yeung, S. (2025). University students' engagement with generative AI-supported automated writing evaluation (AWE) feedback. *Journal of Second Language Writing*, 68, Article 101203. <https://doi.org/10.1016/j.jslw.2025.101203>
- Zhan, Y., & Yan, Z. (2025). Students' engagement with ChatGPT feedback: Implications for student feedback literacy in the context of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 1–14. <https://doi.org/10.1080/02602938.2025.2471821>
- Zhang, J., & Wang, J. (2025). Student perceptions of hybrid feedback: Using gen-AI to enhance engagement with EAP writing feedback. *The JALT CALL Journal*, 21(2), 2175–2175. <https://doi.org/10.29140/jaltcall.v21n2.2175>
- Zhang, Z., Aubrey, S., Huang, X., & Chiu, T. K. (2025). The role of generative AI and hybrid feedback in improving L2 writing skills: A comparative study. *Innovation in Language Learning and Teaching*, 1–19. <https://doi.org/10.1080/17501229.2025.2503890>

- Zhang, Z. V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, 43, Article 100439. <https://doi.org/10.1016/j.asw.2019.100439>
- Zhang, Z. V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90–102. <https://doi.org/10.1016/j.asw.2018.02.004>
- Zou, D., Xie, H., & Wang, F. L. (2023). Effects of technology enhanced peer, teacher and self-feedback on students' collaborative writing, critical thinking tendency and engagement in learning. *Journal of Computing in Higher Education*, 35(1), 166–185. <https://doi.org/10.1007/s12528-022-09337-y>