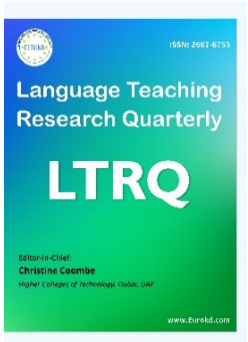


Language Teaching Research Quarterly

2025, Vol. 47, 113–137

Generative AI vs. Teachers: Feedback Quality, Feedback Uptake, and Revision

Hetian Yu[1], Qin Xie[2]*

[1]The Education University of Hong Kong, Hong Kong, China
[2]Faculty of Education, University of Macau, Macau, China

Received *16 December 2024* ***Accepted*** *20 March 2025*

Abstract

Generative AI has stormed the education community worldwide. Gen-AI's potential to assist teachers in providing feedback to students is the focus of the current research. Specifically, the research investigated how similar or different AI-generated feedback was from teacher feedback, and whether there were differences in students' responses to the two feedback sources. The research adopted a 2x2 counter-balance experimental design (Latin Square), whereby the participants were divided into two groups to write on two topics, receive feedback, and resubmit the revised work. The participants were 60 EFL secondary students and four teachers from a high school in China. All teachers and a Gen-AI bot (ChatGPT) marked and provided feedback on the 240 written samples. The analysis of the 1200 records (240*5) found that the AI-bot's surface-level feedback was comprehensive, accurate, and similar to that of one teacher. It outperformed the teacher qualitatively and quantitatively in the meaning-level feedback. Overall, teacher feedback received higher student uptake rates, but the difference was small. Students were capable of adopting a variety of strategies to respond to the feedback regardless of its sources. The findings support adopting ChatGPT feedback as a supplement to teacher feedback given to novice EFL writers.

Keywords: *ChatGPT, Written Corrective Feedback, EFL Writing, Revision Operations, Feedback Uptake*

How to cite this article (APA 7th Edition):

Yu, H., & Xie, Q. (2025). Generative AI vs. teachers: Feedback quality, feedback uptake, and revision. *Language Teaching Research Quarterly*, *47*, 113-137. https://doi.org/10.32038/ltrq.2025.47.07

Introduction

Feedback, a widely used pedagogical activity in second-language writing classrooms, has drawn tremendous scholarly interest in the last two decades (Hyland & Hyland, 2006; Yu & Lee, 2016). A plethora of studies have examined its effectiveness and identified substantial evidence that feedback from an agent such as a teacher, computer, or peer could improve

* Corresponding author.
E-mail address: qinxie@um.edu.mo
https://doi.org/10.32038/ltrq.2025.47.07

writing competence (Zhang, 2020), enhance writing motivation (Zhang & Hyland, 2018), and promote self-regulated learning strategies (Yu & Lee, 2016).

Written corrective feedback (WCF) is a widely used formative assessment technique in English writing classrooms; it is usually given in the form of teachers' written comments on students' written work. WCF could improve students' writing proficiency (Han & Hyland, 2015), accuracy, and effectiveness (Ferris, 2006). However, providing detailed WCF takes considerable teacher time, especially in Asia regions where the class size is typically large (Lee, 2019). Machines and artificial intelligence (AI) can reduce teachers' workload by providing instant feedback on students' work.

Although the emergence of Gen-AI and its application in providing feedback is relatively new, the literature on its predecessor, automated writing evaluation systems (AWEs), is abundant (e.g., Li, 2024), which could provide some conceptual basis for the current research. Many AWE studies investigated learners' perceptions of AWE feedback (Bai & Hu, 2017), its effectiveness on writing skills (Koltovskaia, 2020), and learners' revision (O'Neill & Russell, 2019). The emergence of ChatGPT, with its potential for providing intelligent responses, has prompted the present researchers to wonder how effective its feedback may be and how students respond to GenAI feedback. The present study answered these questions, given their conceptual and pedagogical importance. Gen-AI's considerable efficiency in generating large amounts of intelligent responses to text prompts means it has the potential to free teachers from the time-consuming, repetitive task of providing WCF on students' writing. The study could, therefore, offer new insights into the future educational landscape in the AI era.

Literature Review

Written Corrective Feedback (WCF)

Widely used in second language (L2) writing instruction, teacher WCF is essential to improving EFL learners' writing performance and promoting their long-term writing development (Ferris, 2006). Three types of WCF were identified (Ellis, 2010), namely, *direct*, *indirect*, and *metalinguistic*. Direct WCF refers to providing learners with the correct forms directly; indirect WCF only indicates errors or problems without correction. Metalinguistic WCF offers explanations for errors (Ferris, 2006). WCF is also classified according to whether it is focused or unfocused. Focused feedback targets specific types of errors; unfocused feedback, on the other hand, addresses multiple errors in students' essays.

Although previous research has explored the effectiveness of WCF (e.g., Kang & Han, 2015), the findings have been mixed. On the one hand, there is evidence that teacher WCF is effective in helping learners reduce errors and improve accuracy (Suzuki et al., 2019). Sheen (2007), for instance, reported that focused WCF could reduce learners' attentional load, enhance their understanding of their errors, and guide them in modifying their writing. Rahimi (2021) also argued that comprehensive WCF could improve learners' linguistic accuracy and overall writing development. Similarly, Van Beuningen et al. (2012) reported that direct and indirect WCF had positive long-term impacts on writing accuracy.

On the other hand, studies also report that WCF, focused or unfocused, did not improve learners' writing. According to Lee (2019), WCF focusing on limited language errors did not contribute to the improvement of EFL learners' ability to use more complex linguistic forms, whereas unfocused comprehensive feedback, which identifies a large number of errors, risked

overloading learners' attentional capacity and did not result in significant improvement either. Moreover, sentence-level WCF helped learners correct errors but did not improve the overall quality of writing (Lee, 2019).

As such, the existing research on WCF has concluded that the latter's effects on writing quality depend on many factors, including feedback mode (focused or comprehensive) and quality and individual learner characteristics, such as language proficiency and willingness to take up feedback.

Regarding feedback take-up and revision behaviors, previous research found that EFL learners tended to have a high rate of taking up and adopting a variety of revision strategies (Han & Hyland, 2015; Tian & Zhou, 2020). Specifically, high-proficiency learners addressed 72.09% of surface-level feedback, whereas low-proficiency ones addressed only 40.45% (Cheng & Liu, 2022). High-proficiency learners were also more willing to adopt revision strategies. While high or low-proficiency students were generally open to surface-level feedback (Han & Hyland, 2015), they tended to pay limited attention to the meaning-level feedback (Cheng & Liu, 2022). Such differences were observed in their uptake rates across different writing tasks (Tian & Zhou, 2020).

In summary, EFL learners tended to accept teacher WCF. However, their feedback uptake was influenced by a range of factors, such as the nature of feedback (surface or meaning), language proficiency (Cheng & Liu, 2022), learning goals (Zhang & Hyland, 2022), and the difficulty of the writing tasks (Tian & Zhou, 2020). Further research is needed to examine revision behaviors in response to different teacher feedback.

Automated Feedback

Existing literature on machine-generated feedback primarily concerns automated writing evaluation (AWE) systems predating generative AI, such as *Criterion, Grammarly, Pigai, and MY Access*! However, hitherto, there is little published research regarding feedback from Gen-AI in L2 writing (e.g., ChatGPT). As such, the present researchers surveyed existing research concerning AWE feedback to inform the conceptualization of the present study.

The findings on AWE feedback are mixed. On the one hand, AWE systems could provide instant, specific, and individualized feedback on the content and the language of learners' writing (Dikli & Bleyle, 2014). They were also found capable of improving writing accuracy (Link et al., 2022; Shi et al., 2025), offering suggestions on content (Ranalli, 2018), increasing holistic writing scores (Dikli & Bleyle, 2014) and learner autonomy (Ware, 2011).

On the other hand, AWE feedback was also found to have problems (Karatay & Karatay, 2024), notably focusing on surface-level errors but neglecting meaning-level issues (Guo et al., 2021), being vague, repetitive, and inconsequential to the quality of writing (Koltovskaia, 2020), lacking sufficient accuracy, missing errors, and misleading (Dikli & Bleyle, 2014).

A handful of studies explored EFL learners' uptake of AWE feedback and revision strategies (e.g., Bai & Hu, 2017; Li et al., 2015; Zhang, 2020) and reported varied findings. Specifically, a wide range of uptake rates was identified, from very low (11.5% in Bai & Hu, 2017) and medium (49% in Cotos et al., 2015) to large (73% in Lavolette et al., 2015). In Bai & Hu (2017), feedback uptake was only 11.5%, primarily because the feedback was of low accuracy. In Cotos et al. (2015), almost 51% of feedback was ignored because of the learners'

low language proficiency. Interestingly, however, uptake rates were high in Laviolette et al. (2015); their participants revised successfully based on most feedback.

In terms of revision, Li et al. (2015) reported that their participants submitted their writing to AWE systems multiple times to receive more feedback to inform their revision. Zhang and Hyland (2022) identified six categories of revision: *correction*, *no correction*, *addition*, *substitution*, *reorganization*, and *deletion*. The high-proficiency participants in their study adopted various strategies to make revisions, while the low-proficiency ones did not actively respond to the feedback.

To sum up, existing studies on AWE feedback revealed that students responded to machine-generated feedback in different ways and to different degrees, depending on the quality of feedback and individual factors (e.g., language proficiency and motivation).

Comparison of Teacher- and Machine-Generated Feedback

Only a few studies compared the feedback from teachers and machines regarding their effectiveness and students' feedback uptake and revision actions (Link et al., 2022; Ware, 2011; Zhang & Hyland, 2022); their findings were mixed. Rock (2007) reported that learners who received feedback from Criterion achieved higher scores on their writing than those who received feedback from teachers. However, Ware (2011) found the opposite: Learners who received teacher feedback achieved higher scores than those who received AWE feedback; the latter did not make significant improvements in either language or content because the feedback was vague and inaccurate.

Likewise, few studies have compared students' responses to the two feedback sources. Notably, Zhang and Hyland (2022) explored students' revision operations based on teacher and AWE feedback and found that teacher feedback was highly valued by students and had a higher uptake rate. Similarly, Link et al. (2022) reported that students made more revisions based on teacher feedback than AWE feedback. A similar finding was also reported by Tian and Zhou (2020), who found teacher WCF to have a higher uptake rate than AWE feedback.

ChatGPT and Its Potential for Feedback

The recent advent of generative AI, particularly ChatGPT, has stormed the education community worldwide. The enormous potential offered by ChatGPT to education has attracted considerable attention. Of interest to the present study is its potential to provide feedback to EFL learners' writing. ChatGPT, short for Chat Generative Pre-Trained Transformer, is a large language model (LLM)-driven Chatbot developed by OpenAI. Its latest iteration, GPT-4o, was released in May 2024 (OpenAI, 2024).

ChatGPT can generate texts, grade essays, proofread and suggest specific revisions. Like earlier AWE systems such as Criterion and Grammarly, ChatGPT can generate detailed written feedback, pointing out strengths and weaknesses and offering suggestions for improvement. Unlike the AWE tools that preceded it, ChatGPT relies on the user to prompt it, and depending on the quality of the prompts, ChatGPT feedback varies. Consequently, learners not adequately trained in writing prompts may not obtain satisfactory feedback from ChatGPT.

A few recent publications examined the application of ChatGPT in L2 writing (Amer et al., 2025; Lu et al., 2024; Shi et al., 2025). Researchers highlighted that ChatGPT has the potential to assist learners' L2 writing by providing instant feedback (Lu et al., 2024), enhancing

learners' motivation (Shi et al., 2025), and clarifying the feedback points to support learners' understanding (Su et al., 2023). Also examined is the potential of integrating ChatGPT in writing classrooms (Lu et al., 2024). For instance, Yan (2024) explored ChatGPT's impacts on undergraduates' L2 writing behaviors and reported its potential for assisting writing, such as generating ideas and reducing language errors. Su et al. (2023) found that ChatGPT could use an evaluation rubric to evaluate and provide feedback on students' argumentative writing, which is likely to be a supplement to teacher-written corrective feedback to assess learners' writing. Similarly, Harunasari (2023) also identified potential roles that ChatGPT could serve to facilitate the L2 writing process, such as generating creative ideas, identifying language errors, and proofreading. These studies were optimistic about the potential for generative AI tools to scaffold L2 writing.

A handful of studies investigated the feedback of ChatGPT on learners' written work (e.g., Yan, 2024). Wiboolyasarin et al. (2024) explored the use of ChatGPT feedback in the revision process and found that ChatGPT provided personalized feedback for learners, guiding them to revise their writings and enhancing the depth of their revision. Likewise, Rad et al. (2024) examined the application of ChatGPT in learners' writing process and reported that ChatGPT enhanced their agency in seeking and responding to feedback, promoted feedback engagement, and developed learners' feedback literacy. Banihashem et al. (2024) analyzed the features of peer feedback and ChatGPT feedback on learners' essays, demonstrating that ChatGPT provided more descriptive and overall feedback, while peer feedback was more detailed with specific revision suggestions. They suggest that the combination of ChatGPT and peer feedback could result in better feedback outcomes. Another study by Wang et al. (2024) reported that ChatGPT outperformed other automatic feedback tools (e.g., *Grammarly*) in generating instant and accurate comments on learners' writing.

Also examined are the features of ChatGPT feedback in comparison with teacher feedback and students' perceptions (Guo & Wang, 2023; Lu et al., 2024). For instance, Wei and Li (2023) compared the features of teacher feedback and ChatGPT feedback, revealing that ChatGPT focused on providing suggestions on organization or content, which had the potential to improve learners' overall writing performance. Guo and Wang (2023) conducted a comparative analysis of the quantity and quality of ChatGPT's and teachers' feedback on students' argumentative essays. They found that ChatGPT could provide feedback much more efficiently than teachers and offer a wide range of direct feedback from language and content to the organization. Moreover, it could also support instructors by providing global evaluative feedback on learners' writings and useful suggestions for improvement (Wei & Li, 2023). Similarly, Lu et al. (2024) reported that ChatGPT effectively marked undergraduates' writings; however, ChatGPT provided more general suggestions without clear explanations, whereas teacher comments were more specific. Furthermore, ChatGPT feedback was found to be too generative and incapable of engaging students emotionally. This different finding may be related to the quality of the prompting, which will influence the outcomes. These studies offered valuable insights into the features of ChatGPT and its role as a supplementary feedback practice in English writing instruction. However, there is a noticeable lack of studies concerning learners' uptake of ChatGPT feedback as well as how learners process the feedback. Moreover, it should be noted that these studies were all conducted on university students. Younger students, such as those at schools (K-12), have not yet been studied.

Due to the influence of exam-oriented culture in China, the summative assessment approach is widely used in high schools, where students often receive scores followed by limited feedback on language errors in their writing. As such, it is challenging for these students to improve their English writing performance effectively. The integration of ChatGPT into high school English writing instruction could not only reduce teachers' workload but also enhance learners' motivation to engage in English writing and assist learners in revising their writing (Wei & Li, 2023). Thus, more research is needed to investigate the feasibility of using ChatGPT in high schools.

In addition, most existing studies adopted one writing task in eliciting feedback. Having more than one writing task in one study would be more useful as it would ensure task differences were also considered in evaluating feedback.

Given the above, we believe that more research into ChatGPT feedback is needed to understand its potential for L2 writing instruction and research fully. Research along this line could be extended to multiple aspects of writing performance and processes with L2 writers of different age groups and in different social, educational, and cultural contexts. To join this promising line of research, we systematically compared the written feedback from ChatGPT with that from teachers and closely examined EFL high-school students' responses to the two feedback sources. Specifically, the study answered the following research questions:

RQ$_1$: How does ChatGPT WCF differ from teacher WCF?

RQ$_2$: Are there differences between students' uptake of feedback from the two sources? If so, what are they?

RQ$_3$: Are there differences in students' revisions based on feedback from teachers as opposed to ChatGPT? If so, what are they?

The Study

Participants

The study was conducted in a high school in northwestern China. Four teachers and sixty (31 females and 29 males) Grade 11 students aged 16–17 participated in the study. All student participants were Chinese and had been learning English for eight years, with their overall English proficiency at level 4 (lower-intermediate) according to China's Standards of English Language Ability (CSE). None of them had prior experience with ChatGPT. They were randomly split into Groups A and B, each with 30 students. Based on the English writing scores of the final English examination in the previous semester, there was no significant difference between the two groups (p=0.907), indicating that they were of equivalent English and comparable English writing performance. Participants' demographic information is presented in Table 1.

Four practicing English language teachers from the same school participated in the study. One teacher was the class teacher of the participants and had been teaching them for two years; the other three were similarly qualified and experienced but were teaching other classes. They all held master's degrees in English education or English language literature and had over three years of experience teaching English at the high school level. They were given a copy of the writing samples and gave feedback directly on the scripts. All participants, teachers, and students were informed of their rights to withdraw from the study before signing a consent form.

Research Setting

The present study was conducted in a high school during the second semester of a general English course. The course was aimed at developing learners' English competence, which included listening, speaking, reading, and writing. Students have eight English courses per week. In this semester, students were required to learn how to write argumentative essays. Prior to the semester, all students had learned to write argumentative essays for a semester, and they had a fundamental knowledge of argumentative writing. Normally, the class teacher provided surface-level and meaning-level feedback with summative grades on learners' writings, which was similar as the feedback practice in the present study.

Table 1

Participants' Demographic Information

Group	N	Male	Female	Age	Test score Mean (SD)
Group A	30	14	16	16.92	71.63 (4.04)
Group B	30	15	15	16.95	71.77 (4.70)
Total	60	29	31	16.93	71.70(4.35)

Research Procedures

Based on a pilot study, we adopted a 2x2 Latin-square (counter-balanced) design in conducting the four-week experiment. Specifically, the two groups of students responded to two writing topics and received either feedback from the teacher or ChatGPT in alternating order to balance potential order effects (i.e., whether they received either teacher or ChatGPT feedback first). The counter-balance research design is summarized in Appendix 1.

At the beginning of the study, students were introduced to ChatGPT and briefed on its functionality. Their English teacher demonstrated the application of ChatGPT in English writing in class with the assistance of the researcher. In Week 1, both groups were given Task A: writing a 150–200-word essay describing their views on high school students' Internet usage (see Appendix 2). They were given one day to write the first draft and handed in the drafts to the teacher. Group A received written teacher feedback. Meanwhile, Group B's writing samples were digitized and entered into ChatGPT for feedback; Group B received ChatGPT feedback. Both groups revised their essays based on the feedback and submitted them in Week 2. In Week 3, both groups were given Task B, a similar task to Task A. Specifically, they were asked to describe their views on after-school tutorials (see Appendix 2). Again, they were given one day to write. Group A received feedback from ChatGPT, whereas Group B received feedback from the teacher. Both groups revised their essays and submitted the final work in Week 4.

Reliability Considerations

To ensure reliability in coding the feedback, the researchers adopted the same prompt for ChatGPT to generate feedback for all 120 samples. The prompt is as follows:

> I'd like you to act as a high school English teacher to give feedback on students' essays at the surface and meaning levels. Surface-level feedback refers to feedback on spelling, grammar, word use, mechanics, and sentence structure. You have to select the

surface-level errors, followed by the correct forms. Meaning-level feedback refers to feedback on organization, logic, idea and elaboration, and writing style. You also have to provide revision suggestions on the meaning-level issues.

Appendix 3 presents examples of written feedback generated by ChatGPT using the same prompt.

To ensure that the feedback from teachers and ChatGPT was comparable, all four teachers received similar instructions as those given to ChatGPT and embodied in the prompt. Following the instructions, the teachers provided unfocused and direct feedback on surface-level errors (e.g., grammar, spelling, and mechanics) and indirect and unfocused feedback on meaning-related problems (e.g., content, logic, and coherence) by writing comments to suggest revisions and improvements (Ferris 2006).

However, only the written feedback from one teacher was given to the students in the experiment above; this teacher was labeled as the class teacher thereafter. The class teacher had been the English language teacher of the student participants for two years; differing from the other three teacher participants, who marked the scripts for the present research, he marked the writing samples as part of his teaching duties, as his comments were to return to his students for follow-ups. He, therefore, marked them in accordance with his normal feedback practice and based on his knowledge of the students.

Data Analysis

The students' original and final written samples and the feedback from the teachers and ChatGPT were the data for textual analysis. All written feedback was coded and classified based primarily on Ferris's (2006) taxonomy, with slight modifications. The surface-level feedback was codified as *Spelling*, *Singular-Plural*, *Verb Tense*, *Subject-Verb Agreement*, *Articles*, *Pronouns*, *Collocation*, and *Punctuation*. The meaning-level feedback was codified as *Ideas and Elaboration*, *Organization*, and *Logic* based on the taxonomy proposed by Zhang and Hyland (2018). The total number of feedback and the feedback points per essay were counted and computed to enable subsequent comparison between ChatGPT and the teachers (RQ1). To answer RQ2 regarding students' feedback uptake, the number of feedback given was compared to that of feedback taken up as evidenced by their final drafts.

To answer RQ3 regarding students' revision actions, the revision operations were categorized according to six codes, in line with Sommer (1980) and Chapelle et al. (2015), namely, *No Correction*, *Correction*, *Addition*, *Deletion*, *Substitution*, and *Reorganization*.

To ensure coding reliability, the researchers elaborated the coding scheme with the teachers and coded the data with them initially and later independently. The inter-coder agreement rate for coding revision operations was initially 89.3%. After discussing the disagreements with the teachers, the final inter-coder agreement rates for feedback focus and revision operations were 94.1% and 96.4%, respectively. Meanwhile, the researcher trained a research assistant in the coding scheme; the research assistant then coded all written feedback for the second time independently; the two sets of codes were compared, and the differences were resolved through discussion and revisiting the data. The two rounds of double coding ensured the reliability of the final codes.

Results

Differences between Teacher and ChatGPT-Generated Feedback

Table 2 presents the total feedback points of the class teacher, ChatGPT, and the other three teachers. To our surprise, ChatGPT's feedback resembled the class teacher more than the other three teachers, and the latter provided far more feedback. For instance, in both Task A and B, the mean numbers of surface feedback from ChatGPT and the class teacher were much closer to each other (Class Teacher: 7.4 and 6.87; ChatGPT: 6.55 and 6.48) than the other teachers (T1: 9.48 and 8.23; T2: 10.6 and 8.75; T3: 15.8 and 11.32).

Table 2

The Number of Total Feedback Points Identified in the Two Writing Tasks

Writing tasks	Feedback levels	Class Teacher	ChatGPT	T1	T2	T3
Task A	Surface	**7.40**	**6.55**	9.48	10.6	15.8
	Meaning	1.93	2.57	2.78	4.45	2.75
	sum	9.33	9.12	12.45	15.05	18.55
Task B	Surface	**6.87**	**6.48**	8.23	8.75	11.32
	Meaning	2.08	2.97	2.92	4.32	2.12
	sum	8.95	9.45	11.18	13.07	13.43
Tasks A & B	Surface	**7.13**	**6.52**	8.86	9.68	13.56
	Meaning	2.01	2.78	2.85	4.38	2.43
	sum	**9.14**	**9.29**	11.81	14.06	15.99

Further statistical analysis, as presented in Table 3 (paired t-test comparisons) and Table 4 (correlation analysis among all five feedback providers), also verified the above finding. In Table 3, we found the differences in the number of feedback given were significantly different among all pairs of comparisons (p<.001), except for the pair of ChatGPT and the class teacher (Pair 1 and 5, t=0.842, p=0.201 and 0.402). This finding verified that, first, ChatGPT feedback was similar to the class teacher; second, it differed from the other three teachers; third, the four teachers' feedback differed significantly. Similarly, as Table 4 shows, ChatGPT and the class teacher have the highest positive correlations among all pairs (r=.483, p<.001), whereas the highest correlation among the human teachers is only .404 between T2 and T3.

Table 3

Paired T-Test Results: Comparing the Feedback Points across the Five Raters

		Paired Differences							Significance	
	Comparisons: paired t-tests	Mean	SD	SE	95% Confidence Interval of the Difference		*t*	*df*	One-Sided	Two-Sided
					Lower	Upper			p	p
Pair 1	ChatGPT - Class teacher	0.14	1.84	0.17	-0.19	0.48	0.84	119	0.201	0.402
Pair 2	ChatGPT - T1	-2.35	3.21	0.29	-2.93	-1.77	-8.00	119	<.001	<.001
Pair 3	ChatGPT - T2	-4.76	3.88	0.35	-5.46	-4.06	-13.44	119	<.001	<.001
Pair 4	ChatGPT - T3	-6.75	5.29	0.48	-7.71	-5.79	-13.97	119	<.001	<.001
Pair 5	*Class teacher- ChatGPT*	-0.14	1.84	0.17	-0.48	0.19	-0.84	119	0.201	0.402
Pair 6	*Class teacher - T1*	-2.49	3.12	0.29	-3.05	-1.92	-8.73	119	<.001	<.001
Pair 7	*Class teacher - T2*	-4.90	3.92	0.36	-5.61	-4.19	-13.68	119	<.001	<.001
Pair 8	*Class teacher - T3*	-6.89	5.31	0.48	-7.85	-5.93	-14.23	119	<.001	<.001
Pair 9	T1 - T2	-2.41	4.08	0.37	-3.15	-1.68	-6.48	119	<.001	<.001
Pair 10	T1 - T3	-4.40	4.67	0.43	-5.25	-3.56	-10.33	119	<.001	<.001
Pair 11	T2 - T3	-1.99	4.60	0.42	-2.82	-1.16	-4.75	119	<.001	<.001

Table 4

Correlations among the Codes of Written Feedback among the Five Raters

		Class teacher	ChatGPT	T1	T2	T3
ChatGPT	Pearson *r*	.483**	1			
	Sig. (2-tailed)	*<.001*				
T1	Pearson *r*	.228*	0.123	1		
	Sig. (2-tailed)	*0.012*	*0.181*			
T2	Pearson *r*	-0.1	-0.152	0.133	1	
	Sig. (2-tailed)	*0.277*	*0.097*	*0.147*		
T3	Pearson *r*	-0.068	-0.116	.356**	.404**	1
	Sig. (2-tailed)	*0.457*	*0.206*	*<.001*	*<.001*	

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

Table 5 provides more details by breaking down the feedback into sub-categories. The table shows that both teachers and ChatGPT provided more feedback at the surface level than at the meaning level. This is understandable because surface-level feedback focuses on micro aspects of writing, especially linguistic accuracy, while meaning-level feedback is usually a global and holistic evaluation.

Moreover, all four teachers provided more surface-level feedback (N=856, 1063, 1161, and 1627) than ChatGPT (N=782). The number of meaning-level feedback from ChatGPT (N=333), however, was larger than two teachers (the class teacher and T3: N=241 and 292); therefore, ChatGPT was in the mid-range in providing meaning-level feedback among the assessors.

Regardless of the differences between teachers and ChatGPT, the varied feedback distributions were also related to the participants' fairly low language proficiency; as such, their difficulties were mostly at the morphosyntactic level. At the meaning level, both the teachers and ChatGPT focused on three global qualities of writing, namely, content, organization, and logic. Feedback on content was found to be the most prominent, meaning the students had difficulties expressing their views using examples.

* Corresponding author.
E-mail address: qinxie@um.edu.mo
https://doi.org/10.32038/ltrq.2025.47.07

Table 5

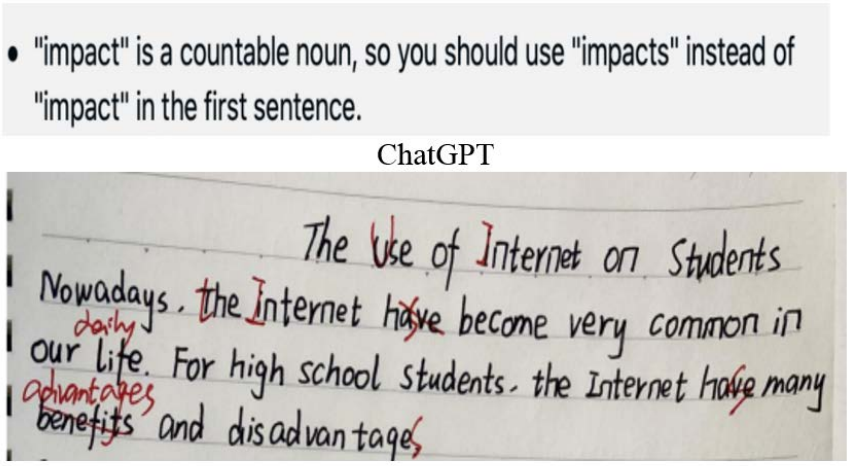
Feedback Points Breakdown: Both Tasks

	Class Teacher	ChatGPT	T1	T2	T3
			Surface-level feedback		
Spelling	262(23.90%)	252(22.60%)	336(23.69%)	425(25.19%)	676(35.23%)
Singular vs Plural	167(15.20%)	159(14.30%)	149(10.51%)	214(12.69%)	173(9.01%)
Verb Tense	102(9.30%)	97(8.70%)	171(12.06%)	163(9.67%)	183(9.54%)
Subject Verb Agreement	46(4.20%)	40(3.60%)	53(3.73%)	78(4.62%)	86(4.80%)
Articles	100(9.10%)	86(7.70%)	92(6.49%)	62(3.68%)	162(8.44%)
Pronouns	60(5.50%)	44(3.90%)	79(5.57%)	59(3.49%)	78(4.06%)
Collocation	45(4.10%)	40(3.60%)	165(11.64%)	81(4.80%)	96(5.00%)
Punctuations	74(6.70%)	64(5.70%)	18(1.27%)	79(4.68%)	173(9.02%)
sum	*856*	*782*	*1063*	*1161*	*1627*
			Meaning-level feedback		
Ideas	104(9.50%)	137(12.29%)	126(8.89%)	197(11.68%)	149(7.76%)
Organization	53(4.80%)	90(8.10%)	156(11.00%)	265(5.71%)	86(4.48%)
Logic	84(7.70%)	106(9.50%)	73(5.15%)	64(3.79%)	57(2.97%)
sum	*241*	*333*	*355*	*526*	*292*
Total	1097	1115	1418	1687	1919

The quality of the feedback was also analyzed and compared. ChatGPT's surface-level feedback not only identified mistakes in the writing but also explained them—a practice that can help learners internalize linguistic knowledge and apply it in subsequent learning. By contrast, the teacher only circled the errors (indirect feedback) or provided the correct forms without explanation. Figure 1 below compares the feedback from ChatGPT and the teacher on one particular error.

Figure 1

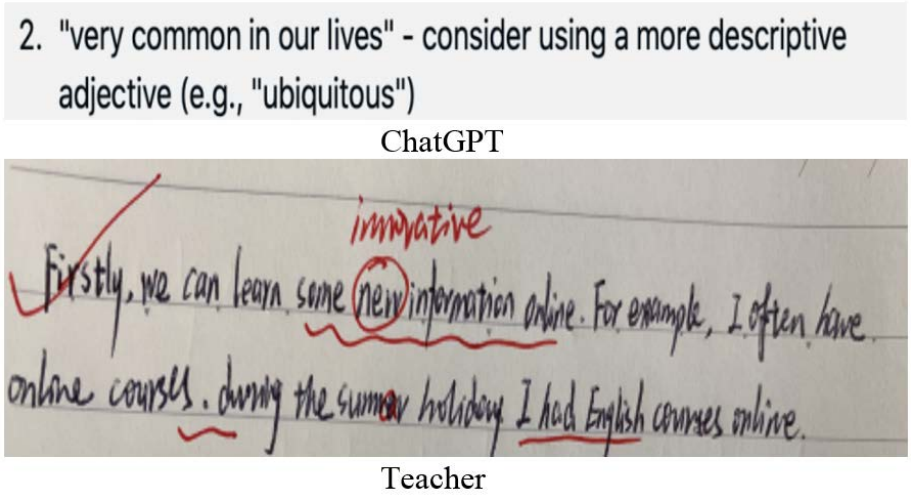
Sample 1



Furthermore, the feedback from ChatGPT used advanced vocabulary and expressions, which could help student writers achieve greater linguistic complexity. However, some of these words were clearly above the participants' language level. For example, the word "ubiquitous" in Figure 2 was new to the participants, who may ignore the feedback. By contrast, the teacher, who had been teaching the participants for two years, had a better understanding of their language level. Accordingly, the vocabulary and expressions he adopted pitched at this level.

Figure 2

Sample 2

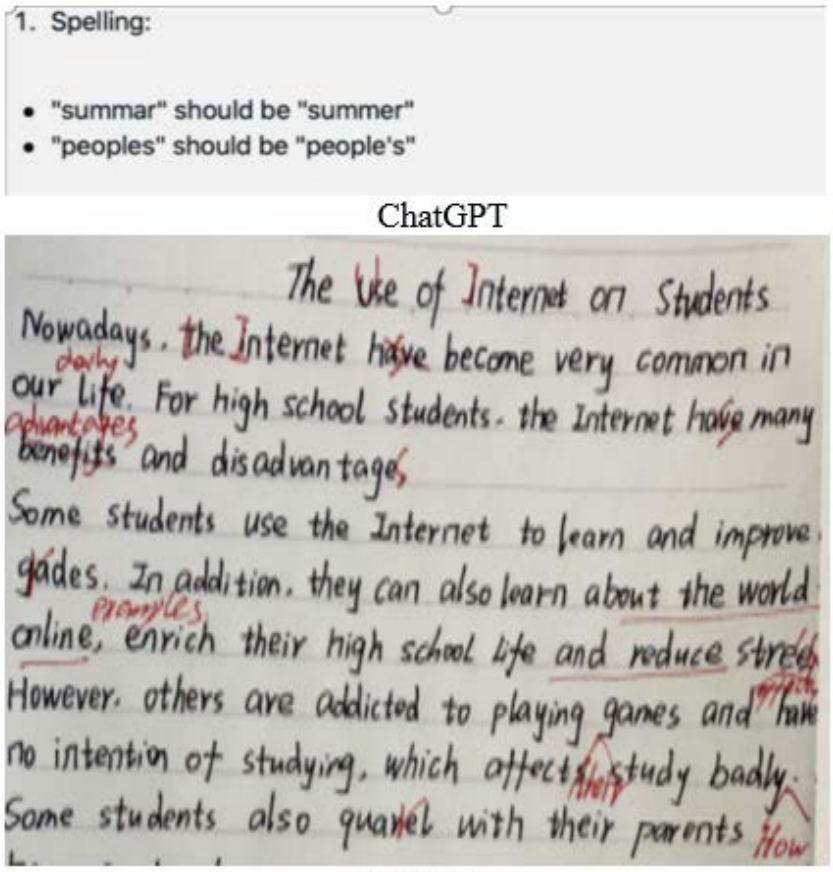


* Corresponding author.
E-mail address: qinxie@um.edu.mo
https://doi.org/10.32038/ltrq.2025.47.07

Regarding the presentation of their feedback, ChatGPT listed the errors it identified one by one, followed by the correct forms. As such, the errors were not embedded in their co-texts (Figure 3). Teachers, on the other hand, circled and corrected the errors directly in the script, making it effortless for learners to see and revise the errors within the text context.

Figure 3

Sample 3

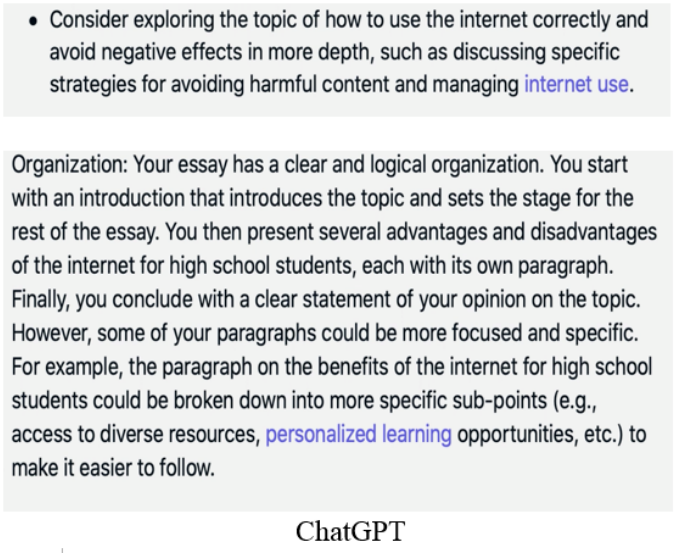


Where meaning-level feedback was concerned, ChatGPT not only provided a significant amount of personalized feedback on the content, organization, and logic, it also gave explanations and examples of how to improve the quality of the writing. Figure 4 shows a ChatGPT feedback item focused on content and organization that offers detailed instructions on discussing one's views in depth, which could guide learners to improve the quality of their writing. The teacher, on the other hand, provided generic revision suggestions. For instance, he wrote, "Pay attention to the logic of your essays; please give more examples. You have to use some complex sentences." Arguably, such feedback is less useful to low-proficiency students.

Furthermore, ChatGPT provided positive, negative, and constructive feedback. In Figure 5, ChatGPT first acknowledges the merits of the essay. In offering suggestions, it adopts hedge words such as *could*, *may*, and *would* to soften the tone. The teacher, in contrast, preferred to draw a line or a checkmark or to write a short comment, e.g., "good", on the well-written part. He also wrote a couple of sentences to suggest how to revise the writing.

Figure 4

Sample 4



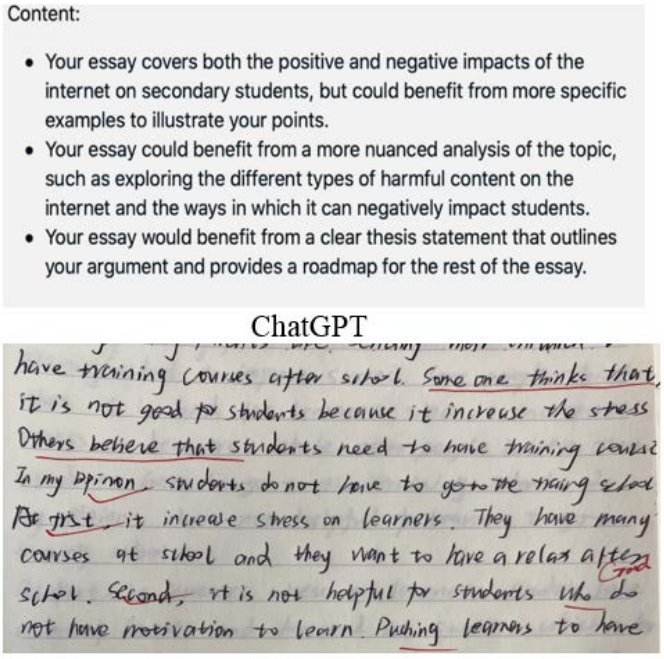
- Consider exploring the topic of how to use the internet correctly and avoid negative effects in more depth, such as discussing specific strategies for avoiding harmful content and managing internet use.

Organization: Your essay has a clear and logical organization. You start with an introduction that introduces the topic and sets the stage for the rest of the essay. You then present several advantages and disadvantages of the internet for high school students, each with its own paragraph. Finally, you conclude with a clear statement of your opinion on the topic. However, some of your paragraphs could be more focused and specific. For example, the paragraph on the benefits of the internet for high school students could be broken down into more specific sub-points (e.g., access to diverse resources, personalized learning opportunities, etc.) to make it easier to follow.

ChatGPT

Figure 5

Sample 5



Content:

- Your essay covers both the positive and negative impacts of the internet on secondary students, but could benefit from more specific examples to illustrate your points.
- Your essay could benefit from a more nuanced analysis of the topic, such as exploring the different types of harmful content on the internet and the ways in which it can negatively impact students.
- Your essay would benefit from a clear thesis statement that outlines your argument and provides a roadmap for the rest of the essay.

ChatGPT

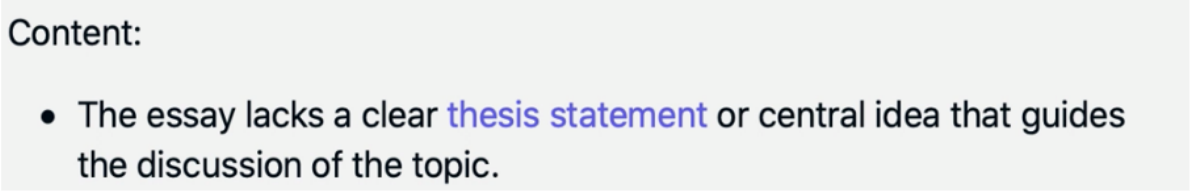
have training courses after school. Some one thinks that, it is not good for students because it increase the stress. Others believe that students need to have training course. In my opinion, students do not have to go to the traing school. At first, it increase stress on learners. They have many courses at school and they want to have a relax after school. Second, it is not helpful for students who do not have motivation to learn. Pushing learners to have

Teacher

Nevertheless, the participants found some of ChatGPT's feedback difficult to understand. In addition to some of the advanced words it used, there were academic terms, such as "thesis statement" (Figure 6), that were unfamiliar to the participants. The teacher's feedback, on the other hand, catered to the participants' language ability and was based on the local curriculum's English writing requirements.

Figure 6

Sample 6

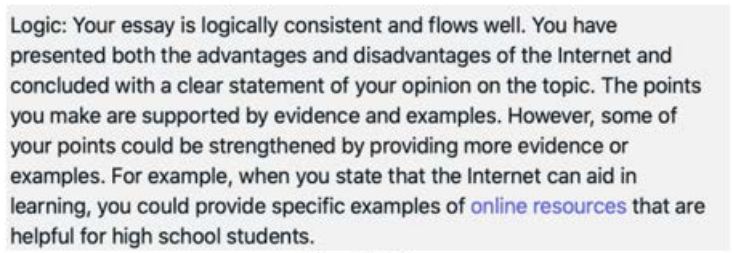


ChatGPT

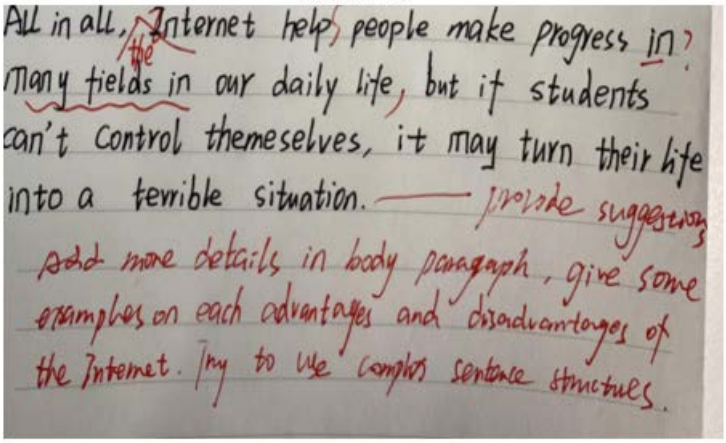
Lastly, ChatGPT feedback was personalized but long (i.e., up to six sentences), especially on meaning-level issues, whereas the teacher's feedback tended to be much shorter, expressed in only two or three sentences. In the example in Figure 7, the teacher wrote: 'Add more details in the body paragraph, give some examples on each advantage and disadvantage of the Internet; Try to use complex sentence structures."

Figure 7

Sample 7



ChatGPT



Teacher

In summary, both ChatGPT and the teacher provided varied surface- and meaning-level feedback. ChatGPT provided personalized and comprehensive feedback with detailed explanations, but some of its feedback used unfamiliar terms or sophisticated words that were beyond the participants' abilities at the time. By contrast, the teacher provided accurate, direct, and easy-to-locate feedback; this practice may contribute to students' utilization of the feedback. That said, some of the teacher's meaning-level feedback was vague, which could also impede its uptake and consequent revision.

Differences between Students' Uptake of Feedback from Teacher and ChatGPT

Table 6 presents the quantity of feedback points, their uptake, and the ratios of the two. Overall, teacher feedback had a higher uptake rate than ChatGPT feedback (78.10% vs. 70.41%). Specifically, 856 of the 1,097 teacher feedback items were addressed in the second submission, whereas only 740 of the 1,051 ChatGPT feedback items were addressed. A closer look at the uptake statistics found that surface-level feedback had far higher uptake rates than meaning-level feedback, a finding applicable to both feedback sources. The uptake rate of teacher-generated surface-level feedback was slightly higher (Teacher: 91.80% vs. ChatGPT: 86.17%), and those of ChatGPT-generated meaning-level feedback was slightly higher (Teacher: 29.10% vs. ChatGPT: 32.03%).

Table 6

Feedback Uptake Rates

	ChatGPT Feedback			Teacher Feedback		
Feedback type	Quantity	Uptake	Uptake Rate	Quantity	Uptake	Uptake Rate
Surface- level	745	642	86.17%	856	786	91.80%
Meaning- level	306	98	32.03%	241	70	29.10%
Total	1,051	740	70.41%	1,097	856	78.10%

Differences between Students' Revisions based on Teacher ChatGPT-Generated Feedback

Overall, students' revision strategies based on teacher and ChatGPT-generated feedback were rather similar. Table 7 shows that students correctly revised 66.60% and 61.50% of the surface-level feedback from the teacher and ChatGPT, respectively. Apart from *correction* (Teacher: 66.60%; ChatGPT: 61.50%), they also adopted *addition* (Teacher: 7.90%; ChatGPT: 7.30%), *substitution* (Teacher: 7.20%; ChatGPT: 9.20%), *deletion* (Teacher: 6.60%; ChatGPT: 6.30%) and *reorganization* (Teacher: 5.30%; ChatGPT: 5.20%) in revising their essays. This reveals that learners could apply a set of revision strategies to use surface-level feedback to modify their essays.

Teacher feedback resulted in higher revision rates than ChatGPT feedback for all strategies except substitution, and correction was the most heavily employed strategy for revision, regardless of the feedback source. Unlike the students' revision operations on surface-level feedback, however, *no correction* was the most frequent response to meaning-level feedback (Table 8), which surprised the researchers. The revision rates tended to be similar for the two feedback sources, with teacher feedback resulting in slightly higher rates of *no correction* (Teacher: 62.40%; ChatGPT: 60.80%), *deletion* (Teacher:5.70%; ChatGPT: 4.10%), and *reorganization* (Teacher: 9.60%; ChatGPT: 8.60%), but lower rates of *addition* (Teacher:13.10%; ChatGPT: 15.30%) and *substitution* (Teacher: 9.20%; ChatGPT: 11.20%).

Table 7

Percentages of Students' Revision Actions on Surface-Level Feedback

	Class Teacher	ChatGPT
Correction	632 (66.60%)	549 (61.50%)
No correction	61 (6.40%)	93 (10.50%)
Addition	75 (7.90%)	65 (7.30%)
Substitution	68(7.20%)	82 (9.20%)
Deletion	63 (6.60%)	57 (6.30%)
Reorganization	50 (5.30%)	46 (5.20%)
Overall	949 (100%)	892 (100%)

Table 8

Percentages of Students' Revision Actions on Meaning-Level Feedback

	Class Teacher	ChatGPT
Correction	0	0
No correction	143 (62.40%)	163 (60.80%)
Addition	30 (13.10%)	41 (15.30%)
Substitution	21 (9.20%)	30 (11.20%)
Deletion	13 (5.70%)	11 (4.10%)
Reorganization	22 (9.60%)	23 (8.60%)
Overall	229 (100%)	268(100%)

Furthermore, we examined the learners' revision actions, examples of which are presented in Appendix 4. The learners frequently used correction in response to surface-level feedback. This is probably because both the teacher and ChatGPT provided the correct forms for spelling, grammar, and punctuation errors, which made it easy for learners to accept the feedback and use the new forms in their final drafts.

No correction indicates that learners ignored the feedback, i.e., made no revisions based on it. This may be attributed to their disagreement with the feedback or to their limited linguistic knowledge. For example, one participant did not replace "improve" with "facilitate" as suggested probably because S/he did not understand the suggested word. Students also ignored general or vague feedback, such as the following comment from the teacher: "Revise this sentence with a more complex structure." In this case, the student may not have known which "complex structure" to be adopted.

Addition and substitution were less frequently employed. Both the teacher and ChatGPT directly provided new or alternative words for students to use in enhancing the linguistic complexity of their writing, which should have been easy to follow.

No correction was the most frequently used revision operation in response to meaning-level feedback. In the example below, ChatGPT provides long and comprehensive feedback with three suggestions for addition. However, the student made no revisions in response to these suggestions. This may be because s/he did not know how to incorporate the suggestions or was not sufficiently motivated to do so.

> In the body paragraphs, provide specific examples or evidence to support your points about the advantages and disadvantages of the internet. This will help strengthen your arguments and make your essay more persuasive. Consider using transition words and phrases to improve the flow between sentences and paragraphs. In your conclusion, you can provide some specific examples or suggestions on how to make reasonable use

> of the internet, which will enhance the clarity and impact of your final thoughts. (ChatGPT)

Some students adopted the *reorganization strategy* in revising. Generally, they reorganized the structure of their essays or reordered their sentences based on suggestions such as "Your arguments on this topic were not placed in a logical way. You should reorder them" (Teacher). The following extracts are taken from a student's first and revised drafts; the logic of the revised one is improved.

First draft:

> Firstly, the training school provides new knowledge that can broaden our horizons. Then, the training school can help us learn information that we have not understand at school. Next, the training school enhances our understanding of the knowledge and helps us how to use the knowledge to do exercise.

Revised draft:

> Firstly, the training school can help us learn information that we have not understood at school. Secondly, the training school also enhances our understanding of the knowledge and helps us how to use the knowledge to do exercise. Finally, the training school provides new knowledge that can broaden our horizons.

In summary, learners adopted various strategies to revise their essays. Their revision operations were influenced by the feedback they received as well as their language ability and willingness to revise.

Discussion

Differences between the WCF from Teacher- and ChatGPT

This study found that teacher- and ChatGPT-generated WCF had distinctive features. ChatGPT can be prompted to generate personalized feedback (Barrot, 2023). In this study, ChatGPT was asked to review EFL students' essays and provide both surface- and meaning-level feedback. Due to the participants' limited language proficiency, their essays showed numerous problems. ChatGPT pointed out these weaknesses and provided many personalized revision suggestions; it outperformed the teacher, especially in providing meaning-level feedback on content. Nevertheless, the content of ChatGPT feedback has to be qualified as ChatGPT cannot have prior knowledge of students' writing abilities to tailor its feedback according to individual student needs. As such, some of the feedback points from ChatGPT are not suitable for students, and they may find them difficult to understand.

At the surface or linguistic level, ChatGPT feedback was comprehensive and comparable to the feedback of the class teacher, though not as thorough as the other teachers. In line with earlier research on AWE feedback (Tian & Zhou, 2020), ChatGPT, an AI-powered chatbot, proved to have significant potential for evaluating the strengths and weaknesses of students' writing and suggesting revisions supported by examples and explanations. Unlike previous

studies (e.g., Tian & Zhou, 2020), wherein AWE systems provided limited and even no feedback on content and organization, this study found that ChatGPT was able to offer a large number of suggestions on meaning-level issues such as content, logic, coherence, and organization. As a matter of fact, it offered more meaning-level feedback than the class teacher did.

Consistent with prior research, such as Mao and Crosthwaite (2019) and Thi and Nikolov (2022), the present study found that the teachers offered far more feedback on language-related errors. The focus of teacher feedback may be influenced by various factors, such as feedback beliefs, time restrictions, and individual factors pertaining to the learners (Zhang, 2020). In this study, the teacher preferred to give direct and comprehensive feedback on surface-level issues such as spelling, vocabulary, grammar, verb tense, and punctuation. Direct feedback on language errors could be more appropriate for students with low language ability, such as the participants in this study, and could scaffold them to improve the accuracy of their writing (Zhang, 2020).

Overall, this study revealed that ChatGPT could serve as a valuable writing assistant to low-proficiency language learners and their teachers. Future research may consider comparing the characteristics of teacher- and ChatGPT-generated feedback on the writing samples of learners at other language proficiency levels to explore whether this study's observations still hold.

Learners' Uptake of and Revision Actions

This study also investigated how EFL learners responded to WCF. Overall, teacher feedback received higher uptake rates than ChatGPT feedback (78.1% vs. 69.6%), but the differences were small. Compared to studies of other widely used AWE systems (e.g., Criterion, Pigai, and Grammarly), the present study found that learners incorporated much more ChatGPT feedback. For instance, in Bai and Hu's (2017) study, learners' uptake of feedback from Pigai was only 11.5%. In Chapelle et al. (2015), learners ignored half of Criterion's feedback. The present study's high uptake rates of ChatGPT feedback indicated the high accuracy of ChatGPT's WCF, which was direct, detailed, and comprehensive. The close uptake rates of ChatGPT- and teacher-generated feedback show that ChatGPT can be adopted as an alternative feedback source to reduce teachers' marking and feedback workload and assist EFL writers as a writing companion, providing instant feedback along the way of writing.

Trust and Credibility of Feedback Sources

The finding regarding the higher uptake rate of teacher feedback (than ChatGPT) echoes previous research findings (e.g., Link et al., 2022; Tian & Zhou, 2020). Several factors have been found to contribute to this preference. Firstly, learners' trust in teachers may play a role (Yang & Carless, 2013). Learners have been found to respect teachers' authority and regard their feedback as credible (Cheng & Liu, 2022) and thus trust the value of their feedback in improving their writing (Yang & Carless, 2013).

By contrast, our participants may place less trust in AI or machines as they had no prior experience with AI. AI or machines had less authority or power over them. As such, they tended to ignore some of ChatGPT's feedback. A trusting atmosphere and a good relationship between the feedback provider and recipients are crucial to promoting learners' agency in the

feedback process and their engagement with the feedback (Cheng & Liu, 2022). Because it was only the first time that the participants used ChatGPT, we believe that with time and more practice, they could place more trust in AI programs' feedback.

Feedback Language vs. Learners' Language Ability

The present study found that the teacher could better cater to the learners' language ability and offer specific and context-sensitive suggestions that fit the local syllabus, whereas ChatGPT's feedback contained vocabulary that may have been unfamiliar to the students and beyond the scope of the curriculum. The teacher had a better understanding of the learners' writing proficiency levels and provided suggestions that were more appropriate to them achieving the target performance required by the local curriculum. His feedback used simple words better suited to low-proficiency high school students. By contrast, ChatGPT had no prior knowledge of the participants' writing abilities or the syllabus they followed. Thus, the language used in its feedback may have exceeded their abilities, and its suggestions may have exceeded the curriculum requirements. Therefore, future use of ChatGPT as a teaching assistant requires further training with local examples and prompt refinement to cater to low-proficiency EFL learners and their local curricula.

Feedback Focus, Form, and Amount vs. Learners' Zone of Proximal Development (ZPD)

Also noteworthy are the substantially high uptake rates of surface-level feedback and low uptake rates of meaning-level feedback. This finding aligns with previous feedback studies such as Cheng & Liu (2022) and Zhang (2020). Several interplaying factors contributed to this result. On the one hand, there are factors associated with the feedback itself, i.e., its focus (surface vs. meaning; local vs. global) and form (direct vs. indirect). On the other hand, there are factors associated with the learners, i.e., language ability and zone of proximal development (ZPD), which are essential in expanding learners' cognitive capacity, enhancing their maturity, and promoting learning motivation.

In this study, the feedback from the teacher and ChatGPT on surface-level issues was direct (i.e., the errors were pointed out directly, and the correct forms were given). In contrast, the feedback on meaning-level issues was indirect and more demanding for learners to process. Moreover, surface-level feedback tended to be at the word, phrase, or clausal level (local), whereas meaning-level feedback tended to be at the paragraph or discourse level (global). It has been suggested that direct and local feedback provides explicit information that helps low-proficiency learners notice and correct their errors with ease (Storch & Wigglesworth, 2010). Conversely, indirect and global feedback is more appropriate for advanced learners, as its comprehension requires high-order abilities and substantially more effort to address global, discourse-level problems (Bitchener & Ferris, 2012).

To help students process indirect meaning-level feedback targeting global problems, examples and detailed explanations are necessary. In contrast to Link et al. (2022), the present study found that EFL writers incorporated 30.30% of ChatGPT's meaning-level feedback and employed strategies to revise their essays with such feedback. This may be because of the examples and explanations in ChatGPT's meaning-level feedback, which could have enabled learners to follow the suggestions. By contrast, most of the teacher's meaning-level feedback

was brief, vague, generic, and abstract, which may have made it difficult for the students to act upon the feedback.

While teachers often do not have the time to provide students with detailed examples, explanations, or personalized suggestions, ChatGPT clearly can do this. Moreover, this study's findings indicate that learners can use the meaning-level feedback from ChatGPT to revise their writing. Therefore, the researchers recommend that teachers harness the power of ChatGPT to generate meaning-level feedback for their students, particularly in large classrooms. That said, they must be strategic and consider learners' language ability (Zhang, 2020) and ZPD.

According to previous feedback research (e.g., Lee, 2019; Han & Hyland, 2015), learners are likely to value, accept, and internalize the feedback when it targets the areas that were taught in the class or were within their existing knowledge system; in other words, feedback that matches their ZPD. Participants in the present study were more familiar with surface-level feedback than extensive meaning-level feedback. The latter may have been beyond their ZPD.

Similarly, while the participants demonstrated the capacity for revising their writing based on both teacher and ChatGPT-generated feedback using revision operations such as *correction*, *addition*, *substitution*, and *reorganization*, the study found that consistent with Zhang (2020) and Koltovskaia (2020), these revisions remained largely at the surface level, and there were far fewer deep revision operations (e.g., *reorganization* and *substitution*). The latter is necessary for the more global, meaning-level problems to modify the writings in a well-organized way.

In this regard, educators must be careful with their feedback's quantity and demands. ChatGPT could be a helpful assistant in L2 writing instruction and learning. However, similarly to comprehensive and detailed teacher feedback (see Lee 2019), the feedback from ChatGPT could be too overwhelming (in quantity) for the learners to take up all at once. Therefore, educators should be strategic in limiting the amount and the demands of the feedback given to the learners in each assignment, taking into careful consideration the learners' ZPD and the curriculum requirements.

This is especially important for low-ability learners. Rather than being released all at once, the feedback, be it teacher or ChatGPT-generated, should be given in multiple batches and planned carefully in accordance with writing pedagogy. For instance, the meaning-level feedback could be given before the surface-level feedback on mechanics and linguistic features. Giving meaning-level feedback on the first draft can help the learners secure sufficient and logically organized content (ideas, examples, and details) within a good structure. The surface-level feedback can be given later (i.e., to a second or final draft) and aims to help the learners edit and polish their language and finalize the writing.

Conclusion

The present study is significant in that it is one of the few empirical studies that systematically compared ChatGPT-generated feedback with teacher feedback as well as students' responses to each feedback source, thereby providing insights for utilizing generative AI (ChatGPT) as an alternative or supplementary source of feedback. Integrating generative AI into L2 writing instruction could resolve a long-standing issue in second-language writing instruction and enable personalized writing instruction and feedback.

The study has several limitations that future research could address. First, the research results should be interpreted cautiously since this study involved only low language proficiency EFL Chinese students at the high school level. Future research on other student populations differing in cognitive maturity and language proficiency could be conducted, as these features could affect how students respond to the feedback. Furthermore, the experiment only lasted four weeks, and the participants only worked on two writing tasks; the longitudinal and accumulative effects of ChatGPT feedback were not examined. In addition, the study only involved four teachers whose feedback showed considerable variations, which, at times, exceeded the differences between teachers and ChatGPT. This finding, on the one hand, indicates the sensitivity of teacher feedback to contextual factors; on the other hand, it shows the differences between ChatGPT and teachers are, at times, not as high as those among teachers. In other words, with training, ChatGPT could be a more reliable teaching assistant, faithfully mimicking a teacher's feedback style and providing students with consistent feedback. That said, future research should involve more teachers to understand the differences between teachers as a group on one side and generative AI on the other.

ORCID

https://orcid.org/0009-0003-0290-2657

https://orcid.org/0000-0001-8528-6092

Acknowledgements

Not applicable.

Funding

This research was partially supported by a research grant given to the second and corresponding author by the Research Grants Council of the Hong Kong SAR Government, China (Project # 18600121), when she worked at the Education University of Hong Kong.

Ethics Declarations

Competing Interests

No, there are no conflicting interests.

References

Amer, M. I. B., Rababah, L. M., & Rababah, M. A. (2025). Exploring ChatGPT as a tool for thesis writing: Perspectives of EFL supervisors in Jordanian universities. *Journal of Language Teaching and Research, 16(1)*, 33–42. https://doi.org/10.17507/jltr.1601.04

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

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

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*, Article 100745. https://doi.org/10.1016/j.asw.2023.100745

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing. Routledge.*

Chapelle, C. A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, *32*(3), 385–405. https://doi.org/10.1177/0265532214565386

Cheng, X., & Liu, Y. (2022). Student engagement with teacher-written feedback: Insights from low-proficiency and high-proficiency L2 learners. *System*, *109*, Article 102880. https://doi.org/10.1016/j.system.2022.102880

Cotos, E., Huffman, S., & Link, S. (2015). Furthering and applying move/step constructs: Technology-driven marshalling of Swalesian genre theory for EAP pedagogy. *Journal of English for Academic Purposes*, *19*, 52–72. https://doi.org/10.1016/j.jeap.2015.05.004

Dikli, S., & Bleyle, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, *22*, 1–17. https://doi.org/10.1016/j.asw.2014.03.006

Ellis, R. (2010). EPILOGUE: A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*, *32*(2), 335–349. https://doi.org/10.1017/S0272263109990544

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland, & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81–104). Cambridge University Press.

Guo, Q., Feng, R., & Hua,Y. (2021). How effectively can EFL students use automated written corrective feedback in research writing? *Computer Assisted Language Learning*, *35*(9). 2312–2331. https://doi.org/10.1080/09588221.2021.1879161

Guo, K., Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education Information Technology*, *29*, 8435–846. https://doi.org/10.1007/s10639-023-12146-0

Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, *30*(4), 31–44. https://doi.org/10.1016/j.jslw.2015.08.002

Harunasari, S. Y. (2023). Examining the effectiveness of AI-integrated approach in EFL writing: A case of ChatGPT. *International Journal of Progressive Sciences and Technology*, *39*(2), 357-368. https://doi.org/10.52155/ijpsat.v39.2.5516

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83–101. https://doi.org/10.1017/S026144480600339

Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal*, *99*(1), 1–18. https://doi.org/10.1111/modl.12189

Karatay, Y., & Karatay, L. (2024). Automated writing evaluation use in second language classrooms: A research synthesis. *System*, *123*, Article 103332. https://doi.org/10.1016/j.system.2024.103332

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

Lavolette, E., Polio, C. & Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language Learning & Technology*, *19*(2), 50–68. http://dx.doi.org/10125/44417

Lee, I. (2019). Teacher written corrective feedback: Less is more. *Language Teaching*, *52*(4), 524–536. https://doi.org/10.1017/S0261444819000247

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation feedback in ESL writing instruction. *Journal of Second Language Writing*, *27*, 1–18. https://doi.org/10.1016/j.jslw.2014.10.004

Li, M. (2024). Leveraging ChatGPT for second language writing feedback and assessment. *International Journal of Computer-Assisted Language Learning and Teaching*, *14*(1), 1–11. https://doi.org/10.4018/IJCALLT.360382

Lu, Q., Yao, Y., Xiao, L., Yuan, M., Wang, J., & Zhu, X. (2024). Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing? *Assessment and Evaluation in Higher Education*, *49*(5), 616–633. https://doi.org/10.1080/02602938.2024.2301722

Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, *35*(4), 605–634. https://doi.org/10.1080/09588221.2020.1743323

Mao, S. S., & Crosthwaite, P. (2019). Investigating written corrective feedback: (Mis)alignment of teachers' beliefs and practice. *Journal of Second Language Writing*, *45*, 46–60. https://doi.org/10.1016/j.jslw.2019.05.004

O' Neill, R., & Russell, A. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, *35*(1), 42–56. https://doi.org/10.14742/ajet.3795

OpenAI. (2024). *ChatGPT: Optimizing language models for dialogue*. https://chatgpt.r4wand.eu.org/

Rahimi, M. (2021). A comparative study of the impact of focused vs. comprehensive corrective feedback and revision on ESL learners' writing accuracy and quality. *Language Teaching Research*, *25*(5), 687–710. https://doi.org/10.1177/1362168819879182

Rad, M. R., & Davis, J. L. (2024). Exploring student perceptions of learning experience in fundamental mechanics courses enhanced by ChatGPT. *2024 ASEE Annual Conference & Exposition*. https://doi.org/10.18260/1-2--47426

Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning*, *31*(7). 653-674. https://doi.org/10.1080/09588221.2018.1428994

Rock, J. L. (2007). The impact of short-term use of Criterions on writing skills in Ninth grade. *ETS Research Report Series*, *1*, 1–24. https://doi.org/10.1002/j.2333-8504.2007.tb02049.x

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, *41*(2), 255–283. https://doi.org/10.1002/j.15457249.2007.tb00059.x

Shi, H., Chai, C. S., Zhou, S., & Aubrey, S. (2025). Comparing the effects of ChatGPT and automated writing evaluation on students' writing and ideal L2 writing self. *Computer Assisted Language Learning*, 1–28. https://doi.org/10.1080/09588221.2025.2454541

Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing: Case studies. *Studies in Second Language Acquisition*, *32*(2), 303–334. https://doi.org/10.1017/S0272263109990532

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, *57*, Article 100752. https://doi.org/10.1016/j.asw.2023.100752

Suzuki, W., Nassaji, H., & Sato, K. (2019). The effects of feedback explicitness and type of target structure on accuracy in revision and new pieces of writing. *System*, *81*, 135–145. https://doi.org/10.1016/j.system.2018.12.017

Tian, L., & Zhou, Y. (2020). Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System*, *91*, Article 102247. https://doi.org/10.1016/j.system.2020.102247

Thi, N. K., & Nikolov, M. (2022). How Teacher and Grammarly Feedback Complement One Another in Myanmar EFL Students' Writing. *The Asia-Pacific Education Researcher*, *31*(6), 767–779. https://doi.org/10.1007/s40299-021-00625-2

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the Effectiveness of Comprehensive Error Correction in Second Language Writing. *Language Learning*, *62*(1), 1–41. https://doi.org/10.1111/j.1467-9922.2011.00674.x

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*, Article 101440. https://doi.org/10.1016/j.tsc.2023.101440

Ware, P. (2011). Computer-generated feedback on student writing. *TESOL Quarterly*, *45*(4), 769–774. https://doi.org/10.5054/tq.2011.272525

Wei, S., & Li, L. (2023). Artificial intelligence-assisted second language writing feedback: A case study of ChatGPT. *Foreign Languages in China*, *20*(3), 33–40. https://doi.org/10.13564/j.cnki.issn.16729382.2023.03.007.

Wiboolyasarin, W., Wiboolyasarin, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. *Computers and Education: Artificial Intelligence*, *6*, Article 100228. https://doi.org/10.1016/j.caeai.2024.100228

Yan, D. (2024). Feedback seeking abilities of L2 writers using ChatGPT: a mixed method multiple case study. *Kybernetes*. https://doi.org/10.1108/K-09-2023-1933

Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education*, *18*(3), 285–297. https://doi.org/10.1080/13562517.2012.719154

Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching*, *49*, 461–493. https://doi.org/10.1017/S0261444816000161

Zhang, Z. (2020). Engaging with automated writing evaluation feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, *43*, 78–91. https://doi.org/10.1016/j.asw.2019.100439

Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, *51*, Article 100586. https://doi.org/10.1016/j.asw.2021.100586