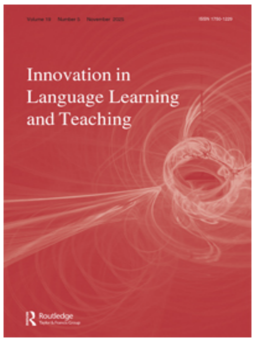

Innovation in Language Learning and Teaching

ISSN: 1750-1229 (Print) 1750-1237 (Online) Journal homepage: www.tandfonline.com/journals/rill20

Exploring the use of ChatGPT-generated model texts as a feedback instrument: EFL students' text quality and perceptions

Dunming (Jason) Lu & Yutian Zeng

To cite this article: Dunming (Jason) Lu & Yutian Zeng (03 Jul 2025): Exploring the use of ChatGPT-generated model texts as a feedback instrument: EFL students' text quality and perceptions, Innovation in Language Learning and Teaching, DOI: 10.1080/17501229.2025.2525341

To link to this article: https://doi.org/10.1080/17501229.2025.2525341

Published online: 03 Jul 2025.

INNOVATION IN LANGUAGE LEARNING AND TEACHING
https://doi.org/10.1080/17501229.2025.2525341

Exploring the use of ChatGPT-generated model texts as a feedback instrument: EFL students' text quality and perceptions

Dunming (Jason) Lu [a] and Yutian Zeng [b]

[a]Department of English, City University of Hong Kong, Kowloon, Hong Kong, People's Republic of China; [b]Department of English and Communication, The Hong Kong Polytechnic University, Kowloon, Hong Kong, People's Republic of China

ABSTRACT

Purpose: This study investigates the usefulness of ChatGPT-generated model texts as a feedback instrument. Specifically, it focuses on the effects of ChatGPT-generated model texts on EFL students' text quality in terms of content, organisation, vocabulary, and grammar. Students' perceptions are also examined to gauge a more comprehensive understanding from a subjective perspective.

Methodology: The study adopted a pretest and post-test quasi-experimental design with three groups: one control group (CG) without any form of feedback, one experimental group (EG1) with a model text written by human teachers, and one experimental group (EG2) receiving a model text generated by ChatGPT. Data were collected from 104 Chinese high school EFL students. Students' writing tests were analysed using descriptive and inferential statistics, while perception questionnaires were analysed using descriptive statistics and content analysis.

Findings: The results showed that while there was no significant difference between the three groups at the outset, both experimental groups demonstrated better performance than the control group in all the analytical measures in the post-test with large effect sizes. In addition, the improvements made by EG1 and EG2 were similar. Students generally held positive perceptions of using ChatGPT-generated model texts as a feedback tool, although they expressed concerns about the potential lack of contextual relevance and stylistic unnaturalness of the AI-generated models.

Originality/value: This study contributes to the growing body of research on AI-assisted feedback and offers implications for integrating emerging technology, such as ChatGPT, into the use of model texts as a feedback instrument in EFL writing classrooms.

ARTICLE HISTORY
Received 14 January 2025
Accepted 22 June 2025

KEYWORDS
Models; feedback; ChatGPT; L2 writing; AI-assisted language learning

CONTACT Dunming (Jason) Lu dunminglu2-c@my.cityu.edu.hk Department of English, City University of Hong Kong, 83, Tat Chee Avenue, Kowloon, Hong Kong, People's Republic of China

1. Introduction

Recent advancements in artificial intelligence (AI) have given rise to improved digital tools that can support feedback provision, such as automated written evaluation (AWE) systems like Pigai, tools that can provide automated written corrective feedback (AWCF) like Grammarly (Godwin-Jones 2022), and generative AI (GenAI) models like ChatGPT (Generative Pre-trained Transformer). Among these tools, ChatGPT has garnered significant scholarly attention, with an emerging body of research investigating multiple aspects of ChatGPT-generated feedback, such as its quality compared to human feedback (e.g. Guo and Wang 2024; Steiss et al. 2024), its effectiveness on students' writing quality (e.g. Tsai, Lin, and Brown 2024), and students' perceptions (Escalante, Pack, and Barrett 2023). One commonality of these studies is that they primarily centre on ChatGPT's potential to provide written corrective feedback (WCF), which involves correcting writers' errors (Bitchener and Storch 2016). However, a notable feature that distinguishes ChatGPT from other AI-powered feedback tools is its ability to produce coherent and grammatically correct human-like text regarding different genres, difficulty levels, lengths, and topics. Such a capability renders ChatGPT a valuable tool for supplying model texts of a writing task, a more discursive form of feedback that has attracted increasing scholarly interest recently (Kang 2020).

Unlike the ChatGPT-generated WCF that mainly offers negative evidence feedback addressing the micro-level (e.g. vocabulary, grammar, spelling) features in writing, using ChatGPT-generated model texts as a feedback instrument (MTFI) can provide learners with a variety of positive evidence feedback (PEF) in relation to content, vocabulary, and organisational structures. It is believed that learners can notice such information when they compare their writing to the model texts. Consequently, they incorporate some features into their subsequent drafts, leading to enhanced writing quality (Kang 2024). As such, ChatGPT-generated model texts can serve as an effective alternative to ChatGPT's WCF, which can be implemented in English-as-a-foreign (EFL) writing classrooms to improve students' writing performance.

Despite the potential of ChatGPT-generated model texts in developing students' text quality, few studies have been conducted to validate its role as a feedback instrument. Some scholars have proposed using ChatGPT to create model texts to support students' revision (Jiang and Hyland 2025; Nguyen, Dao, and Nguyen 2024). However, these proposals remain largely theoretical. To the best of our knowledge, there is no empirical evidence showing the effects of ChatGPT-generated MTFI. Closing this gap is crucial to understanding the integration of ChatGPT to facilitate students' writing, which draws implications for effective human-AI collaboration in EFL writing instruction. Moreover, it can contribute to a more holistic evaluation of ChatGPT's affordances and limitations in different feedback techniques. This, in turn, can inform more varied strategies for capitalising on the AI tool to assist EFL writing. Therefore, this study used ChatGPT-generated model texts as a feedback instrument and ascertained its usefulness by examining the effects on EFL students' text quality and their perceptions.

2. Literature review

2.1. ChatGPT as a feedback tool

Since the launch of ChatGPT in November 2022, many scholars and educators have explored its affordances in language teaching and learning despite concerns over plagiarism and students' unethical use (e.g. Barrot 2023; Kohnke, Moorhouse, and Zou 2023). In the writing context, utilising ChatGPT as a writing feedback tool has attracted increasing scholarly interest (e.g. Guo and Wang 2024; Koltovskaia, Rahmati, and Saeli 2024; Lin and Crosthwaite 2024; Steiss et al. 2024). For instance, comparing teacher feedback and ChatGPT feedback, Guo and Wang (2024) found that ChatGPT generated more feedback than teachers, and its feedback was distributed evenly across three criteria (i.e. content, organisation, and language) with more direct revision guidance. By contrast, the study by Koltovskaia, Rahmati, and Saeli (2024) adopted a human-centred approach to examine six Iranian students' engagement with ChatGPT-generated feedback. The findings revealed that students accepted over half of ChatGPT's suggestions and generally held positive attitudes towards the feedback in making their writing more professional and native-like, which implied the usefulness of the AI tool in delivering feedback. Given its value, it is suggested that teachers can collaborate with the ChatGPT to assist their feedback provision and promote students' writing gains (Guo and Wang 2024).

However, researchers have also raised concerns over incorporating ChatGPT-generated feedback into the classroom. One prominent worry is the limitations of ChatGPT-generated feedback (Guo and Wang 2024; Koltovskaia, Rahmati, and Saeli 2024; Steiss et al. 2024). For example, teachers commented that the AI-generated feedback was lengthy, difficult to read, and sometimes irrelevant (Guo and Wang 2024). Students asserted that ChatGPT's suggestions were vague and repetitive, and they sometimes misinterpreted their intentions, which caused significant barriers to their uptake (Chen et al. 2024). Furthermore, high-quality WCF by ChatGPT demands enhanced literacy, especially in prompt engineering skills, which many EFL teachers might need sufficient training to master (Ma et al. 2024). These issues indicate that, though promising, effective use of ChatGPT-generated WCF in writing instruction might present significant challenges to teachers and students.

Despite this ongoing discussion on leveraging ChatGPT as an automated writing feedback tool, most current research seems to gravitate towards WCF, a feedback technique that mainly deals with writers' lexical and grammatical errors at the sentence level (Kang 2020). Scant studies have explored its affordances in other more discursive feedback strategies, such as MTFI and reformulation. Given the versatile functions of ChatGPT, this lack of scholarship might obscure our understanding of its role in diverse feedback practices, thus limiting its application in writing instruction.

2.2. Research on model texts as a feedback instrument

Models are samples of well-written texts that 'take into consideration the content and genre of the target text type' and align with learners' ages and proficiency levels (Coyle, Cánovas Guirao, and Roca de Larios 2018, 26). Using model texts as a feedback instrument typically involves three stages: composing an initial draft, comparing it with one or two model texts, and rewriting the original draft (Nguyen and Vu 2024). Theoretically, this practice is supported by the Output Hypothesis (Swain 2000) and the Noticing Hypothesis (Schmidt 2001). Framed within the Output Hypothesis (Swain 1985, 2000), learners identify their linguistic limitations when they compose their drafts (output), notice a gap between their language use and the model texts they receive (noticing), and reflect upon the noticed features to choose and incorporate these aspects into their rewritten texts to solve their problems (hypothesis-testing and metalinguistic processing). This process is believed to help learners understand the form-meaning connections and promote L2 development. Throughout these stages, students' noticing plays a pivotal role in language development because higher levels of noticing are associated with enhanced learning gains, as indicated by the Noticing Hypothesis (Schmidt 2001). These cognitive theories justify that model texts can be adopted as a valid feedback strategy for students' improved text quality (Tieu and Baker 2023).

Empirical studies across various contexts have provided evidence that EFL students could notice and incorporate features from the models into their texts, resulting in enhanced text quality (e.g. Cánovas Guirao, Roca de Larios, and Coyle 2015; Hanaoka 2007; Kang 2020; 2024; Nguyen, Nguyen, and Phuong 2024b). These studies typically adopted a quasi-experimental research design, and different writing genres have been deployed (e.g. argumentative essay, picture description, expository writing). For instance, Kang (2020) reported a significant impact of MTFI on 20 Korean high school EFL students' argumentative essay quality regarding vocabulary and content. The study by Nguyen, Nguyen, and Phuong (2024b) with 68 Vietnamese university EFL learners revealed that students with models outperformed their counterparts without receiving the model text in content, vocabulary, organisation, and overall scores with medium effect sizes. These studies established that MTFI could be a feasible and effective tool to promote EFL learners' writing. The value of MTFI was corroborated by students' perceptions, as research generally depicted positive attitudes towards its usefulness and a strong willingness to use MTFI (García Mayo and Loidi Labandibar 2017; Kang 2024; Nguyen and Vu 2024). As such, the MTFI approach is recommended to be implemented in EFL writing classrooms (Nguyen, Dao, and Nguyen 2024).

Although these studies offer insightful findings, using models as a feedback method remains a relatively nascent research area (Nguyen, Dao, and Nguyen 2024). One notable gap in this field is the limited exploration of integrating innovative technologies to assist MTFI. With recent and rapid technological advancements and the emerging scholarship on how AI can support feedback practices, investigating the use of AI tools like ChatGPT for crafting model text is both timely and warranted (Jiang and Hyland 2025; Nguyen, Dao, and Nguyen 2024). Arguably, ChatGPT holds promise as a model text provider. Built on a large language model (LLM), ChatGPT can generate coherent and grammatically correct human-like text in response to various prompts (Su, Lin, and Lai 2023). Its adaptivity and flexibility enable it to produce suitable models for the target learners regarding genre, length, complexity, and topic. More importantly, research evidence has demonstrated that the texts created by ChatGPT are indistinguishable from those by native speakers, indicating that its default writing proficiency level can be equal to that of advanced professional English writers (Nguyen and Barrot 2024). This underscores ChatGPT's ability to replicate high-quality teacher writing. In addition, ChatGPT can produce extended texts at a remarkable speed (Ray 2023), significantly expediting the process of supplying model text. These advantages position ChatGPT as a valuable and efficient tool in generating models to help direct learners to 'what quality performance looks like' (Steiss et al. 2024, 1).

However, researchers have identified some limitations of ChatGPT's output (Goulart et al. 2024; Jiang and Hyland 2025; Kohnke, Moorhouse, and Zou 2023). One is the potential biases embedded in its generated content. Since ChatGPT is trained on an English corpus that may not represent different cultures, it will likely create texts that exhibit biases by chance, potentially offering misleading or inappropriate examples for students (Barrot 2023). ChatGPT may also produce untruthful information and introduce fabricated evidence in the text due to its lack of real-world knowledge (Godwin-Jones 2024). Moreover, ChatGPT might fall short of fully emulating the nuances in human writing, such as enacting personal stances and engaging in persuasive interaction (Goulart et al. 2024; Jiang and Hyland 2025). These constraints highlight the pitfalls of using ChatGPT to generate models, particularly for specific genres like argumentative writing that demands credible evidence and the articulation of personal perspectives.

2.3. The present study

The literature reviewed above suggests that while the potential of AI, especially ChatGPT, in facilitating WCF has been increasingly recognised, its application in other feedback methods, such as MTFI, remains largely underexplored. This gap limits the broader understanding of ChatGPT's pedagogical application in feedback provision to improve EFL learners' writing. Given ChatGPT's advanced text-processing capabilities and flexibility in generating texts resembling human writing, it holds the potential to be an effective tool for MTFI. However, its effectiveness in this domain has yet to be tested. Moreover, it has been identified that most MTFI research has been situated in the European context, with populations in other settings underrepresented (Nguyen, Dao, and Nguyen 2024). This study aimed to address these gaps by examining the usefulness of ChatGPT-generated MTFI with mainland Chinese EFL learners. Doing so contributes to the growing body of research on AI-assisted feedback and offers a more comprehensive picture of ChatGPT's educational value in writing, informing enhanced EFL writing pedagogy in the evolving AI era. Specifically, the research questions were formulated as follows:

(1) Does the use of ChatGPT-generated model texts as a feedback instrument significantly enhance mainland Chinese EFL students' text quality in rewritten essays?
(2) How do these students perceive the use, effectiveness, strengths and weaknesses of ChatGPT-generated model texts as a feedback instrument?

3. Methodology

3.1. Research context and participants

The present study was conducted at a high school in Southeast China. The participants were 104 Grade 11 students (aged 16–17, 55 males and 49 females) from three intact classes. They were all EFL learners with Mandarin or Cantonese as their first language and had been studying English for over 9.8 years ($SD = 1.85$) without any overseas study experience. Their English proficiency levels, determined by the Oxford Placement Test (OPT) (UCLES 2001), ranged from B1 to B2 on the CEFR (Common European Framework of Reference) scale. Results from a one-way ANOVA showed no significant difference between these three classes regarding their proficiency levels ($F_{(2,101)} = 1.292$, $p = .279$), which ensured they were comparable in this regard. Therefore, one class was randomly assigned as the control group (CG, $n = 35$), one as the experimental group 1 (EG1, $n = 33$), and one as the experimental group 2 (EG2, $n = 36$). Importantly, all participants had prior experience with both human-authored and AI-generated model texts as a feedback instrument in their regular writing instruction.

3.2. Design

In line with previous MTFI research (e.g. Kang 2024; Nguyen, Nguyen, and Phuong 2024b), this study adopted a quasi-experimental design with three groups (CG, EG1, and EG2) over a three-stage writing task: (1) composing (pretest), (2) comparing, and (3) rewriting (post-test). Both EG1 and EG2 participated in stage two, where EG1 compared their writing with a model text written by experienced teachers, whereas EG2 compared their text with a model generated by ChatGPT. The CG did not participate in stage 2, receiving no feedback in any form.

3.3. Data collection instruments

3.3.1. Writing task and model texts

An argumentative writing task was adopted in this study because this genre is commonly used to evaluate EFL learners' writing proficiency (Hirvela 2017). The task required students to give their opinions on whether environmental problems are too big for individuals to solve (See Appendix A). This topic was selected because students had learnt about environmental protection by the time of data collection, which equipped them with the background knowledge. A pilot test of the writing task was administered to another group of learners in the same grade at the same school to determine its appropriateness and elicit students' preferred stance. Results indicated that the writing task was suitable, and most students held the view that individuals can contribute to solving environmental problems.

Only one model text was provided to each experimental group to reduce their cognitive load and control the potential effect of working memory on their subsequent incorporations (Kang 2023). Based on the result from the pilot test, both model texts were created arguing against environmental issues being too big for individuals to solve. The model text given to the EG1 (model 1) was written by the class teacher (See Appendix B), who had over seven years of teaching experience at the same school. Then, it was modified by a native speaker and carefully reviewed by two other EFL teachers. The model supplied to the EG2 (model 2) was generated by ChatGPT (See Appendix C). The first author developed a prompt that described the specific context and requirements (Table 1) and used it to request the ChatGPT 4o version to produce the model. This prompt is believed to reflect the prompting strategy typically adopted by practising EFL teachers who may have little training in prompt engineering. We are aware that advanced prompting techniques could elicit model texts of higher quality. However, this might misalign with authentic instructional settings because teachers are observed to lack the ability to write effective, complex prompts (Ma et al. 2024) at the time of data collection.

Both model texts were pilot-tested with a group of 20 students with a similar profile, who were asked to rate the comprehensibility on a seven-point Likert scale (1 = very easy to understand and 7 = very difficult to understand). The results showed that both models were of a suitable difficulty level for the participants ($M = 3.62$, $SD = 1.15$ for model 1 and $M = 3.77$, $SD = 1.25$ for model 2).

Table 1. Prompt used to generate model text by ChatGPT.

Prompt
Your job is to write a model argumentative essay for a group of learners so that they can learn from your model text in terms of content, organisation, vocabulary, and grammar. The target learners are mainland Chinese high school Grade 11 students with B1 to B2 proficiency levels (CEFR). You need to write on the topic of whether environmental problems are too big for individuals to solve. Please adopt the argument that environmental problems are not too big for individuals to solve. Write at least 150 words.

3.3.2. Questionnaires

A two-part questionnaire was adopted in this study to answer RQ2 (See Appendix D). The first part collected students' demographic information, including age, gender, and years of studying English. The second part included question items adapted from Nguyen and Vu (2024), which were intended to capture participants' perceptions of the model's usefulness as a way to enhance their writing with regard to content, organisation, vocabulary, and grammar. The items were designed on a five-point Likert scale (1 = strongly disagree and 5 = strongly agree). Two open-ended questions were also included to elicit participants' views on the general strengths and weaknesses of ChatGPT-generated model texts, as well as their overall preferences between model texts produced by teachers and those generated by ChatGPT.

3.4. Procedures

The data collection spanned four weeks. In the first week, all participants were informed about the research, signed a consent form, and completed the OPT. In the second week, they were engaged in Stage 1, where they were required to finish the argumentative writing task in class within 40 min. Upon completion, they were asked to report in a paper the writing problems they encountered while composing. In the following week, both EG groups attended Stage 2, in which they received their first draft, the corresponding models, and a sheet to finish a comparison task. The authorship of the model text was not disclosed at this stage to avoid biasing students' engagement. They were asked to write down the similarities and differences between their writing and the model, as well as the features they perceived as useful in the model. Both groups were allowed 30 min to finish the comparison task, either in English or Mandarin. The CG was encouraged to identify and correct their own mistakes within the same time frame. One week later, in Stage 3, all groups (CG, EG1, and EG2) were asked to write on the same prompt without external support (e.g. materials or models). Following this, the EG2 completed the questionnaire in English or Mandarin, which took approximately 15 min. Figure 1 summarises the procedures of this study.

3.5. Data analysis

3.5.1. Scoring of the writing tests

The writing tests were scored based on an analytical rubric adopted by Kang (2024). Four dimensions were assessed: content, organisation, vocabulary, and grammar (See Appendix E). For each dimension, 0–5 points were possible, and the scores of the four dimensions were combined to yield a total score. The two authors were involved in the rating. The Pearson correlation coefficients (r) for the scores given by the two raters were 0.89 and 0.93 for the first and rewritten essays, respectively. The final score of each participant was the aggregated average value of the scores by the two raters.

3.5.2. Questionnaire

Descriptive statistics (i.e. mean, standard deviation, and frequencies) were reported for the questionnaire. Following the interpretation adopted by Nguyen and Vu (2024), the mean values were categorised as low (under 2.5), medium (2.5–3.49), and high (over 3.5).

For the coding of the open-ended responses, a conventional content analysis approach was employed (Hsieh and Shannon 2018). This method involved a systematic inductive coding process aimed at identifying emerging themes under three broad categories: strengths, weaknesses, and comparison between human- and ChatGPT-generated model texts. Within each category, the first author conducted a preliminary round of open coding, where responses were broken down into smaller units of meaning (e.g. phrases or sentences) linked to specific perceptions or ideas. To ensure the reliability of the coding process, once the first author completed the initial coding, the responses, with the initial codes removed, were sent to the second author for an independent analysis. The authors then met to compare their coding, resolve discrepancies, and examine relationships between codes. Through this iterative process, the codes were refined and organised into the key themes presented in this study. According to Cohen's kappa coefficient (*k*), the intercoder agreement rate yielded a value of 0.84, reflecting a substantial level of agreement in content analysis (Neuendorf 2017).

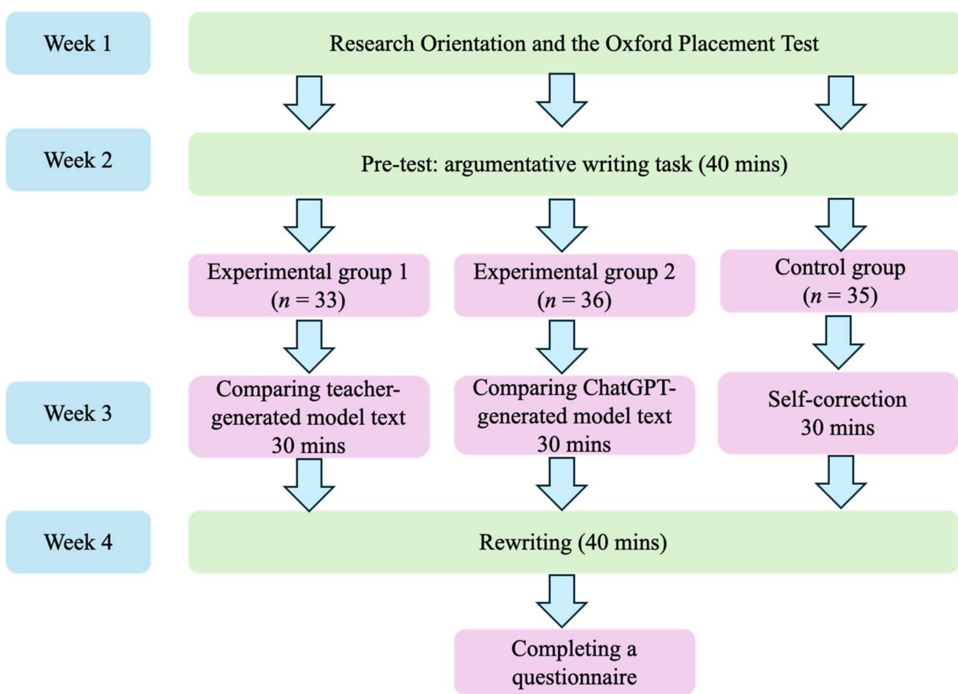

Figure 1. The study procedure, Figure 1.

3.5.3. Statistical analysis

To evaluate the effects of the ChatGPT-generated model text in any of the dimensions to be studied (RQ3), a mixed ANOVA was conducted, incorporating both a within-groups variable (time) and a between-groups variable (group). Assumptions of normality, homogeneity of variance, and sphericity were met. Simple effect analysis and paired-sample *t*-tests were used as the post-hoc test within groups. The alpha level was set at .01 for the main effects, interaction effects, and pairwise comparisons. Where there were multiple comparisons, Bonferroni corrections were used to avoid Type I errors.

Effect sizes were also calculated for each statistical analysis to measure the magnitude of observed effects. Partial eta squared was used to assess effect sizes for mixed ANOVAs, and Cohen's *d* was calculated for post-hoc comparisons. Following Cohen's (1992) recommendation, η_p^2 values of .01, .06, .14; and *d* values of .20, .50, .80 were interpreted as small, medium, and large effect sizes, respectively.

3.5.4. Textual analysis

In addition to statistical analysis, a textual analysis was conducted to gain a more nuanced understanding of the effects of the ChatGPT-generated model text. Specifically, EG2 students' essays from Stages 1 (pretest) and 3 (post-test) were systematically compared to identify textual revisions. These changes were then cross-referenced with the ChatGPT-generated model text to determine whether specific model-derived features had been incorporated. In alignment with the analytical rubric, the identified features were categorised into four dimensions: content, organisation, vocabulary, and grammar. To enhance the credibility of the analysis, all classifications were independently reviewed by two researchers, with any discrepancies resolved through discussion. This qualitative process supplemented the quantitative analysis by enabling an in-depth examination of how the model text addressed students' writing challenges. The Results section below presents representative examples of these incorporations and their contributions to improved writing quality.

4. Results

4.1. Quantitative effects of the ChatGPT-generated model text

Table 2 displays the descriptive statistics for the scores of the initial and rewritten essays among the three groups. It can be seen that the control group (CG) showed minimal improvement in total scores from the pretest ($M = 8.31$, $SD = 1.68$) to the post-test ($M = 8.67$, $SD = 1.54$). In contrast, experimental groups (EG1 and EG2) showed significant increases, with EG2 achieving the highest post-test total score ($M = 13.34$, $SD = 1.98$). The results from a one-way ANOVA (See Appendix F) suggested no significant difference between the three groups across the four dimensions and total scores in the pretest. Figure 2 visually presents the changes in mean scores across time.

The results of the mixed-design ANOVA (Table 3) revealed significant main effects for time and group, as well as significant interaction effects for all measured variables, including total scores ($F = 43.795$, $p < .001$, $\eta_p^2 = .464$), content ($F = 14.969$, $p < .001$, $\eta_p^2 = .229$), organisation ($F = 43.435$, $p < .001$, $\eta_p^2 = .244$), vocabulary ($F = 12.779$, $p < .001$, $= .202$), and grammar ($F = 14.992$, $p < .001$, $\eta_p^2 = .229$). Post-hoc pairwise comparisons at the post-test

Table 2. Descriptive statistics for the pre and post-test scores.

	CG ($n = 35$)				EG1 ($n = 33$)				EG2 ($n = 36$)			
	Pretest		Posttest		Pretest		Posttest		Pretest		Posttest	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Total	8.31	1.68	8.67	1.54	8.58	1.76	12.67	2.10	8.90	1.92	13.34	1.98
Content	1.99	.70	2.29	.66	1.98	.55	3.26	.56	2.29	.11	3.50	.10
Organisation	2.14	.38	2.39	.47	2.00	.35	3.26	.52	2.11	.51	3.47	.62
Vocabulary	2.07	.61	2.27	.74	2.24	.59	3.15	.63	2.25	.49	3.22	.57
Grammar	2.20	.47	2.27	.51	2.35	.59	3.00	.67	2.25	.50	3.15	.57

stage (Table 4) showed that both experimental groups (EG1 and EG2) significantly outperformed the control group (CG) across the analytical measures, with large effect sizes ($d > .80$). EG2 consistently achieved the highest scores, but the differences between EG1 and EG2 in all dimensions were smaller and non-significant.

From the pretest to post-test, both experimental groups demonstrated significant improvements in total scores (Table 5), with large within-group effect sizes (EG1: $d = 1.92$; EG2: $d = 1.61$), while the control group showed no significant change ($p = .396$, $d = 0.42$). Similar patterns were observed for the specific aspects, where EG1 and EG2

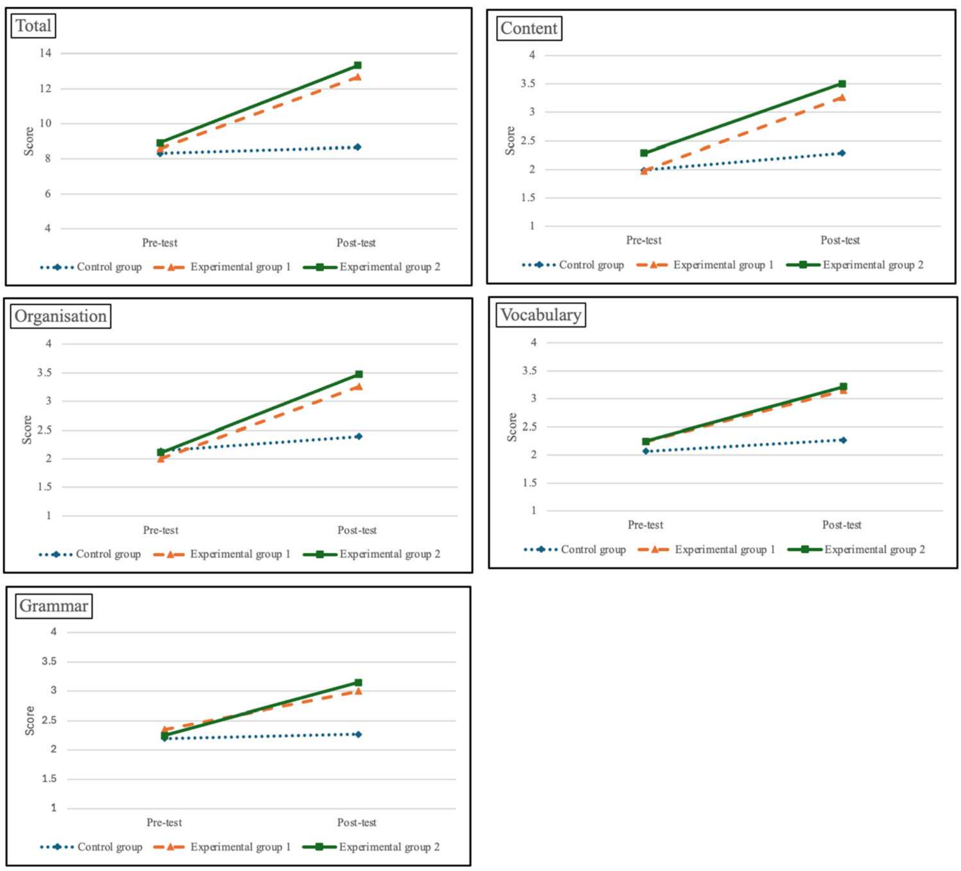

Figure 2. Group means across time for total and sub-dimension scores, Figure 2.

Table 3. Results of the mixed ANOVA.

Effect	*F*	*df*	*p*	η_p^2
Total				
Time	222.685	(1, 101)	0.000*	.688
Group	28.962	(2, 101)	0.000*	.364
Time × Group	43.795	(2, 101)	0.000*	.464
Content				
Time	130.024	(1,101)	.000*	.563
Group	23.411	(2,101)	.000*	.317
Time × Group	14.969	(2,101)	.000*	.229
Organisation				
Time	307.365	(1,101)	.000*	.753
Group	16.305	(2,101)	.000*	.462
Time × Group	43.435	(2,101)	.000*	.244
Vocabulary				
Time	99.393	(1,101)	.000*	.496
Group	14.287	(2,101)	.000*	.221
Time × Group	12.779	(2,101)	.000*	.202
Grammar				
Time	71.271	(1,101)	.000*	.414
Group	11.881	(2,101)	.000*	.190
Time × Group	14.992	(2,101)	.000*	.229

** $p < 0.01$.

exhibited moderate within-group improvements (Cohen's *d* ranging from 0.49 to 0.68), and the CG made no significant progress.

4.2. Qualitative evidence of ChatGPT-generated model text's impact on student writing

A closer comparison between EG2 learners' initial and rewritten texts revealed that the ChatGPT-generated model text addressed students' writing challenges by serving as comprehensible input, enabling them to incorporate key features that improve their writing quality. For instance, one major weakness found in students' initial drafts was the lack of elaboration and evidence to validate their claims, probably due to insufficient knowledge of effective argumentative essays. However, the ChatGPT-generated model text provided these students with detailed examples to support their arguments that individuals can help solve environmental problems, such as starting with small moves and influencing their social circles and companies to take action. Noticing this in the comparison task, students included these supporting details in their subsequent writing, leading to higher scores in content. Table 6 shows a concrete instance of how a student revised the content by adding illustrative examples of small actions based on the ChatGPT-generated model text.

Table 4. Pairwise comparisons between groups at the post-test stage.

	CG – EG1			CG – EG2			EG1 – EG2		
	MD	*p*	Cohen's *d*	MD	*p*	Cohen's *d*	MD	*p*	Cohen's *d*
Total	−2.138	.000*	−1.389	−2.632	.000*	−1.710	−0.504	.532	−0.328
Content	−0.486	.000*	−1.027	−0.760	.000*	−1.606	−0.275	.053	−0.581
Organisation	−0.365	.001*	−0.918	−0.527	.000*	−1.326	−0.163	.276	−0.410
Vocabulary	−0.526	.000*	−1.067	−0.565	.000*	−1.146	−0.039	1.000	−0.079
Grammar	−0.439	0.000*	−0.979	−0.466	.000*	−1.039	−0.027	1.000	−0.060

MD = mean difference, * $p < 0.01$.

Table 5. Comparisons of the scores within each group over two testing times.

	CG ($n = 35$)			EG1 ($n = 33$)			EG2 ($n = 36$)		
Pretest – post test	SE	*p*	Cohen's *d*	SE	*p*	Cohen's *d*	SE	*p*	Cohen's *d*
Total	.42	.396	2.46	.33	.000*	1.92	.27	.000*	1.61
Content	.18	.107	1.07	.12	.000*	.66	.11	.000*	.68
Organisation	.10	.025	.61	.09	.000*	.53	.09	.000*	.52
Vocabulary	.16	.210	.93	.09	.000*	.54	.10	.000*	.59
Grammar	.12	.566	.73	.13	.000*	.72	.08	.000*	.49

SE = standard error, * $p < 0.01$.

Vocabulary was another dimension that showed noticeable differences between the initial and revised drafts. Specifically, students in the EG2 only used some basic vocabulary in their essays in the pretest, and some phrases seemed to be direct translations from Mandarin that deviated from English conventions (e.g. *everyone's effort gathered together*). By contrast, they included a wider range of lexical items suitable for environmental protection (e.g. *conserve natural resources*, *green initiatives*), a feature demonstrated in the ChatGPT-generated model text. In addition, students demonstrated more appropriate and precise word choices (e.g. *collectively, broader societal changes*) after engaging with the model (Table 7). These enhancements in lexical variety and accuracy contributed to improved vocabulary performance.

Regarding organisation, many of the initial drafts consisted of a single, lengthy body paragraph that included multiple subclaims with clear separation. However, students' rewritten texts typically featured distinct body paragraphs, each presenting a focused subclaim supported by the relevant evidence, mirroring the structure of the ChatGPT-generated model text. This structural improvement reflects the influence of the model text, which provided a clear exemplar of effective organisational conventions that students were able to emulate in their revisions. In terms of grammar, the ChatGPT-generated model text supported students' development by exposing them to more complex syntactic structures, which they subsequently adopted in their rewritten texts. Table 8 presents an example of a student who drew on the structure of a sentence from the model text to construct a complex sentence in the revised version.

4.3. Students' perceptions of the ChatGPT-generated model text

Table 9 shows the results of participants' perceptions of the usefulness of the ChatGPT-generated model text adopted in this study. As shown in the table, our participants generally regarded using the ChatGPT-generated model text as a feedback technique as somewhat helpful in improving their writing ($M = 3.89$, $SD = .67$), with a significant majority of participants (77.8%) agreeing or strongly agreeing that the model text was useful. In terms of specific aspects, content was perceived as the most beneficial, with 63.9% of participants agreeing or strongly agreeing that model text aided their

Table 6. An example of a student's revision in the content aspect (the EG2).

Original draft	Don't be afraid that starting small counts little. Small drops can gather to become a strong flood. It is every tiny effort from individuals that makes a different.
Rewritten essay	Then, reducing various kinds of waste by recycling, avoiding single-use plastic bags and bottles means a lot to environment, greatly easing the pressure of breaking down waste.

Note: The extracts were presented in the student's own words.

Table 7. A student's revision on vocabulary use.

Original draft	Individual's effort can lead to crucial change, and everyone's effort gathered together can also form a surge of superpower.
Rewritten essay	Initially, small actions taken by individuals can collectively accumulate to enormous changes.

Note. The extracts were presented in the student's own words.

improvement in this area. Vocabulary followed, with 58.3% of students endorsing its usefulness. Organisation and grammar were rated similarly. Almost half of the participants chose either *agree* or *strongly agree* options regarding the usefulness of the model for revising the organisational structures of their writing (47.8%) and grammar (50%), respectively. Only a small minority of students disagreed with the usefulness of the model across any aspect, reinforcing the overall positive perception of ChatGPT-generated texts as a feedback tool. Figure 3 visually illustrates these trends, offering a clearer view of response distributions across dimensions.

A qualitative analysis of participants' responses to the open-ended questions painted a more detailed picture regarding participants' perspectives on ChatGPT-generated model texts. As depicted in Table 10, two-thirds of participants ($n = 24$) noted that ChatGPT-generated texts contained a variety of words and expressions for paraphrasing, which enriched their lexical repertoire and improved their vocabulary in rewriting. Over 60% of participants ($n = 23$) mentioned that reading the model texts helped clarify their thinking and reasoning. Other notable strengths included the inclusion of a range of grammatical structures ($n = 11$), the ability to provide more ideas and arguments ($n = 10$), and their convenience and efficiency ($n = 10$). About 20% of participants ($n = 8$) also appreciated the support the texts provided in structuring essays, while three students highlighted flexibility as a unique advantage of ChatGPT-generated model texts.

Regarding weaknesses, the most commonly reported issue was that ChatGPT-generated texts could restrict their content development and creativity during revision ($n = 17$). Additionally, about 28% of participants ($n = 10$) described the output from ChatGPT as formulaic, unnatural, and monotonous. A few participants also commented that the texts included difficult words ($n = 7$) and lacked emotional depth, life experiences ($n = 5$), and logical reasoning ($n = 6$). Three students also indicated that the model texts did not offer concrete suggestions for improving their writing. Two students pointed out the need for advanced prompting skills to use the tool to generate satisfactory model texts.

The comparison between teacher-generated and ChatGPT-generated model texts revealed distinct advantages for each type. Approximately 20% of participants highlighted human insights into the topic as the crucial feature of teacher-generated model texts. In addition, models from teachers were praised for better alignment with test requirements and students' existing knowledge ($n = 9$), greater adaptability to their

Table 8. A student's use of a complex sentence structure inspired by the ChatGPT-generated model.

Original draft	In addition, individuals have the ability to raise society's awareness of the importance to protect environment.
Rewritten essay	By adopting eco-friendly habits like using less plastic or saving energy, individuals can inspire their family or friends to do the same things, thereby escalating the personal action into society efforts.

Note: The extracts were presented in the student's own words.

Table 9. Students' responses on the usefulness of the ChatGPT-generated model text ($n = 36$).

Area	*M*	*SD*	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Overall	3.89	.67	0 (0%)	1 (2.8%)	7 (19.4%)	23 (63.9%)	5 (13.9%)
Content	3.69	.67	0 (0%)	1 (2.8%)	12 (33.3%)	20 (55.6%)	3 (8.3%)
Organisation	3.50	.81	0 (0%)	3 (8.3%)	16 (44.4%)	13 (36.1%)	4 (11.1%)
Vocabulary	3.61	.84	1 (2.8%)	1 (2.8%)	13 (36.1%)	17 (47.2%)	4 (11.1%)
Grammar	3.50	.70	0 (0%)	2 (5.6%)	16 (44.4%)	16 (44.4%)	2 (5.6%)

Note: Numbers in parentheses indicate the percentage of each response.

backgrounds and levels ($n = 6$), and their authenticity, emotional resonance ($n = 8$) and flexibility ($n = 6$). Conversely, ChatGPT-generated texts were valued for having fewer errors ($n = 10$) and a standardised writing style ($n = 9$). Some participants also articulated that ChatGPT tended to use more sophisticated language ($n = 7$) and had the potential to customise the model text to meet personalised needs ($n = 5$). Notably, five students reported that they preferred a combination of ChatGPT and human expertise in generating model texts, acknowledging the benefits and drawbacks of each type and underscoring the need for human-AI collaboration.

5. Discussion

This study aimed to investigate the usefulness of ChatGPT-generated model texts as a feedback instrument by exploring its effects on EFL students' text quality and their perceptions. The quantitative analysis of this study highlights the effectiveness of the ChatGPT-generated model text on learners' overall text quality and specific dimensions, including content, organisation, vocabulary, and grammar. Specifically, the control and experimental groups in our study were comparable in terms of English proficiency levels and pretest scores at the outset. However, in the post-test, learners who received the ChatGPT-generated model text outperformed the control group across all dimensions with large effect sizes. These findings corroborate previous intervention studies that

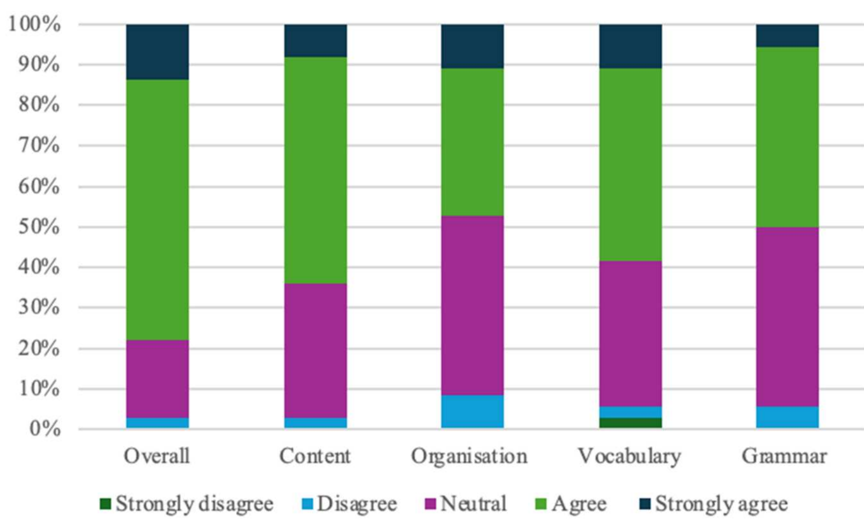

Figure 3. Percentage distribution of students' responses, Figure 3.

Table 10. Themes from the content analysis of the open-ended questions ($n = 36$).

Category	Theme	Number of survey respondents (percentage)
Strengths	• Clarifies thinking and reasoning	23 (63.9%)
	• Provides more ideas and arguments	10 (27.8%)
	• Contains a range of topic-based, sophisticated words and expressions	24 (66.7%)
	• Helps structure the essay	8 (22.2)
	• Includes a range of grammatical structures	11 (30.6%)
	• Flexible to produce	3 (8.3%)
	• Convenient and efficient	10 (27.8%)
Weaknesses	• Places limits on content development and creative thinking	17 (47.2%)
	• Lacks emotions, life experiences, and depth of thinking	5 (13.9%)
	• Demands advanced prompting skills	2 (5.6%)
	• Difficulty in understanding some words	7 (19.4%)
	• Does not have explicit comments on directions for improvement	3 (8.3%)
	• Formulaic, unnatural, and monotonous content and language	10 (27.8%)
	• Illogical reasoning	6 (16.7%)
Comparison between human- and ChatGPT-generated models	Teacher-generated model texts • More tailored to students' backgrounds and levels	6 (16.7%)
	• Align with test requirements and the acquired knowledge	9 (25%)
	• Fewer errors	1 (2.8%)
	• Has in-depth insights and clear reasoning	7 (19.4%)
	• Authentic and emotionally resonant	8 (22.2%)
	• Adaptable and less rigid	6 (16.7%)
	ChatGPT-generated model texts • Fewer errors	10 (27.8%)
	• More standardised writing style	9 (25%)
	• More advanced language	7 (19.4%)
	• Can be suitable for high school student	1 (2.8%)
	• Can be customised to meet personalised needs	5 (13.9%)

found significant improvements in text quality through MTFI treatment, whether in narrative (e.g. Luquin and García Mayo 2024) or expository writing (e.g. Kang 2024; Tieu and Baker 2023). Comparing the two experimental groups, although students with the ChatGPT-generated model showed slightly higher scores than those with model text produced by human teachers in all four analytical measures, no significant differences were observed, indicating that the two experimental groups made similar improvements from the initial writing to the rewriting. This suggests that, under comparable conditions, ChatGPT-generated model texts can be as effective as human-produced ones in supporting student writing across multiple dimensions. The qualitative analysis of students' initial and revised essays further reinforces the quantitative findings by illustrating how the model text served as concrete, comprehensible input that addressed students' writing difficulties – features learners subsequently adopted and applied in their revisions.

The comparable effectiveness might stem from the similar quality and relevance of the two types of model texts. In this study, the prompt given to ChatGPT was carefully crafted to specify the pedagogical purpose of the model (i.e. to serve as a feedback instrument for students' learning in the content, organisation, vocabulary, and grammar) and the writer's stance. As a result, ChatGPT was able to produce a model text that addressed students' needs and challenges in a targeted way, largely mirroring the kind of support offered by the human-written text. More importantly, the prompt also specified contextual factors such as students' grade levels and proficiencies, which enabled ChatGPT to generate a suitable model text aligning well with the target learners. Indeed, the pilot test results confirmed that the complexity of the ChatGPT-generated text was equally appropriate to the teacher-authored version. This alignment with students' proficiency levels facilitated their comprehension and likely sustained their motivation to notice the useful features and incorporate them into their subsequent drafts (García Mayo and Loidi Labandibar 2017), thus contributing to improved text quality. These findings highlight the importance of mindful prompt design that clearly articulates learning goals and accounts for learner context in order to ensure that ChatGPT-generated model texts are pedagogically effective and on par with those created by human instructors.

Concerning learners' perceptions of the effectiveness of ChatGPT-generated MTFI, they generally acknowledged its usefulness in enhancing their overall writing quality and performance in specific aspects. This finding aligns with previous studies, which revealed learners' positive attitudes towards MTFI (García Mayo and Loidi Labandibar 2017; Hanaoka 2007; Nguyen and Vu 2024). In particular, our participants expressed that using the ChatGPT-generated model texts clarified their reasoning, provided a diverse repertoire of vocabulary and sentence structures, and expanded the scope of their ideas. These benefits echo the MTFI research using teacher-generated model texts, where EFL learners appreciated the models in allowing them to enrich essay content and acquire new lexical items (e.g. Kang 2024; Nguyen, Nguyen, and Phuong 2024b). Notably, students' observed revisions incorporating content and vocabulary features from the model also triangulate these self-reported perceptions, reinforcing the perceived usefulness of the model text. From this perspective, ChatGPT, with its exceptional capability to produce error-free, consistent, coherent human-like text, could serve as a satisfactory alternative to human-generated model texts, demonstrating quality writing to students and lessening the burden on teachers as the sole feedback provider (Steiss et al. 2024).

However, students mentioned some constraints of ChatGPT's models compared with human-generated text, such as the unnaturalness and rigidity of the language and insufficient depth in emotion and content. These comments were in line with researchers' findings that the style of ChatGPT's output tends to be formulaic with a formal structure and neutral tone, lacking sentiment elements and involvement (Goulart et al. 2024; Jiang and Hyland 2025; Nguyen and Barrot 2024). Thus, although AI-generated texts might excel in certain aspects, teachers are still not substitutable, especially when the target writing task requires higher-order thinking skills, engagement with the subject matter, and establishment of authors' distinctive voices (Barrot 2023). Another limitation of ChatGPT-generated model texts perceived by our participants is the lack of contextual relevance, such as the test requirements, students' prior knowledge and cultural backgrounds, and the target audience. This is because ChatGPT, as an LLM that generates content based on mathematical calculation, lacks real-world knowledge, experience, and true situational awareness (Godwin-Jones 2024). Therefore, rather than fully outsourcing the creation of model texts to AI, it is suggested that teachers collaborate with ChatGPT to develop model texts that are pedagogically sound and contextually relevant. Teachers can strategically guide the text-generation process by crafting well-defined prompts that include essential contextual factors. After receiving the ChatGPT-generated draft, they need to review it critically and engage in iterative interaction with the AI tool to refine the output until it meets the desired standard. Teachers should adjust the writing style manually when necessary and supplement their real-world perspectives, cultural references, and subject-specific nuances that AI may overlook. Such a hybrid approach, where teachers actively apply their pedagogical content knowledge (PCK) and evaluative judgement while harnessing the strengths of AI, might strike a balance between human expertise and technological innovation, optimising the use of model texts as a feedback instrument.

Notably, our participants highlighted ChatGPT's potential to generate individualised model texts. Cao and Mao (2024) emphasise that model texts, as a form of positive evidence feedback, should be tailored to individual needs and appropriately aligned with students' proficiency levels to ensure they clearly understand how to effectively act upon the feedback provided. Furthermore, the provision of model texts should be attuned to the learner's Zone of Proximal Development (ZPD), considering their current and potential performance level (Nguyen and Vu 2024). We argue that ChatGPT, with its adaptivity and flexibility, can present new opportunities for breaking the one-size-fits-all philosophy (Weng and Chiu 2023) by producing model texts that can cater to learners' abilities and preferences, therefore supporting their subsequent noticing and rewriting. This personalisation can be achieved by incorporating students' profiles (e.g. proficiency levels, writing challenges, and learning goals) when prompting ChatGPT to generate model texts. As such, teachers can leverage ChatGPT to provide differentiated model texts that accommodate diverse proficiencies and learning needs within a heterogeneous classroom, which allows students to receive appropriate scaffolding. Similarly, through proper prompts that specify factors such as language complexity and writing conventions, teachers can adjust the model texts for different educational contexts (e.g. secondary schools or higher education) to match institutional expectations. Still, as our participants pointed out, maximising ChatGPT's customisation potential might require advanced prompting skills, raising the need for AI literacy training for teachers.

Interestingly, one major weakness of ChatGPT-generated MTFI reported by our participants was that this feedback technique would likely hinder their creativity and content development in rewriting. This contrasts with the results in Kang's (2024) study, which illustrated that students were mainly concerned about the difficulty of understanding some words and grammatical structures and the lack of explicit suggestions in directing their improvement when using MTFI. Since the ChatGPT-generated model texts can offer writing samples that appear difficult to surpass, students might become overly reliant on them, limiting their opportunities for independent improvement in their rewritten texts. To alleviate such an issue, teachers should educate students to view models as a feedback tool rather than an absolute standard. Providing demonstrations on effectively using model texts during revision and rewriting is also encouraged.

6. Conclusion

To the best of our knowledge, this study is the first to investigate the use of ChatGPT in feedback techniques beyond written corrective feedback, with a particular focus on model texts as a feedback instrument. The findings of this study revealed that utilising ChatGPT-generated model texts could yield comparable positive effects to using model texts composed by human teachers. Thus, ChatGPT-generated model texts could be a valuable feedback tool to enhance EFL students' text quality in content, organisation, vocabulary, and grammar. Students' perceptions also supported the usefulness of ChatGPT-generated model texts as a feedback instrument, as they generally held positive attitudes towards such a practice despite concerns about the potential lack of contextual relevance and the stylistic unnaturalness of the AI-generated model texts.

However, some limitations of this study are acknowledged. First, this study only utilised one model text, which may not expose participants to sufficient input to notice and incorporate into their subsequent writing. It is suggested that future studies administer two model texts to provide richer input resources to mitigate this issue. In addition, this study may fail to capture the sustained effect of ChatGPT-generated MTFI because it only adopted a one-shot experiment. Hence, longitudinal studies can be conducted to investigate the long-term effect. It should be noted that the effectiveness of ChatGPT-generated model texts was specific to the same writing prompts. Whether the positive effects can be retained and transferred to a new writing task with a different topic warrants further investigation. Besides this, the exclusive focus on a single argumentative writing task in this study might limit the applicability of the findings to other genres, such as narrative or expository writing. Future studies can consider exploring the use of ChatGPT-generated model texts across different genres to provide a more comprehensive understanding. Moreover, this study relied solely on holistic scoring to measure students' text quality. Employing diverse measurements (e.g. the CAF measures) could offer a more nuanced assessment of the effects of model texts on learners' written output. Regarding ChatGPT, this study used only a single contextual template prompt to generate the model text. Although it is believed that this prompt can reflect the current prompting literacy of a language teacher without substantial training in prompt engineering, future studies can use more diverse prompting formats (e.g. role-based, zero-shot, etc.) and scenarios to improve ChatGPT's performance to generate suitable model texts that align with students' profiles. Future studies could also adopt a within-subjects design to explore how learners engage with both human-authored and AI-generated model texts, which may contribute to a more nuanced understanding of their role in academic writing.

Despite these limitations, this study may have some implications. First, instead of seeing model texts and error correction as a dichotomy, teachers can combine both methods when giving feedback, contingent on the focus of teaching and assessment (e.g. form-focused or meaning-oriented). For instance, teachers can provide model texts after students finish their first drafts, guiding them to compare different aspects of their writing and the models and encouraging self-editing. Teachers' WCF can be given to students' revised texts to furnish their written product. Second, it is recommended that teachers collaborate with ChatGPT in MTFI, especially in classrooms with students of varying proficiency levels and needs. They can capitalise on ChatGPT to efficiently produce different sets of appropriate model texts regarding students' levels, topics, and lengths. However, teachers should examine and refine the AI-generated models using their pedagogical content knowledge, such as awareness of the teaching context (e.g. students' cultural backgrounds and prior knowledge), familiarity with the genre, and insights into the subject matter and the specific writing tasks. This 'human-in-the-loop' approach, which involves active human oversight, ensures an effective synergy between human expertise and technology assistance. Third, when employing ChatGPT-generated model texts as a feedback instrument, teachers are encouraged to maintain transparency with students and foster their critical evaluation and proper use of these texts. Demonstrations and training on leveraging ChatGPT-generated model texts can also be provided to students to maximise the benefits and promote a balanced and informed approach to utilising AI tools in their writing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statment

Data will be made available on request.

ORCID

Dunming (Jason) Lu http://orcid.org/0009-0008-0160-3707

References

Barrot, J. S. 2023. "Using ChatGPT for Second Language Writing: Pitfalls and Potentials." *Assessing Writing* 57:100745. https://doi.org/10.1016/j.asw.2023.100745.

Bitchener, J., and N. Storch. 2016. *Written Corrective Feedback for L2 Development*. Bristol: Multilingual Matters.

Cánovas Guirao, J., J. Roca de Larios, and Y. Coyle. 2015. "Models Can Be Used as a Written Feedback Technique with Young EFL Learners." *System* 52:63–77. https://doi.org/10.1016/j.system.2015.04.002.

Cao, Z., and Z. Mao. 2024. "Studies on Positive Evidence Feedback in Second Language Writing: Status Quo and Implications." *RELC Journal* 55 (1): 209–218. https://doi.org/10.1177/00336882211072905.

Chen, Z., X. Zhu, Q. Lu, and W. Wei. 2024. "L2 Students' Barriers in Engaging with Form and Content-Focused AI-Generated Feedback in Revising Their Compositions." *Computer Assisted Language Learning* 1–21. https://doi.org/10.1080/09588221.2024.2422478.

Cohen, J. 1992. "A Power Primer." *Psychological Bulletin* 112 (1): 155–159. https://doi.org/10.1037/0033-2909.112.1.155.

Coyle, Y., J. Cánovas Guirao, and J. Roca de Larios. 2018. "Identifying the Trajectories of Young EFL Learners across Multi-Stage Writing and Feedback Processing Tasks with Model Texts." *Journal of Second Language Writing* 42:25–43. https://doi.org/10.1016/j.jslw.2018.09.002.

Escalante, J., A. Pack, and A. Barrett. 2023. "AI-Generated Feedback on Writing: Insights into Efficacy and ENL Student Preference." *International Journal of Educational Technology in Higher Education* 20 (1): 57–20. https://doi.org/10.1186/s41239-023-00425-2.

García Mayo, M. P., and U. Loidi Labandibar. 2017. "The Use of Models as Written Corrective Feedback in English as a Foreign Language (EFL) Writing." *Annual Review of Applied Linguistics* 37:110–127. https://doi.org/10.1017/S0267190517000071.

Godwin-Jones, R. 2022. "Partnering with AI: Intelligent Writing Assistance and Instructed Language Learning." *Language Learning & Technology* 26 (2): 5–24. https://doi.org/10.10125/73474.

Godwin-Jones, R. 2024. *Distributed Agency in Second Language Learning and Teaching through Generative AI*. https://doi.org/10.48550/arxiv.2403.20216.

Goulart, L., M. L. Matte, A. Mendoza, L. Alvarado, and I. Veloso. 2024. "AI or Student Writing? Analyzing the Situational and Linguistic Characteristics of Undergraduate Student Writing and AI-Generated Assignments." *Journal of Second Language Writing* 66:101160. https://doi.org/10.1016/j.jslw.2024.101160.

Guo, K., and D. Wang. 2024. "To Resist It or to Embrace It? Examining ChatGPT's Potential to Support Teacher Feedback in EFL Writing." *Education and Information Technologies* 29 (7): 8435–8463. https://doi.org/10.1007/s10639-023-12146-0.

Hanaoka, O. 2007. "Output, Noticing, and Learning: An Investigation into the Role of Spontaneous Attention to Form in a Four-Stage Writing Task." *Language Teaching Research* 11 (4): 459–479. https://doi.org/10.1177/1362168807080963.

Hirvela, A. 2017. "Argumentation & Second Language Writing: Are We Missing the Boat?" *Journal of Second Language Writing* 36:69–74. https://doi.org/10.1016/j.jslw.2017.05.002.

Hsieh, H.-F., and S. E. Shannon. 2018. "Content Analysis." In *The Sage Encyclopedia of Educational Research, Measurement, and Evaluation*, edited by B. B. Frey, 393–394. Thousand Oaks: Sage.

Jiang, F., and K. Hyland. 2025. "Does ChatGPT Argue Like Students? Bundles in Argumentative Essays." *Applied Linguistics* 46 (3):375–391 https://doi.org/10.1093/applin/amae052.

Kang, E. Y. 2020. "Using Model Texts as a Form of Feedback in L2 Writing." *System* 89:102196–11. https://doi.org/10.1016/j.system.2019.102196.

Kang, E. Y. 2023. "The Role of Working Memory in the Effects of Models as a Written Corrective Strategy." *International Review of Applied Linguistics in Language Teaching* 61 (4): 1725–1753. https://doi.org/10.1515/iral-2021-0207.

Kang, E. Y. 2024. "EFL Learners' Perceptions and Their Association with the Effectiveness of Model Texts as a Feedback Tool." *Innovation in Language Learning and Teaching* 18 (1): 47–58. https://doi.org/10.1080/17501229.2023.2226144.

Kohnke, L., B. L. Moorhouse, and D. Zou. 2023. "ChatGPT for Language Teaching and Learning." *RELC Journal* 54 (2): 537–550. https://doi.org/10.1177/00336882231162868.

Koltovskaia, S., P. Rahmati, and H. Saeli. 2024. "Graduate Students' Use of ChatGPT for Academic Text Revision: Behavioral, Cognitive, and Affective Engagement." *Journal of Second Language Writing* 65:101130. https://doi.org/10.1016/j.jslw.2024.101130.

Lin, S., and P. Crosthwaite. 2024. "The Grass Is Not Always Greener: Teacher vs. GPT-Assisted Written Corrective Feedback." *System* 127:103529. https://doi.org/10.1016/j.system.2024.103529.

Luquin, M., and M. P. García Mayo. 2024. "A Longitudinal Study of the Effects of Model Texts on EFL Children's Written Production." *System* 120:103190. https://doi.org/10.1016/j.system.2023.103190.

Ma, Q., P. Crosthwaite, D. Sun, and D. Zou. 2024. "Exploring ChatGPT Literacy in Language Education: A Global Perspective and Comprehensive Approach." *Computers and Education. Artificial Intelligence* 7:100278. https://doi.org/10.1016/j.caeai.2024.100278.

Neuendorf, K. A. 2017. *The Content Analysis Guidebook*. Thousand Oaks: Sage.

Nguyen, L., and J. S. Barrot. 2024. "Detecting and Assessing AI-Generated and Human-Produced Texts: The Case of Second Language Writing Teachers." *Assessing Writing* 62:100899. https://doi.org/10.1016/j.asw.2024.100899.

Nguyen, L. Q., P. Dao, and B. T. T. Nguyen. 2024. "Model Texts as a Feedback Instrument in Second Language Writing: A Systematic Review." *Language Teaching Research* 1–26. https://doi.org/10.1177/13621688241291336.

Nguyen, L. Q., B. T. T. Nguyen, and H. Y. Phuong. 2024. "Exploring the Use of Model Texts as a Feedback Instrument in Expository Writing: EFL Learners' Noticing, Incorporations, and Text Quality." *Assessing Writing* 62:100890. https://doi.org/10.1016/j.asw.2024.100890.

Nguyen, L. Q., and D. V. Vu. 2024. "Exploring EFL Learners' Engagement and Draft Quality in a Multi-Stage Expository Writing Task Using Model Texts as a Feedback Facilitator: A Mixed-Methods Study." *Journal of Second Language Writing* 66:101161. https://doi.org/10.1016/j.jslw.2024.101161.

Ray, P. P. 2023. "ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations and Future Scope." *Internet of Things and Cyber-Physical Systems* 3:121–154. https://doi.org/10.1016/j.iotcps.2023.04.003.

Schmidt, R. 2001. "Attention." In *Cognition and Second Language Instruction*, edited by P. Robinson, 3–32. Cambridge: Cambridge University Press.

Steiss, J., T. Tate, S. Graham, J. Cruz, M. Hebert, J. Wang, Y. Moon, W. Tseng, M. Warschauer, and C. B. Olson. 2024. "Comparing the Quality of Human and ChatGPT Feedback of Students' Writing." *Learning and Instruction* 91:101894. https://doi.org/10.1016/j.learninstruc.2024.101894.

Su, Y., Y. Lin, and C. Lai. 2023. "Collaborating with ChatGPT in Argumentative Writing Classrooms." *Assessing Writing* 57:100752. https://doi.org/10.1016/j.asw.2023.100752.

Swain, M. 1985. "Communicative Competence: Some Roles of Comprehensible Input and Comprehensible Output in Its Development." In *Input in Second Language Acquisition*, edited by S. Gass and C. Madden, 235–253. Rowley, MA: Newbury House.

Swain, M. 2000. "The Output Hypotheses and beyond: Mediating Acquisition through Collaborative Dialogue." In *Sociocultural Theory and Second Language Learning*, edited by J. Lantolf, 97–114. Oxford: Oxford University Press.

Tieu, L. T. C., and J. R. Baker. 2023. "Using Model Essays in Conjunction with Noticing as a Feedback Instrument in IELTS Writing Preparation." *Innovation in Language Learning and Teaching* 17 (2): 380–392. https://doi.org/10.1080/17501229.2022.2049799.

Tsai, C.-Y., Y.-T. Lin, and I. K. Brown. 2024. "Impacts of ChatGPT-Assisted Writing for EFL English Majors: Feasibility and Challenges." *Education and Information Technologies* 29 (17):22427–22445. https://doi.org/10.1007/s10639-024-12722-y.

UCLES. 2001. *Oxford Quick Placement Test*. Oxford: Oxford University Press.

Weng, X., and T. K. F. Chiu. 2023. "Instructional Design and Learning Outcomes of Intelligent Computer Assisted Language Learning: Systematic Review in the Field." *Computers and Education. Artificial Intelligence* 4:100117. https://doi.org/10.1016/j.caeai.2022.100117.