

## Comparisons of AI and Human Writing Feedback: Student Perceptions in a University EFL Workshop

Nicholas Moore (CHA University)  
Christian Kim (CHA University)

Received: 3 April 2025  
Revised: 20 April 2025  
Accepted: 26 April 2025

Moore, Nicholas & Kim, Christian. (2025). Comparisons of AI and human writing feedback: Student perceptions in a university EFL workshop. *Modern English Education*, 26, 230-241.

### Keywords

AI writing feedback,  
writing instruction, ChatGPT,  
student perception  
AI 글쓰기 피드백, 글쓰기  
지도, ChatGPT, 학습자 인식

### Nicholas Moore (First author)

Assistant Professor  
College of Liberal Arts  
CHA University  
PhD candidate  
Department of English Education,  
Korea University  
[nicholas@cha.ac.kr](mailto:nicholas@cha.ac.kr)

### Christian Kim (Corresponding author)

Associate Professor  
College of Liberal Arts  
CHA University  
[conestar@cha.ac.kr](mailto:conestar@cha.ac.kr)  
ISNI: 0000 0004 7781 2015

\* This work was supported by the industry academic cooperation foundation fund, CHA University Grant 202300200001.

### Abstract

This study examined the perceptions of university-level EFL students toward feedback on their persuasive writing from humans and artificial intelligence (AI). Post-feedback student writing was also analyzed to find any measurable differences attributable to the feedback sources. Students were not provided with the feedback sources in order to eliminate potential bias. The study was conducted through an asynchronous writing workshop with 32 students receiving feedback on a first draft of a writing assignment. They were then asked to produce a second draft and respond to a ten-item Likert questionnaire about perceived dimensions of clarity, motivation, and usefulness of the feedback. Second draft samples were analyzed for grammar and usage mistakes, readability, and general proficiency. Independent samples t-tests were performed, and no statistically significant differences were found between the groups on any measure, despite several medium to large effect sizes in four survey responses, three in favor of AI feedback. The results are discussed with regard to AI's potential for integration into writing instruction. The importance of teaching skeptical attitudes towards AI-generated material is underscored, and further research is encouraged to investigate effects beyond the short term.

## INTRODUCTION

The integration of generative AI tools such as ChatGPT with education has raised new questions in a broad range of fields, from ethics to genetic engineering. The role of providing students with formative feedback on their writing has been deemed necessary to advance their writing quality and to foster self-efficacy (Wilson & Roscoe, 2020). However, educators often struggle to provide individualized feedback at scale due to the substantial time commitment this requires (Applebee & Langer, 2011). Automated Writing Evaluation (AWE) systems were proposed as one possible solution, but these systems have tended to focus more on length, grammar, and surface-level corrections rather than on deeper writing concerns, such as argumentation strength, discourse coherence, and organization (Grimes & Warschauer, 2010; Wilson et al., 2021).

The emergence of generative AI has presented English educators with a new opportunity to provide students with convenient and adaptive feedback that seems to break through those limitations of AWE (Gayed et al., 2022). Compared to AWE systems, ChatGPT generates satisfactorily human-like responses adapted to user input. This process can likely benefit from, but not require, prior training on engineering optimized writing prompts. Studies suggest that students perceive AI-generated feedback as beneficial in varied educational contexts (Carrier et al., 2024; Cho & Choe, 2024; Elwood & Bode, 2014; Liang et al., 2024; Timson & Grow, 1999). Some studies report positive effects on writing motivation and engagement (Chan & Hu, 2023). However, research into student perceptions of AI-generated feedback, particularly as compared to human feedback in blind experiments, remains limited.

This study focuses on ChatGPT and compares the perceptions of university-level EFL learners who received feedback from large language models (LLM) or human university instructors during an asynchronous writing workshop. Specifically, it examines whether those students randomly assigned to receive AI-generated feedback find that feedback as useful, motivating, and actionable as the students in the control group found their human feedback. As opposed to prior research where students knew they were interacting with AI tools (Gayed et al., 2022), participants in this study were not informed whether their feedback was AI- or human-generated to eliminate potential bias.

The findings of this study are anticipated to contribute to the growing discourse on AI in writing instruction. If it can be more conclusively shown that students perceive no difference between human and AI feedback in terms of effectiveness and motivation, AI-assisted feedback could be considered for supplementation in EFL contexts where educator resources are limited. Finally, this paper expands on the unclear results to discuss the practical integration of generative AI into university writing instruction, which can help educators to determine when and how AI tools might best encourage learning.

## LITERATURE REVIEW

### Historical Background of AI-Assisted Writing in L2 Writing Learning

Artificial Intelligence has long been integrated into language learning, especially in the fields of writing instruction and feedback. The earliest AI-powered writing tools date back to the 1960s with AWE systems. These were designed to assess writing quality and provide very basic corrections (Shermis & Burstein, 2002). AWE systems such as Project Essay Grade (PEG) and e-rater utilized statistical models to analyze patterns in their linguistic input and to assign scores based on human-graded reference samples (Page, 2003). These tools were often effective in speeding up grading, but they were limited in providing writing feedback more substantive or nuanced than surface-level corrections.

AWE systems possess high accuracy in scoring (Wilson & Roscoe, 2020). However, they do not provide meaning-focused feedback, which is necessary for improving writing proficiency. Many AWE tools tend to prioritize length and syntactic complexity over aspects like clarity and strength of argumentation (Bridgeman & Ramineni, 2017; Wilson et al., 2021). The inability of AWE tools to judge high-level factors in writing, like argumentation, is a shortcoming that students have learned to exploit, and this is a justification for giving particular attention to AI and persuasive student writing. Research suggested that students in higher-resource school districts can get used to the style of writing assessment given by AWE, and can even use machine-tricking strategies, rendering the tools less effective in promoting deeper learning (National Council of Teachers of English, 2013, p. 3).

The field of AI has recently been revolutionized by the development of LLMs, bringing about significant advancement in what is possible for AI-assisted writing. Whereas AWE requires pre-trained models on specific prompts, LLMs can generate dynamic and adaptive feedback customized to user input. If LLMs are integrated into education, this advancement makes possible a more personalized and contextual approach to writing instruction (Cho & Choe, 2024; Wilson et al., 2021). Educators can begin to scrutinize whether ChatGPT, or another language model, will break past the limitations of AWE and

fill a pedagogical need for detailed and adaptive feedback in language learning.

## Previous Studies on AI-Assisted Writing Workshops

Writing workshops integrating AI feedback have gained attention as potential learning environments where students can revise and refine their work (Mun, 2024; Song & Song, 2023). Repeated interactions with a chatbot alone can enhance students' perceptions of writing tasks and boost confidence (Gayed et al., 2022). Unlike standard AWE-based classroom settings, AI-assisted writing workshops encourage students to engage with feedback in a more self-directed manner, allowing them to assess and apply suggestions independently. Students polled for a study after using generative AI technologies say that they enjoy the personalized feedback and have come to expect learning resources that are available 24/7 (Chan & Hu, 2023). Students can assess and apply, on their own, suggestions given by a self-directed learning resource that they feel is on the cutting edge of educational service.

One previous study on student perceptions of the AI feedback tool Grammarly found that students have positive associations with AI technology as a writing tool, especially under two top conditions- if they already think it is useful for their writing, and if they perceive that the tool is already popular (Liang et al, 2024). Another found that students perceive Grammarly as very useful mainly for catching lower-level errors, while, on the other hand, they recognize a potential for over-reliance (Kim & Han, 2021). ChatGPT has specifically been studied for its ability to teach higher-level concerns. A recent study by Hong and Shin (2024) saw a group of EFL writing students improve their effective usage of conjunctive adjuncts as compared to a control group that utilized peer feedback only.

Perceptions of ChatGPT used in English lessons is occasionally mixed, depending on the student populations, or the designs of the studies or lessons. One ten-session ESP course that balanced ChatGPT-based activities with individualized coaching saw instructor evaluations rise most among the lowest-level students, while middle-tier students experienced the largest increase in perceived self-efficacy, reflecting changes in “personal confidence rather than just objective learning gains” (Cho & Choe, 2024, p. 471). Another study on 53 Korean undergraduate students found that, after completing a course integrating AI-assisted writing into the curriculum, the students gained confidence and were motivated to use AI in their English writing assignments (Carrier et al., 2024). There is, however, a chance that this confidence is misapplied. A skeptical approach may be required to avoid mistakes and accidental plagiarism. Burkhard (2022) found that about 61% of the 266 freshman students enrolled in a Swiss introductory academic writing course had an overly confident and unreflective attitude towards AI-powered writing tools, putting them at risk of making such mistakes.

In summary, previous research has found that students recognize the ability of AI tools to catch lower-level errors in their writing. There is also potential for AI to be used in conjunction with peer or instructor feedback to improve higher-level skills. Students seem to recognize some benefits to integrating AI into their writing processes, boosting their comfort levels and motivation. Confidence, however, does not automatically result in objective gains in writing ability. In fact, confidence can be inappropriate, especially if it is not accompanied by a conscientious attitude. AI technology remains prone to errors and plagiarism. The current study continues these lines of inquiry by attempting to measure both subjective (motivation, confidence) and objective (readability, usage errors) changes among Korean university students when they are not aware AI is being applied.

## The Role of Feedback in Writing Instruction

In writing instruction, feedback given to students can be divided broadly into two types- form-focused (e.g., grammar, spelling) and content-focused feedback (e.g., argumentation, organization) (Williams, 2005). Previous studies have found that students often prefer actionable feedback that directly points out mistakes (Ferris & Roberts, 2001), but more indirect feedback that motivates students towards self-revision is also valuable (Timson & Grow, 1999).

Studies have found that, historically, AI feedback is especially effective at focusing on low-level grammar, vocabulary, and spelling mistakes (Elwood & Bode, 2014). Furthermore, EFL students have been shown to prefer such details and explicit feedback on the level of grammar, while at the same time either tolerating or resisting higher-level concerns like argument development (Radecki & Swales, 1988). Such noted preferences align with the surface-level feedback AI can tirelessly provide, which means there may be a favorable coincidence of available tools and consumer preferences, though at the expense of occasionally necessary structural feedback.

The amount of time passing between the submission of student work and the reception of feedback also plays an important role in student interest and development. It has been shown that students benefit highly from timely feedback to assessments, since penalties to motivation can arise the longer feedback is delayed (Black & Wiliam, 1998). With the instantaneous nature

of AI feedback, its integration into writing pedagogy is a field that continues to develop, alongside the testing of new AI models for accurate suggestions (Tate et al., 2024).

## Orientation of the Current Study

From the perspectives of educators in the literature, human and AI feedback seem to have different strengths. Human feedback seems more accurate for higher-level concerns. AI remains an important resource to consider for immediate feedback with a focus on lower-order mistakes, such as grammar. To provide insight into both types of mistakes, it is beneficial to examine genres of writing that utilize both, such as persuasive writing and IELTS Academic Writing Task 2.

While previous studies explore some aspects of the usefulness or perceptions of AI, the extent to which AI and human feedback are directly compared is comparatively underexplored. The current study aims to fill this gap by surveying students under both conditions for how they perceive ChatGPT feedback and human feedback in terms of value, actionability, and motivation. Specifically, this study aims to answer the following research questions: 1) Do EFL students perceive any differences in the quality, clarity, and motivational impact between AI and human writing feedback? 2) Does the source of feedback (AI or human) lead to measurable differences in students' revision of their writing? 3) Can AI-generated feedback be used effectively in university EFL writing instruction?

## METHODOLOGY

### Participants

A total of 32 Korean undergraduate students voluntarily participated after finding a recruitment notice located on a shared university noticeboard. Notice was also given in all classes taught by native speakers, and students were informed that a small gift certificate would be given to those who completed two writings and a survey. All participants were enrolled in English writing courses at the freshman or sophomore level at the time and had varying levels of English proficiency. The courses are part of a general academic English curriculum that includes descriptive and argumentative writing. 10 were male and 22 were female. Ages ranged from 18 to 22. None of the students were majoring in English; the most common majors were nursing, pharmacy, and an undifferentiated general education major that most students are required to have during their first year of study.

When potential participants followed the recruitment notice, it linked to a consent notice, and finally to the sign-up sheet. Participants were assigned to a human feedback (control) group or an AI feedback (experimental) group based on the order in which they signed up. The group assignment changed every two students. For example, participants 1 and 2 were assigned to the human feedback group, and participants 3 and 4 to the AI feedback group. This method ensured a balanced and randomized distribution. Group assignment procedures were not disclosed to participants. Gender representation was even across the two groups.

To maintain ethical considerations, students were informed in advance, in writing, of how their data would be anonymized, and of their right to withdraw from the study at any time. Students gave signed consent for their responses to be used in this paper upon completing the survey. Data was stored on a university-provided, password-protected computer, following the university's research ethics guidelines.

### Procedure

The workshop utilized asynchronous timing, and was conducted via email. Each participant was sent one IELTS-style writing prompt, received feedback after submitting their writing sample, and then was assigned another similar prompt to give them a chance to apply what they learned. The same two prompts were used with all participants, but the order was randomized to control for any difficulty-related effect. Odd-numbered participants were assigned Prompt A first, followed by Prompt B, while even-numbered participants followed the reverse order. The prompts asked students to respond to societal issues using approximately 250 words:

(Prompt A) "The average standard of people's health is likely to be lower in the future than it is now." To what extent do you agree or disagree with this statement?

(Prompt B) "Many people go through life doing work that they hate or have no talent for." Why does this happen?

What are the consequences of this situation?

Prompt A asks for an opinion and Prompt B for a description of a real or imagined situation, in line with the students' writing classes, allowing students to practice what they had recently learned. There was a 48-hour time limit given in the instruction email. As mentioned, the second step in the process was for participants to receive feedback from either humans or AI, depending on their randomly-assigned group. Samples of human feedback and AI feedback can be found in Appendices 2 and 3, respectively. After receiving feedback, students were given their second prompt. Upon submitting the second writing assignment, students were directed to an online, post-workshop questionnaire. The survey was composed of ten items with 5-level Likert scale possible responses, along with the option to add qualitative explanations.

The five-point scale was chosen to allow students to express partial and full agreement and disagreement with the items. Seven levels were considered, but not chosen since the increased mental load of deciding partial agreement may lead to a lower survey completion rate. Ten survey items were constructed to capture various perspectives towards feedback that are not necessarily included when asking about overall like or dislike. Items 1 to 3 address surface-level features such as clarity, detail, personalization, criteria commonly found and emphasized in formative feedback literature (King et al., 2008; Morris & Chikwa, 2016). Items 4 and 5 distinguish between lower- and higher-order writing concerns. Items 6 and 7 ask about immediate practicality and implementation, while 8 to 10 address long-term value and classroom integration concerns. The survey is attached as Appendix 1.

## Data Analysis Procedures

Students in the experimental group received AI feedback from ChatGPT 3.5, and the other half was given human feedback from the authors' EFL instructor colleagues. These human raters have many years of experience and advanced degrees in TESOL or related fields. Feedback was minimally prompted to better match self-studying conditions, and to favor neither AI (with well-worded prompts) nor humans (with clear objectives or personal knowledge of the students). No specific guidelines were given except for the broad, untrained prompt: "Please give me feedback on my writing." This means that everything from spelling to the coherence of the student writing was a possible target for feedback, as long as the human raters could work within approximately 15 minutes, though this was not strictly regulated.

Once the Likert scale student responses were collected, other tests were also applied to help determine whether the student writing samples measurably changed after feedback. The writing software Grammarly was used to measure grammar and usage errors. In order to categorize all writing samples according to IELTS band descriptors, ChatGPT 4 was chosen over version 3.5 since its capabilities are more suitable for this task. Koraiishi (2024a) studied model 4 for evaluating IELTS Task 2 writing and the LLM was found to assign the same band as trained human raters with a weighted kappa value of 0.811, indicating substantial agreement. Additionally, Flesch-Kincaid readability scores were derived based on word length, sentence length, and syllables per word.

Statistical analysis was conducted in Excel using independent samples t-tests to determine whether the mean differences across AI and human feedback conditions were significantly different. Effects sizes were calculated using Cohen's *d* for independent samples, where group means were divided by pooled standard deviation. For interpretation, this study followed the common benchmarks set out by Cohen (1988), classifying 0.2 and smaller as a small effect, 0.5 as medium, and 0.8 and higher as large.

## RESULTS

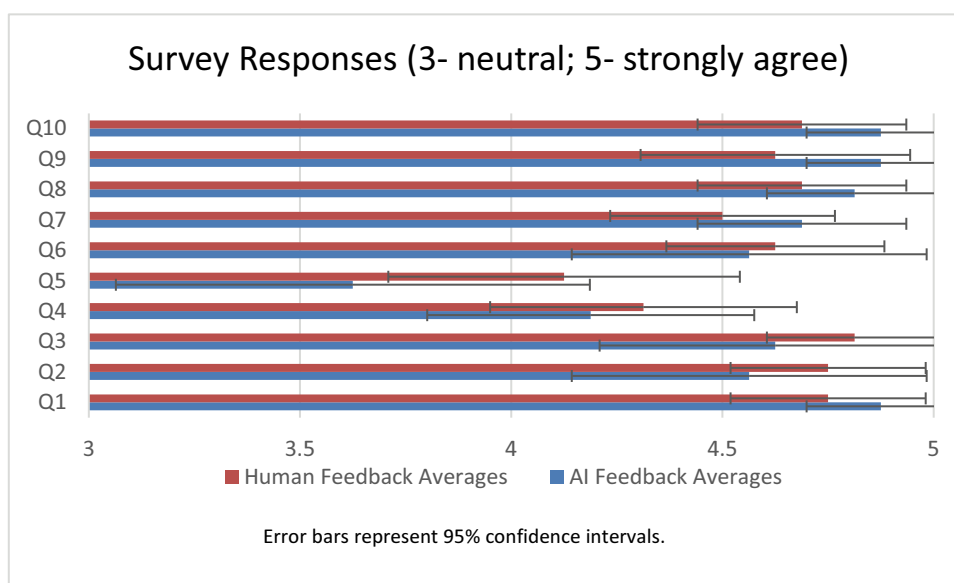
### Feedback Quality Analysis

No statistically significant differences were found among any of the survey items or the writing performance measures. The lowest *p*-value was 0.150, found in reference to Item 5 ("Before getting the feedback, I hadn't realized large-scale mistakes, like with my argument, organization, flow, and coherence"). Even in this case, however, the null hypothesis cannot be rejected, and *p*-values for the other measures were even higher. All *p*-values can be found in Figure 3. The lack of significant differences indicates that students found both AI-generated and human-generated feedback clear, useful, and motivating to a similar degree.

As shown in Figure 1, the average student perception of human feedback was higher than that of AI feedback in exactly half of the cases, whereas AI received better scores in the other half. AI feedback received equally high average scores of

4.88 in statements related to clarity (“The feedback was clear and easy to understand”) and motivation (“The feedback motivated me to improve my writing skills further” and “I want more of this kind of feedback in my English classes”). The highest average score received by human feedback was 4.81 on the statement about personalization: “I felt that the feedback was personalized to my writing style and needs.”

The largest gap between groups was .50 in favor of human feedback on the statement about how well large-scale mistakes were addressed (“Before getting the feedback, I hadn't realized large-scale mistakes, like with my argument, organization, flow, and coherence”). The second largest gap was .25 in favor of AI feedback in reference to the statement: “The feedback motivated me to improve my writing skills further.” However, response variability was considerable, with standard deviation in the responses ranging from .33 to 1.05. When 95% confidence intervals are applied, all the differences overlap considerably, evidencing once again that neither type of feedback was judged superior to any objective extent. Overall, while some patterns can be seen among average ratings, the answer to the first research question, “Do EFL students perceive any differences in the quality, clarity, and motivational impact between AI and human writing feedback?” cannot be answered in the affirmative.



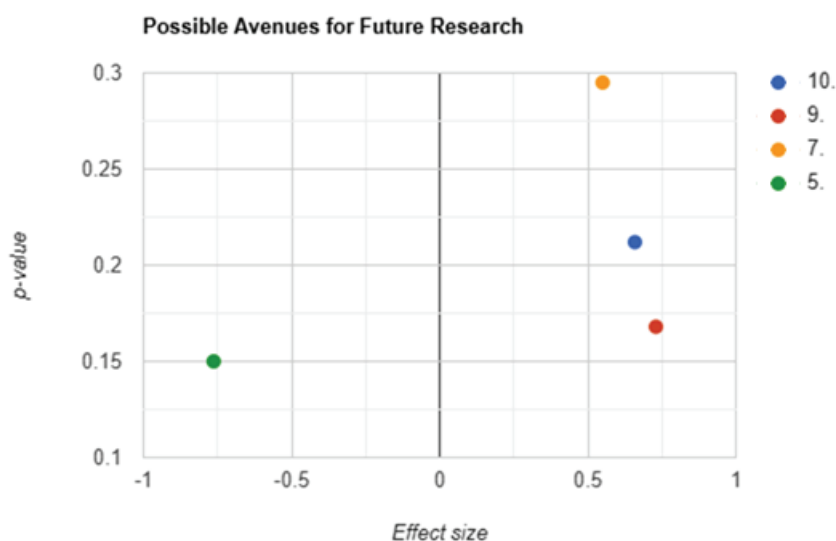
**FIGURE 1**

*Averages for All Post-Workshop Survey Responses*

Figure 2 charts the most significant findings using  $p$ -values and effect sizes, which were calculated by dividing the difference in group means by the pooled standard deviation (Cohen's  $d$ ). The largest effect size ( $d = 0.763$ ) was observed for Item 5 (“Before getting the feedback, I hadn't realized large-scale mistakes...”), favoring human feedback. Favoring AI feedback are Items 7 (“The feedback gave me practical advice...”), 9 (“The feedback motivated me...”), and 10 (“I want more of this kind of feedback...”), showing moderate-to-large effect sizes ( $d = 0.550$ – $0.730$ ) in favor of AI feedback. Although these differences were not statistically significant (all  $p > .15$ ), the magnitude of the effect sizes may suggest patterns worthy of future research. For example, AI feedback may be seen as more useful and motivating, while human feedback may be better at drawing attention to higher-order writing issues. The current dataset, however, may be too limited and can only justify the finding that AI and human feedback are similarly perceived.

## Performance Impact Assessment

To assess whether either type of feedback had a better impact on writing improvement, IELTS band descriptors were applied to both pre- and post-feedback writing samples by the more advanced AI model, ChatGPT 4. The AI feedback group saw a mean improvement in their band score of .344, and the human feedback group demonstrated an improvement of .125. Though the average improvement in the AI group was nearly three times the improvement in the human group, it did not reach a level of statistical significance when an independent samples t-test was run ( $p = .243$ ,  $d = .615$ ). This lack of significance adds some evidence that the types of feedback were comparably effective at improving student writing.



**FIGURE 2**  
*Survey Items with the Largest Group Differences*

Likewise, no meaningful distinctions were found through measuring differences in grammar/usage mistakes or Flesch-Kincaid readability scores. Both groups saw improvements in both measures, however, the negligible *p*-values of greater than .91 indicate that the AI and human feedback could not be found different in these regards.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Difference in IELTS band ratings	Difference in Flesch-Kincaid readability scores	Difference in grammar and usage errors
AI Averages	4.875	4.563	4.625	4.188	3.625	4.563	4.688	4.813	4.875	4.875	0.344	0.313	0.875
Standard Deviation ( $\sigma$ )	0.331	0.788	0.781	0.726	1.053	0.788	0.464	0.390	0.331	0.331	0.551	14.844	4.923
AI Confidence Interval (95%)	0.176	0.420	0.416	0.387	0.561	0.420	0.247	0.208	0.176	0.176	0.294	7.910	2.623
AI range - max	5.051	4.982	5.041	4.574	4.186	4.982	4.934	5.020	5.051	5.051	0.637	8.222	3.498
AI range - min	4.699	4.143	4.209	3.801	3.064	4.143	4.441	4.605	4.699	4.699	0.050	-7.597	-1.748
Human Averages	4.750	4.750	4.813	4.313	4.125	4.625	4.500	4.688	4.625	4.688	0.125	0.250	0.688
Standard Deviation ( $\sigma$ )	0.433	0.433	0.390	0.682	0.781	0.484	0.500	0.464	0.599	0.464	0.451	15.947	5.108
Human Confidence Interval (95%)	0.231	0.231	0.208	0.363	0.416	0.258	0.266	0.247	0.319	0.247	0.240	8.498	2.722
Human range - max	4.981	4.981	5.020	4.676	4.541	4.883	4.766	4.934	4.944	4.934	0.365	8.748	3.409
Human range - min	4.519	4.519	4.605	3.949	3.709	4.367	4.234	4.441	4.306	4.441	-0.115	-8.248	-2.034
Difference in Avg	0.125	-0.188	-0.188	-0.125	-0.500	-0.063	0.188	0.125	0.250	0.188	0.219	0.063	0.188
Pooled Standard Deviation	0.272	0.450	0.436	0.498	0.656	0.462	0.341	0.303	0.342	0.285	0.356	10.893	3.547
Effect size	0.459	-0.417	-0.430	-0.251	-0.763	-0.135	0.550	0.413	0.730	0.659	0.615	0.006	0.053
p-value	0.381	0.426	0.412	0.630	0.150	0.795	0.295	0.431	0.168	0.212	0.243	0.991	0.919
t-statistic	0.889	-0.808	-0.832	-0.486	-1.477	-0.262	1.065	0.799	1.414	1.275	0.000	1.239	1.190

Note. Highlighted results are displayed in Figure 2.

**FIGURE 3**  
*All Derived Statistics*

### Qualitative Responses

Qualitative responses in the survey were optional, and the number received was minimal ( $n = 5$ ). The comments were uniformly positive. Of note is that four responses came from students assigned to receive human feedback while only one came from a student who got AI feedback. The greatest level of engagement occurred in Q7, “This kind of feedback is quite valuable to me.” Students wrote statements such as “It’s my first time when I write my opinion in English. Through this opportunity, I actually learned many things about writing.” and “[This feedback] gave me the courage to do it alone.” The difference in the number of responses is insufficient for formal analysis.

## DISCUSSION

### Interpretation of Results

The primary focus of this study was on whether EFL university students would perceive writing feedback from AI and experienced human instructors differently. In response to the first research question, quality, clarity, and motivational impact were rated similarly across both conditions. No statistically significant differences were found, suggesting that students found neither source of feedback preferable in any dimension asked by the survey items. This finding may indicate that students could not detect the source of the feedback, or that they did not regard the source as affecting the quality of their writing workshop experience.

Small measurable improvements in students' revision of their writing were found (RQ2) in terms of grammar, readability, and IELTS-based writing measures. Interestingly, the largest effect size was found in the IELTS band category, which has the smallest range of possible outputs (band 1-9) in favor of AI feedback. None of these improvements, however, reached the level of statistical significance. Writing skills are developed through consistent practice, and with only one iteration of feedback and student writing adjustment, this result may not be surprising.

### Pedagogical Implications

This study has found that polled students likely rate AI feedback as equally useful and motivating as that from highly educated and experienced human instructors. This result recalls previous studies which found that AI-assisted writing tools can aid in engagement and self-efficacy (Chan & Hu, 2023; Wilson & Roscoe, 2020). The seeming parity in usefulness and motivation combats a common perception among educators- that automated feedback tools could have a negative impact on student motivation (Easter et al., 2024).

The near-instantaneous nature of AI feedback must be one of its most appealing features. While human feedback takes several minutes to write out, and delivery takes longer, AI feedback is delivered on demand, providing the possibility of a back-and-forth educational process with unlimited iterations. Research has demonstrated that timely feedback is essential to student learning (Black & Wiliam, 1998), and now AI is poised to fill in gaps in formative assessment. AI tools can be particularly useful at scale, such as in large EFL classrooms where trained instructors often have to struggle with time constraints.

More theoretically, these implications for iterative learning align with the foundations of modern educational theory, in which we scaffold student development through attainable and interactive challenges using activities, tools, and feedback mechanisms (Vygotsky, 1978). This study found that a minimal prompt ("Please give me feedback on my writing.") was enough to initiate an effective feedback mechanism, showing the potential for writing development scaffolding to be within the means of students individually, without reliance on paid instructors.

### Integration Challenges and Solutions

The convenience and potential of AI feedback exists alongside evidenced concerns about accuracy and overreliance. Hallucinations are a common issue, where AI generates correct-sounding, but ultimately misleading or wrong information by predicting text where it lacks specific training (Easter et al., 2024). Human instructors necessarily ground their observations and recommendations within their theoretical training and experience, whereas AI, as a generative language model, cannot process words in that way, often leading to high-order reasoning mistakes in argument, coherence, or tone. Higher-order mistakes can happen in response to a single query, but they may be more common when AI is pushed to provide multiple and deeper responses into the same topic. This study did not address AI mistakes or user perceptions of mistakes, which remain an open area for inquiry.

There also exists a tendency to over-rely on AI tools, leading to passive learning behaviors in place of active engagement with one's own skill development (Godwin-Jones, 2022). It seems likely that the risk of passivity developing would also occur over time with repeated inquiries. Though the risk is unclear, it indicates that longer-term iterative writing experiments may encounter more obstacles to learning than this workshop could examine as a one-time use of AI. It may be prudent for educators to not overly rely on AI feedback, instead using it to supplement human interaction.

A hybrid model of feedback would be a promising application of the currently known limitations of AI feedback. The immediacy of AI can be leveraged in students' free time, and then class or consultation time can be used for human commentary. The richness and reliability of that commentary, on an individual or class level, can catch higher-level errors.

Engaging socially may also provide a motive for engagement. Additionally, teachers should be trained in AI literacy for themselves and in order to pass that training on to students.

## CONCLUSION

This study found no significant differences between university-level EFL students' perceptions of AI-generated versus human feedback in an asynchronous email-based writing workshop. Both types of feedback were rated highly by participants in response to survey items inquiring about feedback clarity, usefulness, and motivational impact. This is a potentially promising result for EFL writing educators who anticipate AI-integration. AI or AI-assisted feedback could potentially perform as a supplement to limited and relatively impersonal, in-class writing instruction, filling instructional gaps, as in Cho and Choe (2024). It is important to note, however, that the absence of significant gaps in responses in this single case is sufficient to show that the two types of feedback are equally valuable in quality or effect. It is solely argued from these results that students did not perceive a meaningful distinction in this limited and particular context.

The most substantial limitation of this study is likely its small sample size ( $n = 32$ ) and limited scope. The study was conducted through email correspondence, within a single university in a South Korean province. Furthermore, the proffered feedback was in response to only a single piece of student writing each over a short timeframe, and so the findings may not generalize, either to other educational contexts or to students of different demographics, with different proficiency levels, or even with greater familiarity to more advanced AI technology. Given the rapid pace of AI development and adoption, the offerings of tools like ChatGPT are likely to evolve, and students' perceptions can change with each new iteration of the technology. Indeed, OpenAI's ChatGPT 3.5, which supplied the feedback in this study, is no longer in common use. Future research should seek to measure long-term effects of AI-assisted feedback on writing, address differences between LLMs, and solicit student opinions of AI in a range of learning environments that better match with classroom instruction and self-study. There is inconclusive evidence shown in Figure 2 which shows that it may be possible in the future to find a distinction between the types of feedback in terms of higher-level concerns, practicality, or motivation.

While the data in the current study has not provided definitive evidence in favor of either AI or human feedback, the discussion proposed and examined practical methods for the integration of AI into everyday writing instruction. As the data suggest that students may be open to AI-generated feedback without bias, educators with large numbers of students may endeavor to use AI technology to scale up their feedback procedures. Nevertheless, this recommendation comes with warnings. Reasonable and educated concerns remain surrounding AI's inaccuracies, and the potential of overreliance on AI, and a consensus can be found in the literature for more cautious, hybrid approaches that are less vulnerable to critical AI weaknesses (Easter, et al., 2024; Hong & Shin, 2023; Kim & Han, 2021). In the future, studies examining hybrid feedback models, ones which leverage both the efficiency of AI and the experience of humans, seem likely to give students a more comprehensive understanding of how to improve their writing skills.

While AI integration into EFL instruction may carry risks, the behavior of students in this study indicates that the short-term benefits are significant. A thoughtful integration of widely-available AI tools, like ChatGPT, into the classroom will likely enhance writing development. Future studies are called on to refine and test AI usage for accuracy, positive student engagement, and long-term skill enhancement.

## References

- Applebee, A. N., & Langer, J. A. (2011). "EJ" extra: a snapshot of writing instruction in middle schools and high schools. *The English Journal*, 100(6), 14–27.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Bridgeman, B., & Ramineni, C. (2017). Design and evaluation of automated writing evaluation models: Relationships with writing in naturalistic settings. *Assessing Writing*, 34, 62–71.
- Burkhard, M. (2022). *Student perceptions of AI-powered writing tools: Towards individualized teaching strategies*. Presented at the International Conference on Cognition and Exploratory Learning in Digital Age (CELDA), Portugal, 19, 73–81.
- Carrier, J., Heywood, D., & Hwang, Kyu-Hong. (2024). ChatGPT in the EFL writing classroom: Perceptions of AI chatbot assistance among Korean university students. *Journal of Language Sciences*, 31(3), 81–110.
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(43), 1–18.

- Cho, Kyung Hee & Choe, Mun-Hong. (2024). A study on the effectiveness of an ESP course for airline cabin services with aid of ChatGPT. *Modern English Education*, 25, 455–475.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.
- Easter, T., Lee, A., & Bailey, D. (2024). Artificial intelligence in language education: A mixed-methods study of teacher perspectives and challenges. *National Social Science Technology Journal*, 12(1), 203–234.
- Elwood, J. & Bode, J. (2014). Student preferences vis-à-vis teacher feedback in university EFL writing classes in Japan. *System*, 42, 333–343.
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10, 161–184.
- Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, 3, 100055. <https://doi.org/10.1016/j.caeai.2022.100055>
- Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology*, 26(2), 5–24.
- Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment*, 8(6).
- Hong, Seungjin, & Shin, Yu Kyoung. (2024). Exploring the efficacy of ChatGPT in improving L2 writers' use of English conjunctive adjuncts. *Modern English Education*, 25, 212–223.
- Kim, Hye-Kyung, & Han, Sumi. (2021). College students' perceptions of AI-based writing learning tools: With a focus on Google Translate, Naver Papago, and Grammarly. *Modern English Education*, 22(4), 90–100.
- King, D., McGugan, S., & Bunyan, N. (2008) Does it make a difference? Replacing text with audio feedback. *Practice and Evidence of Scholarship of Teaching and Learning in Higher Education*, 3(2), 145–63.
- Koraishi, O. (2024). The intersection of AI and language assessment: A study on the reliability of ChatGPT in grading IELTS Writing Task 2. *Language Teaching Research Quarterly*, 43, 22–42.
- Liang, J., Huang, F., & Teo, T. (2024). Understanding Chinese university EFL learners' perceptions of AI in English writing. *International Journal of Computer-Assisted Language Learning and Teaching*, 14, 1–16.
- Morris, C., & Chikwa, G. (2016). Audio versus written feedback: Exploring learners' preference and the impact of feedback format on students' academic performance. *Active Learning in Higher Education*, 17(2), 125–137.
- Mun, Chae-young. (2024). EFL learners' English writing feedback and their perception of using ChatGPT. *Journal of English Teaching through Movies and Media*, 25(2), 26–39.
- National Council of Teachers of English. (2013). *Machine scoring fails the test*. NCTE Position Statement, 1–12.
- Page, E. B. (2003). The Project Essay Grade: From simplex to complex. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 77–88). Routledge.
- Radecki, P. M., Swales, J. M. (1988). ESL student reaction to written comments on their written work. *System* 16(3), 355–365.
- Shermis, M. D., & Burstein, J. C. (Eds.) (2002). *Automated essay scoring: A cross-disciplinary perspective*. Routledge.
- Song, C. & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Tate, T. P., Steiss, J., Bailey, D., Graham, S., Moon, Y., Ritchie, D., Tseng, W., & Warschauer, M. (2024). Can AI provide useful holistic essay scoring? *Computers and Education: Artificial Intelligence*, 7, 100255. <https://doi.org/10.1016/j.caeai.2024.100255>
- Timson, S., & Grow, A. (1999). Error correction preferences of second language learners: A Japanese perspective. *JACET Bulletin*, 30, 135–149.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Williams, J. (2005). Form-focused instruction. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 671–691). Routledge.
- Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, 58(1), 87–125.
- Wilson, J., Huang, Y., Palermo, C., Beard, G., & MacArthur, C. A. (2021). Automated feedback and automated scoring in the elementary grades: Usage, attitudes, and associations with writing outcomes in a districtwide implementation of MI Write. *The International Journal of Artificial Intelligence in Education*, 31(2), 234–276.

## Appendix 1

### Writing Workshop Survey Items

1. The feedback was clear and easy to understand.
2. The feedback was detailed enough.
3. I felt that the feedback was personalized to my writing style and needs.
4. Before getting the feedback, I hadn't realized small-scale mistakes, like with spelling, vocabulary, grammar, and punctuation.
5. Before getting the feedback, I hadn't realized large-scale mistakes, like with my argument, organization, flow, and coherence.
6. After reading the feedback, I felt more confident about the second writing task.
7. The feedback gave me practical advice. I then used the advice on the second writing task.
8. This kind of feedback is quite valuable to me.
9. The feedback motivated me to improve my writing skills further.
10. I want more of this kind of feedback in my English classes.

## Appendix 2

### Sample Human Writing Feedback

This was great and interesting writing! It's clear that you spent a good deal of time thinking about the competitive natural work. That leads to people stressing out and berating themselves.

The best way to make your insight clear would be by breaking up the constituent thoughts into an organized, multi-paragraph essay. I recommend each point (competition, mental illness, and poor job mobility) have a dedicated body paragraph. Then, you can add an introduction, which should introduce your overall argument and preview these individual points. Finally, a conclusion restates and synthesizes what came before.

In terms of grammar, I recommend not beginning any sentence with a conjunction, like "So" or "But". Also, if there's a way to avoid "not" in multi-verb expressions, that would be better. The following is a list of corrections, beginning with examples of what I mean by these expressions:

I think people are sensitive for not keeping up with somebody. → I think people are sensitive about falling behind.

So, people choose to get a job and try not to be behind. → This is why they choose to get jobs and try to keep up.

...they are failed(extremely), and they will be left alone when others go forward the future. → ...they have failed, and will be stuck while others advance in their careers.

...and go to be exhausted finally. → ...and end up totally exhausted.

They will spend lethargic time about quarter of day. → They can sometimes waste up to a quarter of a day in a depressive slump.

Finally, in place of "Sequentially", I recommend "Consequently" (which you used before), or else "Therefore", "After this", "What's more", "Beyond this", or "Ergo".

Again, this was great overall. Please keep putting this kind of energy into your writing.

## Appendix 3

### Sample ChatGPT 3.5 Writing Feedback

Your writing addresses an important topic about individuals not being able to pursue work aligned with their talents and interests. Here are some feedback points to enhance your writing:

1. **Clarity and Organization:** The ideas are somewhat scattered throughout the text. Consider organizing your thoughts into clearer paragraphs with each addressing a specific aspect of the issue.
2. **Grammar and Punctuation:** There are several grammatical errors and punctuation issues in your writing. For instance, “Although some people have special talents, even they know that, can not try” could be revised to “Although some people have special talents, they may not be able to pursue them.”
3. **Depth of Analysis:** While you touch upon reasons why people might not pursue their talents (such as fear of failure and lack of resources), consider delving deeper into each reason. For instance, explore how societal pressures contribute to individuals not pursuing their passions.
4. **Personal Example:** Your personal example about not being able to pursue singing adds depth to your argument. Consider expanding upon it to illustrate the emotional impact of not being able to pursue one's talents.
5. **Solution-Oriented Approach:** You suggest solutions towards the end of your writing, such as implementing programs to help individuals discover their talents and providing financial support. Expand on these solutions and discuss how they could be implemented effectively.
6. **Language and Tone:** Ensure consistency in your tone throughout the writing. Your tone should be professional and empathetic, addressing the issue seriously while offering constructive solutions.

Here's a revised version incorporating some of these suggestions:

Many individuals find themselves unable to pursue work aligned with their talents and interests, often due to a variety of factors. One prominent reason is the societal pressure to conform to a certain pace of life, leading people to fear failure and avoid taking risks. Additionally, limited access to resources and opportunities can hinder individuals from exploring their passions. For example, I personally experienced this limitation when I was unable to pursue singing due to my circumstances.

The consequences of this situation are profound, leading to decreased individual fulfillment and societal well-being. Many individuals suffer from depression and lethargy as a result of not being able to pursue their passions. To address this issue, it is essential to implement programs aimed at helping individuals discover their talents and provide financial support to enable them to pursue their interests. By doing so, we can foster a society where individuals are empowered to pursue fulfilling careers aligned with their talents, leading to increased overall happiness and societal prosperity.

Overall, your writing addresses an important issue and offers valuable insights. By refining the organization, depth of analysis, and language, you can further enhance the effectiveness of your message.