Teacher- Versus AI-Generated (Poe Application) Corrective Feedback and Language Learners' Writing Anxiety, Complexity, Fluency, and Accuracy

Dan Wang
School of Foreign Studies, East China University of Political Science and Law, ShangHai, China

Abstract

This study examines the effects of corrective feedback (CF) on language learners' writing anxiety, writing complexity, fluency, and accuracy, and compares the effectiveness of feedback from human teachers with an AI-driven application called Poe. The study included three intact classes, each with 25 language learners. Using a quasi-experimental design with pretest and posttest measures, one class received feedback from the teacher, one from the Poe application, and the third received no response to their writing. Data were generated though tests and a writing anxiety scale developed for the study. Data analysis, conducted using one-way ANOVA tests, revealed significant effects of teacher and AI-generated feedback on learners' writing anxiety, accuracy, and fluency. Interestingly, the group that received AI-generated feedback performed better than the group that received teacher feedback or no AI support. Additionally, learners in the AI-generated feedback group experienced a more significant reduction in writing anxiety than their peers. These results highlight the remarkable impact of AI-generated CF on improving writing outcomes and alleviating anxiety in undergraduate language learners at East China University of Political Science and Law . The study demonstrates the benefits of integrating AI applications into language learning contexts, particularly by promoting a supportive environment for students to develop writing skills. Educators, researchers, and developers can use these findings to inform pedagogical practices and technological interventions to optimize the language learning experience in primary school settings. This research highlights the effectiveness of AI-driven applications in language teaching. It highlights the importance of considering learners' psychological well-being, particularly anxiety levels, when developing effective language learning interventions.

Keywords: artificial intelligence, AI, corrective feedback, writing anxiety, complexity, fluency, accuracy

Introduction

Providing written feedback to correct errors made by language learners has long been a fundamental practice in teaching writing, capturing considerable attention in writing research in Second Language (L2). Given the multifaceted and complex nature of writing, feedback encompasses a broad array of responses, offering insights into the accuracy, successful communication, and content of learners' expressions or discourse (Li & Vuono, 2019; Thi & Nikolov, 2021). Pedagogically, feedback is a crucial link between assessment and teaching, providing appropriate information about the language learners' correct performance and guidance to achieve the target learning goals. Consequently, considerable focus has been directed toward understanding the significant contribution of CF to the language learners' writing performance. Written Corrective Feedback (WCF) plays a pivotal role in enhancing learners' writing performance, accuracy, fluency, organizational skills, and task achievement (Karim & Nassaji, 2018, 2019; Leeman, 2010; Lim & Renandya, 2020; Liu & Brown, 2015; Liu & Huang, 2020; Luo & Liu, 2017).

Numerous studies have investigated the impact of written CF in the writing ability of foreign language learners, comparing various types of feedback (Han & Hyland, 2015; Truscott, 2010; Truscott & Hsu, 2008). The findings from these studies have been synthesized through quantitative and qualitative systematic reviews, commonly known as meta-analyses. Earlier meta-analyses, including the work of Russell and Spada (2006), underscored the significant contribution of CF to the development of grammatical knowledge in language learners.

In subsequent analyses conducted by Hyland and Hyland (2006), the emphasis was on the diversity evident in student populations, writing genres, feedback practices, and research designs. Notably, Truscott (2010) posited a negative impact of error correction on students' ability to accurately report information, drawing this conclusion from an examination of twelve published studies. In later meta-analyses conducted by researchers (e.g., Biber et al., 2011; Kang & Han 2015; Liu & Brown, 2015; Sia & Cheung, 2017), valuable perspectives were offered regarding the efficacy of written CF. These analyses considered individual differences among learners and addressed methodological limitations in the existing literature.

Recent meta-analyses by Lim and Renandya (2020) presented evidence supporting the potential of written WCF to enhance L2 writing skills of English as a Foreign Language (EFL) learners, with a specific focus on improving grammatical accuracy. However, ongoing debates persist, addressing questions about the extent of the benefits derived from WCF and the sustained efficacy of various feedback treatments, particularly in comparing implicit and explicit approaches. The existing body of literature on CF in language learning predominantly revolves around traditional teacher-generated feedback, with limited exploration into the consequences of feedback generated by AI. Notably, there is a restricted investigation into the impact of AI-generated feedback, primarily through applications, highlighting a gap in research in this area. While studies have examined the effectiveness of feedback on aspects such as accuracy, complexity, and anxiety, there is a noticeable gap in the research regarding a direct comparison between teacher-generated and AI-generated CF.

One specific instance of an AI-powered application is the Personalized Online Experience (Poe). This application integrates AI technologies to tailor online interactions based on individual user behavior, preferences, and historical data. The Poe application optimizes content recommendations, user interfaces,

and overall interaction design by continuously learning from user engagement. Through machine-learning algorithms, Poe evolves to anticipate user needs, delivering a more personalized and efficient online experience. This enhances user satisfaction and exemplifies the potential of AI-powered applications to revolutionize how we interact with digital platforms, creating a more intelligent and user-centric digital landscape.

Understanding these two feedback sources' potential differences and implications is essential for informing pedagogical practices and optimizing language learning experiences. The rationale for this study stems from the increasing integration of AI technologies in language education and the need to evaluate their efficacy compared to traditional teaching methods. With the emergence of AI applications such as Poe, which claim to provide nuanced and personalized corrective feedback, it is crucial to assess their impact on language learners. This study investigates how learners respond to feedback from AI applications compared to feedback from human teachers, specifically regarding writing anxiety, writing complexity, and accuracy. The rationale behind this comparative analysis lies in the potential benefits and drawbacks of AI-generated feedback, which may differ from the interpersonal and contextual aspects associated with teacher-generated feedback. By addressing this gap, the research seeks to contribute valuable insights into the evolving landscape of language education. More specifically, this study attempts to answer the following questions:

1. Do AI-generated (through Poe application) and teacher corrective feedbacks equally reduce the EFL learners' writing anxiety?

2. Do AI-generated (through Poe application) and teacher corrective feedbacks equally foster EFL learners' writing fluency?

3. Do AI-generated (through Poe application) and teacher corrective feedbacks equally foster the EFL learners' writing accuracy?

4. Do AI-generated (through Poe application) and teacher corrective feedbacks equally foster the EFL learners' writing complexity?

Literature Review

Empirical considerations do not solely drive the research on WCF; it is also underpinned by theoretical frameworks that highlight its potential contributions to L2 development (Polio, 2012). Skill-acquisition theories, such as DeKeyser's (2007), underscore the importance of practice and explicit instruction in developing accuracy, a concept aligned with WCF's role in aiding learners to store and retrieve declarative knowledge. The theoretical foundations of the noticing hypothesis (Schmidt, 2012) and the interaction hypothesis (Long, 1980) further support WCF by helping learners identify gaps in interlanguage by providing the needed evidence.

In L2 writing, the distinction between corrective and non-corrective feedback, focusing on form and content, is emphasized (Luo & Liu, 2017; Zhang, 2021). Corrective feedback targets language learning by

providing negative evidence to enhance accuracy, while non-corrective feedback addresses broader aspects such as content, organization, and linguistic performance. The role of WCF in L2 writing goes beyond traditional written commentary feedback, with feedback strategies ranging from direct to indirect and metalinguistic (Ellis, 2009a, 2009b, 2009c). The choice between comprehensive and focused feedback treatments has been explored, with recent studies highlighting the benefits of focused feedback. However, some still argue for a comprehensive approach (Benson & DeKeyser, 2018).

Considering the significance of writing tasks in L2 writing, factors such as task types and complexity play a vital role (Liu & Huang, 2020). Both unfocused and focused writing tasks are used for evaluating the language learners' writing proficiency. The impact of different writing task genres on language use, each with distinct communicative and functional requirements, contributes to a deeper understanding of how diverse writing tasks influence linguistic performance, encompassing aspects such as accuracy and complexity (Polio & Yoon, 2018).

Research into WCF is empirically driven and rooted in theoretical frameworks exploring its potential contributions to L2 development (Zhang, 2021). Skill-acquisition theories, like DeKeyser's (2007), posit that accuracy results from practice, explicit instruction, and extensive practice, prerequisites for transforming declarative knowledge into procedural knowledge. The concept of written CF is in harmony with these concepts, with the objective of aiding learners in storing and retrieving declarative knowledge, specifically explicit knowledge related to the target language. Stressing the importance of both practice and feedback, Evans et al. (2014) and Hartshorn and Evans (2015) highlighted automatization's crucial role within the skill acquisition theory framework.

In L2 writing, scholars highlight the role of WCF in fostering students' writing skills and abilities. The crucial distinction between corrective and non-corrective feedback, which focuses on form and content, is essential (Long, 1980; Luo & Liu, 2017). Corrective feedback aims at negative evidence to promote learning the target language, specifically addressing accuracy. Conversely, non-corrective feedback provides commentary on broader aspects, including organization, linguistic performance, and format. Exploring different types of WCF beyond traditional written commentary feedback is gaining interest.

Empirical investigations into WCF examine its facilitative role through the comparison of feedback strategies against no-feedback conditions (Ellis, 2009a, 2009b, 2009c; Kurzer, 2018) and assess the relative effectiveness of various feedback strategies (Riazantseva, 2012). Feedback interventions delineate between comprehensive and focused approaches, determining the extent of WCF provided to students. While earlier studies leaned towards comprehensive error correction, recent research underscores the advantages of focused feedback. Scholars (e.g., Benson & DeKeyser, 2018; Stefanou & Révész, 2015) propose that correcting the errors in a focused way yields more significant benefits than addressing all errors indiscriminately. However, certain studies advocate for comprehensive feedback that addresses a range of errors rather than concentrating on a specific sort of error (Bonilla López et al., 2018).

Ellis (2009c) and Robinson (2011) advocated for focusing on meaning in tasks, classifying them as unfocused or focused, based on general language use or specific linguistic features in L2 writing, in which both unfocused tasks and focused writing tasks serve as assessments of learners' proficiency. They also believe that recognizing how task demands influence L2 writing is paramount because tasks establish

contexts which provide opportunities to uptake CF. While the impact of cognitive demands imposed by tasks on learners' accuracy has received limited attention in WCF research, empirical investigations into various genres of writing tasks reveal distinct communicative and functional requirements.

Studies on AI

AI constitutes a domain with a rich historical and philosophical background (Bozkurt et al., 2023; Cao, 2023). Its evolution has raised fundamental inquiries about machine cognition and the capacity for independent creativity beyond programmed instructions (Kurzweil, 2014; Winterson, 2022, pp. 9–32). These inquiries led to the adoption of the concept of AI technologies (Benavides et al., 2020; Bozkurt et al, 2023; Winterson, 2022, pp. 9–32).

Despite these strides, AI has become so profoundly integrated into daily life that there are expectations of an era where human and artificial intelligence converge (Kurzweil, 2014). The ubiquitous influence of AI extends to communication and advisory roles across various professions, including media, accounting, and copywriting (Bozkurt et al., 2021). Since the inception of computerized AI, educators have expressed concerns about the potential obsolescence of their roles (Goksel & Bozkurt, 2019; Selwyn et al., 2023). More recently, apprehensions have emerged regarding the possibility of students completing assignments or responding to questions with undetectable AI assistance, raising concerns about academic integrity and the authenticity of students' work (Diebold, 2023; Luan et al., 2020; Ouyang et al., 2022).

The systematic exploration of AI technologies in educational settings predominantly revolves around their application for forecasting learner outcomes and behaviors which establish adaptive learning environments, improve academic performance, and enhance overall learning achievements and experiences (Chu et al., 2022; Zawacki-Richter et al., 2019). Recent examinations of literature in K–12 education reveal a broadening range of applications for artificial intelligence in education (AIED), including collaborative learning, modeling approaches, and visualization. This signifies a shift beyond conventional pedagogical methods, as noted by Humble and Mozelius (2022) and Zawacki-Richter et al. (2019). However, the multifaceted adoption of AI in education necessitates a comprehensive understanding of its potential implications within the broader social, cultural, pedagogical, and organizational contexts. Despite the potential advantages of integrating AI into education, numerous persistent challenges and ethical considerations warrant attention. These challenges encompass attitudes toward AI, educators' proficiency in effectively using technological tools, ethical concerns, and various technological hurdles (Sharma et al., 2019).

Ethical considerations form a cornerstone in the discourse surrounding AI in education. Scholars argue for a comprehensive examination of ethical concerns, including privacy issues and data ownership, before the widespread adoption of AI technologies (Humble & Mozelius, 2022). The potential influence of major ed-tech organizations over educational institutions raises additional ethical questions, particularly regarding privacy and corporate control, as these organizations may have access to student and staff data for corporate gains (Bozkurt et al., 2021).

Moreover, the existing literature on AI in education often consists of descriptive studies, indicating a need for a robust theoretical foundation to propel the field forward (Chen et al., 2020). Establishing a theoretical

framework is essential for advancing our understanding of the implications of AI in educational settings and guiding future research and implementation strategies. Therefore, a concerted effort to address these challenges and ethical considerations is imperative to harness responsibly AI's full potential in education.

In a focused examination of the AI-powered chatbot ChatGPT from an educational standpoint, Tlili et al. (2023) supported the use of ChatGPT in education. They advocated for a new teaching philosophy to effectively integrate AI-powered technologies into education, emphasizing the importance of responsible, humanized chatbots and the development of digital literacy competencies (Ng et al., 2021). Concerns about academic integrity have prompted discussions on AI tools' ethical and responsible use in education (Cox, 2021), with calls for updated policies and strategies. Researchers, instructors, and policymakers are cautioned to proactively address potential disruptions caused by integrating AI technologies (Tang et al., 2021). Additionally, it is acknowledged that novel assessment formats emphasizing creativity and critical thinking, areas where AI cannot entirely replace human judgment, may be essential (Dogan et al., 2023).

Methodology

Sample and Procedure

The study involved three intact classes of language learners enrolled in online courses at the School of Foreign Studies, East China University of Political Science and Law, China. Each class consisted of 25 members. As both researcher and instructor, I recruited participants from these classes, all taking a writing course as part of their language curriculum. In order to homogenize the language learners based on writing accuracy, fluency, and complexity, a writing test comprising three tasks was administered to the entire pool of participants. Following the initial evaluation, participants were kept on for the treatment; however, the final analysis only included those whose writing test results were within the range of +/- 1 standard deviation (SD) from the mean. The purpose of implementing this criterion was to guarantee a study group that was relatively homogeneous. The analysis comprised 75 language learners' pretest and posttest results, representing those who satisfied the predetermined requirements. All participants were native speakers of Mandarin, with English being their second language. The study focused on this specific population to explore the impact of the proposed treatment on the writing skills of Chinese English language learners at the University of X. In the initial phase of the study, three intact classes underwent a pre-assessment involving the administration of the Writing Anxiety Scale and a Writing Test. The classes were randomly assigned to three groups for the subsequent intervention. The first group received corrective feedback from the teacher, focusing on addressing issues in their written work. In the second group, students were trained to use the Poe application. This involved submitting their writings to Poe and requesting revisions, edits, and paraphrasing, specifically emphasizing grammar, accuracy, coherence, and complexity. The third group received no corrective written feedback during this intervention period.

Weekly assignments that matched the course material were given to participants, with the understanding that their submissions would be evaluated for correctness, organization, and content. Punctuation, spelling, and recognizable grammatical errors that might obstruct clear communication were all evaluated. Assignments from the start, middle, and end of the semester were chosen for comparison in order to

assess any changes in the intricacy, accuracy, and fluency of participants' work over the course of the semester. For this study, three assignments were selected from each participant, for a total of 225 assignments analyzed. All groups engaged in their specific interventions over the course of 14 sessions. The same Writing Anxiety Scale and Writing Test were used for the post-assessment in all three intact classes after this intervention. After the data was gathered, it was analyzed to look for variations in the groups' performance and anxiety levels when writing.

Data Analysis

A T-unit is characterized as one main clause along with any subordinate clauses that are attached to or embedded within it. The classification of clauses involves the distinction between dependent and independent clauses, with an independent clause being self-sufficient. In contrast, a dependent clause, which includes adverbial, nominal, and adjectival clauses, comprises a finite verb and a subject (Wolfe-Quintero,1998). Following the frameworks proposed by Storch (2009), the evaluation of complexity was conducted through the examination of the ratios of clauses per T-unit (C/T) and dependent clauses per T-unit (DC/T).

The assessment of accuracy took into account the proportion of error-free T-units (EFT/T), the proportion of error-free clauses (EFC/C), and the total number of errors per total number of words (E/W). Errors were categorized into syntactic errors (e.g., word order, incomplete sentences), morphological errors (e.g., tense, agreement, use of articles), and errors in word choice. Notably, spelling and mechanical errors such as punctuation were excluded from consideration. Fluency metrics included the total number of words (W), the count of T-units, and the length of T-units measured in words per T-unit (W/T).

An additional coder was engaged to ensure coding reliability. The inter-coder reliability achieved high scores of .92 for T-unit identification and .97 for clause identification. Regarding the identification of error-free clauses and T-units, the reliability scores were .91 and .93, respectively. The data analysis involved employing a one-way analysis of variance (ANOVA) to compare the scores of the three groups on writing accuracy, fluency, complexity, and writing anxiety tests. This method was chosen to determine whether there were any statistically significant differences among the means of the three independent groups. Before delving into ANOVA, key assumptions, including normality and homogeneity of variances, were thoroughly examined to ensure the reliability of the subsequent results. The null hypothesis assumed no significant differences existed between the group means, while the alternative hypothesis posited that at least one group mean differed.

Post hoc tests, Bonferroni, were then employed to pinpoint specific group differences if the ANOVA results were significant. This ensured that effect size measures, such as eta-squared or omega-squared, could be computed to offer insights into the practical significance of the observed differences. The same one-way ANOVA procedure was applied to posttest scores, allowing for an examination of changes or improvements within each group over the intervention period.

Results

Research Question 1

The first research question delved into how different corrective written feedbacks—teacher, AI generated, and no correction—affect language learners' anxiety levels in writing. The data analysis yielded noteworthy results, presented in tables 1 and 2.

Table 1

ANOVA for Groups' Scores on Writing Anxiety

Variable	Correction type	N	M	F	p	η^2
Writing anxiety	AI generated	25	1.60	53.54	.001	0.70
	Teacher	25	2.10			
	No correction	25	2.70			

Table 1 shows the mean scores for writing anxiety differed significantly across the three groups. The AI-generated group (Poe group) exhibited the lowest anxiety levels, with a mean score of 1.60, followed by the teacher group at 2.10 and the no-correction group at 2.70. The effect size (η^2) of 0.70 suggests a significant impact. The findings in Table 2 reveal a significant reduction in writing anxiety among learners who received AI-generated CF compared to those receiving teacher-generated feedback or no correction ($p = .001$).

Table 2

Bonferroni for Comparisons Between the Groups' Writing Anxiety

Dependent variable	(I) Correction type	(J) Correction type	Mean difference
Writing anxiety	AI generated	Teacher	-.50
		No correction	-1.6
	Teacher	No correction	-.60

Note. I = x; J = x.
$p = .001$

Research Question 2

The second research question centered on a comparison of the writing fluency of students in the three intact classes. Results are presented in tables 3 and 4.

Table 3

ANOVA for Groups' Scores on Writing Fluency

Variable	Correction type	N	M	F	p	η²
EFC/C	AI generated	25	0.8	3.56	.001	0.75
	Teacher	25	0.6			
	No correction	25	0.4			
EFT/T	AI generated	25	0.84	4.21	.001	0.61
	Teacher	25	0.64			
	No correction	25	0.40			
E/W	AI generated	25	0.85	6.25	.001	0.51
	Teacher	25	0.62			
	No correction	25	0.41			

Note. EFC/C= proportion of error-free clauses, FFT/T= error-free T-units, E/W= the total number of errors per total number of words.

Concerning fluency, a one-way analysis of variance (ANOVA) revealed noteworthy distinctions in the proportion of error-free T-units (F (2, 72) = 3.56, p < .05, η² = 0.71). Upon a more detailed investigation into the proportion of error-free clauses, significant differences emerged among distinct groups (F (2, 72) = 4.12, p < .05, η² = 0.61). The results presented in Table 4 from pairwise comparisons underscored that both AI-generated and teacher-generated CF exhibited significantly higher proportions of error-free clauses compared to the no-correction group. Furthermore, the total number of errors per total number of words exhibited notable differences between groups (F (2, 72) = 6.25, p < .05, η² = .51). Subsequent pairwise comparisons elucidated that the error rate within the AI-generated feedback group was lower than that within the teacher-generated feedback group. Additionally, the error rate in the teacher-generated group was lower than in the no-correction group. As a result, these findings suggest an improvement in syntactic accuracy across the various feedback groups.

Table 4

Bonferroni for Comparisons Between the Groups' Writing Fluency

Dependent variable	(I) Correction type	(J) Correction type	Mean difference (I-J)
EFC/C	AI generated	Teacher	0.20
		No correction	0.40
	Teacher generated	No correction	0.20
EFT/T	AI generated	Teacher	0.24
		No correction	0.44
	Teacher	No correction	0.20
E/W	AI generated	Teacher	0.23
		No correction	0.34
	Teacher	No correction	0.21

Note. EFC/C= proportion of error-free clauses, FFT/T= error-free T-units, E/W= the total number of errors per total number of words

p = .001

Research Question 3

The third research question concerned the writing complexity of students in the three intact classes. Results are presented in tables 5 and 6.

Table 5

ANOVA for Groups' Scores on Writing Complexity

Variable	Correction type	*N*	*M*	*F*	*p*	η²
CT	AI generated	25	2.5	5.62	.001	0.62
	Teacher	25	2.00			
	No correction	25	1.60			
DCT	AI generated	25	1.20	4.95	.001	0.62
	Teacher	25	0.80			
	No correction	25	0.60			

Note. CT = the ratios of clauses per T-unit; DCT = dependent clauses per T-unit

Table 5 presents discernible variations in the ratio of clauses per T-unit (F (2, 75) = 5.62, $p < .05$, η² = 0.62) and the ratio of dependent clauses per T-unit (F (2, 72) = 4.95, $p < .05$, η² = 0.62) among the three cohorts. A more in-depth examination, facilitated by a post hoc analysis (as outlined in Table 6), provides additional evidence affirming the superior performance of AI-generated CF over teacher-generated CF. Furthermore, the teacher-generated feedback, in comparison, demonstrates higher effectiveness when contrasted with the no-correction group. These findings shed light on the nuanced impact of different feedback approaches on the syntactic structure, offering valuable insights into the intricate dynamics of language learning and correction methods.

Table 6

Bonferroni for Comparisons Between the Groups' Writing Complexity

Dependent variable	(I) Correction type	(J) Correction type	Mean difference (I-J)
CT	AI generated	Teacher	0.50
		No correction	0.90
	Teacher generated	No correction	0.40
DCT	AI generated	Teacher	0.60
		No correction	0.40
	Teacher	No correction	0.20

Note. CT = the ratios of clauses per T-unit; DCT = dependent clauses per T-unit

p = .001

As seen in Table 6, the mean differences in scores were computed for each combination of correction methods, revealing significant differences across all comparisons ($p < .001$). In the CT, AI-generated correction yielded the highest mean difference of 0.50 compared to teacher-generated with no correction (0.40). In the DCT, AI-generated correction also resulted in the highest mean difference of 0.60, followed by teacher-generated correction at 0.20 and no correction at 0.40. These findings suggest that AI-generated correction is more effective than teacher-generated or no correction in improving learners' performance in writing complexity.

Research Question 4

The fourth research question concerned comparing the writing fluency of students in the three intact classes. Results are presented in tables 7 and 8.

Table 7

ANOVA for Groups' Scores on Writing Fluency

Variable	Correction type	N	M	F	p	η^2
Total of T-units	AI generated	25	31	80.149	.001	0.56
	Teacher	25	28			
	No correction	25	24			
Total of words	AI generated	27	750	95.38	.000	0.61
	Teacher	25	640			
	No correction	35	590			
W/T	AI generated	27	19.50	51.64	.000	0.51
	Teacher	25	17.30			
	No correction	35	15.10			

Note. W/T= the length of T-units measured in words per T-unit

The findings of the study reveal significant differences across the three types of corrections (AI generated, teacher generated, and no correction) regarding various linguistic variables. The total number of T-units produced by participants under the AI-generated correction condition was significantly higher ($M = 31$) compared to the teacher correction condition ($M = 28$) and the no-correction condition ($M = 24$). This difference was statistically significant ($F = 80.149$, $p < .001$), indicating a substantial impact of the correction method on the overall syntactic structure. Similarly, the total number of words in the AI-generated correction condition ($M = 750$) surpassed those in the teacher correction condition ($M = 640$) and the no-correction condition ($M = 590$), with a significant overall difference ($F = 95.38$, $p < .001$). This suggests that AI-generated corrections influenced the participants to produce more words in their writing.

The words per T-unit (W/T) ratio significantly differed among the three conditions. Participants in the AI-generated correction condition had a higher W/T ratio ($M = 19.50$) compared to the teacher correction condition ($M = 17.30$) and the no-correction condition ($M = 15.10$). This difference was statistically significant ($F = 51.64$, $p < .001$), indicating that AI-generated corrections influenced the number of words and the distribution of words within T-units. Results of the post hoc test (Table 8) also verified that the

differences between the AI-generated and teacher-generated CF on three aspects of writing fluency were statistically significant (p = .001), favoring the AI-generated feedback group. The writing fluency of the no-correction group was significantly lower than the writing fluency of the teacher-generated corrective feedback (p = .001).

Table 8

Bonferroni for Comparisons Between the Groups' Writing Fluency

Dependent variable	(I) Correction type	(J) Correction type	Mean difference (I-J)
Total of T-units	AI generated	Teacher	3.00
		No correction	7.00
	Teacher generated	No correction	4.00
Total of words	AI generated	Teacher	110
		No correction	150
	Teacher	No correction	50
W/T	AI generated	Teacher	2.20
		No correction	4.40
	Teacher	No correction	2.20

Note. p = .001

Discussion

Incorporating AI into educational settings has become a focal point of scholarly inquiry, with researchers delving into its potential to augment learning outcomes. This discourse consolidates insights derived from a quasi-experimental study examining the influence of AI-generated CF on writing anxiety, fluency, accuracy, and complexity among EFL learners. The study systematically compares the efficacy of AI-generated feedback against feedback provided by teachers. The results are contextualized within the broader literature on AI in education, the digital transformation in higher education, and the overarching domain of second language acquisition. The results indicate that both AI-generated and teacher-provided feedback significantly affected the language learners' writing accuracy, fluency, and complexity.

Interestingly, AI-generated feedback proves to be more effective than teacher-generated feedback. This aligns with the broader discourse on the efficacy of AI in education, as discussed by Bozkurt et al. (2021) and Chu et al. (2022). These studies emphasized the transformative potential of AI in enhancing educational practices and suggested that AI could provide personalized and timely feedback, addressing individual learning needs.

A noteworthy outcome of the study is the reduction in learners' writing anxiety facilitated by teacher and AI-generated feedback. This finding resonates with the work of Ellis (2009a), who highlighted the importance of feedback in creating a supportive learning environment and reducing learners' anxiety. The study contributes to the growing body of research that recognizes the emotional aspects of language

learning and emphasizes the role of technology, including AI, in fostering a positive learning experience (Han & Hyland, 2015).

The comparison between AI- and teacher-generated feedback draws attention to the unique advantages of AI, as evidenced by the study's results. The AI system used in the research, the Poe application, outperformed human teachers in enhancing writing skills and reducing anxiety. This aligns with the findings of Bonilla López et al. (2018) who investigated the differential effects of feedback forms in second-language writing. The discussion here underscores the potential of AI to provide consistent and objective feedback, addressing some limitations associated with human feedback, such as variability and subjectivity.

The theoretical underpinnings of the study draw support from skill acquisition theory (DeKeyser, 2007) and the task complexity framework (Robinson, 2011). Skill acquisition theory underscores the importance of practice and feedback in language learning, aligning with the study's focus on corrective feedback's impact on writing skills. Insights from the task complexity framework shed light on how cognitive demands embedded within writing tasks influence learners' language development, providing a valuable perspective for interpreting outcomes related to complexity.

Beyond its theoretical contributions, this study enriches the ongoing discourse on AI in education. In line with trends discussed by Bozkurt et al. (2023) and Chen et al. (2020), it emphasizes the need for a nuanced understanding of AI's role in education, considering both practical applications and theoretical implications. Bozkurt et al.'s (2023) systematic review and exploration of speculative futures for ChatGPT contributed to the broader dialogue on responsibly integrating generative AI into education. Therefore, this study not only provides insights into language learning dynamics but also aligns with and extends the broader conversation on the integration of AI into educational contexts.

Implications and Conclusions

The findings of this study hold practical implications for language educators and policymakers. Integrating AI-generated feedback systems, such as the Poe application, into language classrooms could enhance the quality and efficiency of feedback provision. However, as discussed by Kurzweil (2014) and Tlili et al. (2023), ethical considerations should guide the responsible implementation of AI in education. Teachers may need to adapt their roles to incorporate AI as a supportive tool rather than a replacement. In conclusion, the quasi-experimental study on the impact of AI-generated corrective feedback on EFL learners' writing skills contributes valuable insights to the evolving landscape of AI in education. The effectiveness of AI in fostering writing accuracy, fluency, and complexity, and reducing anxiety positions it as a promising tool for language learning. However, the responsible integration of AI into educational practices requires a thoughtful and ethical approach. This discussion bridges the study's findings with existing literature, providing a comprehensive understanding of the implications for language education in the digital transformation era.

Limitations and Suggestions

While the study presents compelling insights into the efficacy of AI-generated corrective feedback, certain limitations warrant consideration. First, while valuable for initial exploration, the quasi-experimental design may lack the robustness of a randomized controlled trial. The study's focus on one specific AI application, Poe, raises questions about the generalizability of findings to other AI platforms. Moreover, the study primarily gauges short-term impacts, leaving the long-term effects of AI-generated feedback on language acquisition unexplored. Additionally, the absence of qualitative data may limit a nuanced understanding of learners' perceptions and experiences with AI feedback. Future research could employ mixed-methods approaches, incorporating qualitative insights to complement quantitative findings. Furthermore, the study needs to delve deeper into the sociocultural aspects influencing the reception of AI in diverse educational contexts, an avenue ripe for exploration. Addressing these limitations will enhance the robustness and applicability of research on AI in language education.

Funding

References

Benavides, L. M. C., Tamayo Arias, J. A., Arango Serna, M. D., Branch Bedoya, J. W., & Burgos, D. (2020). Digital transformation in higher education institutions: A systematic literature review. *Sensors, 20*(11), Article 3291. https://doi.org/10.3390/s20113291

Benson, S., & DeKeyser, R. (2018). Effects of written corrective feedback and language aptitude on verb tense accuracy. *Language Teaching Research, 23*(6), 702–726. https://doi.org/10.1177/1362168818770921

Biber, D., Nekrasova, T., & Horn, B. (2011). The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis. *ETS Research Report Series, 2011*(1), i–99.https://doi.org/10.1002/j.2333-8504.2011.tb02241.x

Bonilla López, M., Van Steendam, E., Speelman, D., & Buyse, K. (2018). The differential effects of comprehensive feedback forms in the second language writing class. *Language Learning, 68*(3), 813–850. https://doi.org/10.1111/lang.12295

Bozkurt, A., Karadeniz, K., Baneres, D., Guerrero-Roldán, A. E., & Rodríguez, M. E. (2021). Artificial intelligence and reflections from educational landscape: A review of AI studies in half a century. *Sustainability, 13*(2), Article 800. https://doi.org/10.3390/su13020800

Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., Farrow, R., Bond, M., Nerantzi, C., Honeychurch, S., Bali, M., Dron, J., Mir, K., Stewart, B., Costello, E., Mason, J., Stracke, C. M., Romero-Hall, E., Koutropoulos, A., … Jandrić, P. (2023). Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. *Asian Journal of Distance Education, 18*(1), 53–130. https://www.asianjde.com/ojs/index.php/AsianJDE/article/view/709

Cao, L. (2023). Trans-AI/DS: Transformative, transdisciplinary and translational artificial intelligence and data science. *International Journal of Data Science and Analytics,15*(1), 119–132. https://doi.org/10.1007/s41060-023-00383-y

Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*(2020), Article 100002. https://doi.org/10.1016/j.caeai.2020.100002

Chu, H.-C., Hwang, G.-H., Tu, Y.-F., & Yang, K.-H. (2022). Roles and research trends of artificial intelligence in higher education: A systematic review of the top 50 most-cited articles. *Australasian Journal of Educational Technology, 38*(3), 22–42. https://ajet.org.au/index.php/AJET/article/view/7526

Cox, A. M. (2021). Exploring the impact of artificial intelligence and robots on higher education through literature-based design fictions. *International Journal of Educational Technology in Higher Education, 18*(2021), Article 3. https://doi.org/10.1186/s41239-020-00237-8

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 94–112). Routledge.

Diebold, G. (2023, January 17). *Higher education will have to adapt to generative AI—And that's a good thing*. Center for Data Innovation. https://datainnovation.org/2023/01/higher-education-will-have-to-adapt-to-generative-ai-and-thats-a-good-thing/

Dogan, M. E., Goru Dogan, T., & Bozkurt, A. (2023). The use of artificial intelligence (AI) in online learning and distance education processes: A systematic review of empirical studies. *Applied Sciences, 13*(5), Article 3056. https://doi.org/10.3390/app13053056

Ellis, R. (2009a). A typology of written corrective feedback types. *ELT Journal, 63*(2), 97–107. https://doi.org/10.1093/elt/ccn023

Ellis, R. (2009b). Corrective feedback and teacher development. *L2 Journal: An electronic refereed journal for foreign and second language educators, 1*(1), 2–18. https://doi.org/10.5070/l2.v1i1.9054

Ellis, R. (2009c). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics, 19*(3), 221–246. https://doi.org/10.1111/j.1473-4192.2009.00231.x

Evans, N. W., Hartshorn, K. J., Cox, T. L., & Martin de Jel, T. (2014). Measuring written linguistic accuracy with weighted clause ratios: A question of validity. *Journal of Second Language Writing, 24*(1), 33–50. https://doi.org/10.1016/j.jslw.2014.02.005

Goksel, N., & Bozkurt, A. (2019). Artificial intelligence in education: Current insights and future perspectives. In S. Sisman-Ugur & G. Kurubacak (Eds.), *Handbook of research on learning in the age of transhumanism* (pp. 224–236). IGI Global. https://doi.org/10.4018/978-1-5225-8431-5.ch014

Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing, 30*(1), 31–44. https://www.doi.org/10.1016/j.jslw.2015.08.002

Hartshorn, K. J., & Evans, N. W. (2015). The effects of dynamic written corrective feedback: A 30-week study. *Journal of Response to Writing, 1*(2), 6–34.

Humble, N., & Mozelius, P. (2022). The threat, hype, and promise of artificial intelligence in education. *Discover Artificial Intelligence, 2*(2022), Article 22. https://doi.org/10.1007/s44163-022-00039-z

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching, 39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

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

Karim, K., & Nassaji, H. (2018). The revision and transfer effects of direct and indirect comprehensive corrective feedback on ESL students' writing. *Language Teaching Research, 24*(4), 519–539. https://doi.org/10.1177/1362168818802469

Karim, K., & Nassaji, H. (2019). The effects of written corrective feedback: A critical synthesis of past and present research. *Instructed Second Language Acquisition, 3*(1), 28–52. https://doi.org/10.1558/isla.37949

Kurzer, K. (2018). Dynamic written corrective feedback in developmental multilingual writing classes. *TESOL Quarterly, 52*(1), 5–33. https://doi.org/10.1002/tesq.366

Kurzweil, R. (2014). The singularity is near. In R. L. Sandler (Ed.), *Ethics and emerging technologies* (pp. 393–406). Palgrave Macmillan UK. https://doi.org/10.1057/9781137349088_26

Leeman, J. (2010). Feedback in L2 learning: Responding to errors during practice. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 111–138). Cambridge University Press. https://doi.org/10.1017/cbo9780511667275.007

Li, S., & Vuono, A. (2019). Twenty-five years of research on oral and written corrective feedback in *System*. *System, 84*(1), 93–109. https://doi.org/10.1016/j.system.2019.05.006

Lim, S. C., & Renandya, W. A. (2020). Efficacy of written corrective feedback in writing instruction: A meta-analysis. *TESL-EJ, 24*(3), 1–26. https://tesl-ej.org/wordpress/issues/volume24/ej95/ej95a3/

Liu, Q., & Brown, D. (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing, 30*(1), 66–81. https://doi.org/10.1016/j.jslw.2015.08.011

Liu, Y., & Huang, J. (2020). The quality assurance of a national English writing assessment: Policy implications for quality improvement. *Studies in Educational Evaluation, 67*(2), Article 100941. https://doi.org/10.1016/j.stueduc.2020.100941

Long, M. H. (1980). *Input, interaction, and second language acquisition*. University of California.

Luo, Y., & Liu, Y. (2017). Comparison between peer feedback and automated feedback in college English writing: A case study. *Open Journal of Modern Linguistics, 7*(04), 197–215. https://doi.org/10.4236/ojml.2017.74015

Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J. H., Ogata, H., Baltes, J., Guerra, R., Li, P., & Tsai, C.-C. (2020). Challenges and future directions of big data and artificial intelligence in education. *Frontiers in Psychology*, *11*, Article 580820. https://doi.org/10.3389/fpsyg.2020.580820

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*(2021), Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies, 27*(6), 7893–7925. https://doi.org/10.1007/s10639-022-10925-9

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing, 21*(4), 375–389. https://doi.org/10.1016/j.jslw.2012.09.004

Polio, C., & Yoon, H. J. (2018). The reliability and validity of automated tools for examining variation in syntactic complexity across genres. *International Journal of Applied Linguistics (United Kingdom), 28*(1), 165–188. https://doi.org/10.1111/ijal.12200

Riazantseva, A. (2012). Outcome measure of L2 writing as a mediator of the effects of corrective feedback on students' ability to write accurately. *System, 40*(3), 421–430. https://doi.org/10.1016/j.system.2012.07.005

Robinson, P. (2011). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3–38). John Benjamins.

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133–164). John Benjamins.

Schmidt, R. (2012). Attention, awareness, and individual differences in language learning. *Perspectives on Individual Characteristics and Foreign Language Education*, *6*(27), 27-49.

Selwyn, N., Hillman, T., Bergviken-Rensfeldt, A., & Perrotta, C. (2023). Making sense of the digital automation of education. *Postdigital Science and Education, 5*(1), 1–14. https://doi.org/10.1007/s42438-022-00362-9

Sharma, R. C., Kawachi, P., & Bozkurt, A. (2019). The landscape of artificial intelligence in open, online and distance education: Promises and concerns. *Asian Journal of Distance Education, 14*(2), 1–2. https://doi.org/10.5281/zenodo.3730631

Sia, P. F. D., & Cheung, Y. L. (2017). Written corrective feedback in writing instruction: A qualitative synthesis of recent research. *Issues in Language Studies, 6*(1), 61–80. https://www.ils.unimas.my/images/pdf/v6n1/ILS_Vol6No1_Sia.pdf

Stefanou, C., & Révész, A. (2015). Direct written corrective feedback, learner differences, and the acquisition of second language article use for generic and specific plural reference. *The Modern Language Journal, 99*(2), 263–282. https://doi.org/10.1111/modl.12212

Storch, N. (2009). The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing*, *18*(2), 103–118. https://api.semanticscholar.org/CorpusID:62520708

Tang, K.-Y., Chang, C.-Y., & Hwang, G.-J. (2021). Trends in artificial intelligence-supported e-learning: A systematic review and co-citation network analysis (1998–2019). *Interactive Learning Environments*, *31*(4), 2134–2152. https://doi.org/10.1080/10494820.2021.1875001 Thi, K, N., & Nikolov, M. (2021). Feedback treatments, writing tasks, and accuracy measures: A critical review of research on written corrective feedback. TESL-E, 25(3), 1-25. https://www.tesl-ej.org/pdf/ej99/a16.pdf

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(1), Article 15. https://doi.org/10.1186/s40561-023-00237-x

Truscott, J. (2010). Further thoughts on Anthony Bruton's critique of the correction debate. *System, 38*(4), 626–633. https://doi.org/10.1016/j.system.2010.10.003

Truscott, J., & Hsu, A. Y.-P. (2008). Error correction, revision, and learning. *Journal of Second Language Writing, 17*(4), 292–305. https://doi.org/10.1016/j.jslw.2008.05.003

Winterson, J. (2022). *12 Bytes: How artificial intelligence will change the way we live and love*. Penguin Random House.

Wolfe-Quintero, K. (1998). The connection between verbs and argument structures: Native speaker production of the double object dative. *Applied Psycholinguistics*, *19*(2), 225–257. https://doi.org/10.1017/S0142716400010055

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zhang, T. (2021). The effect of highly focused versus mid-focused written corrective feedback on EFL learners' explicit and implicit knowledge development. *System, 99*(2), Article 102493. https://doi.org/10.1016/j.system.2021.102493