Education Reform and Development, 2026, Volume 8, Issue 7
https://ojs.bbwpublisher.com/index.php/ERD
Online ISSN: 2652-5372
Print ISSN: 2652-5364

Comparing Teacher-Led and AI-Assisted Dynamic Written Corrective Feedback: Effects on Self-Correction Awareness and L2 Writing Proficiency

Liuyan Xian*, **Linping Huang**

School of Foreign Studies, Lingnan Normal University, Zhanjiang 524048, Guangdong, China

*Corresponding author: Liuyan Xian, xianliuyan56913@126.com

Abstract: The present study aims to compare teacher-led and AI-assisted dynamic written corrective feedback (DWCF) in relation to L2 learners' self-correction awareness and writing performance. Participants were randomly assigned to the teacher-led DWCF group (N = 51) or the AI-assisted DWCF group (N = 50). A writing pretest was administered before the intervention, and a writing posttest and a questionnaire measuring learners' self-correction awareness were administered after the intervention. No significant difference was found in self-correction awareness between the teacher-led and AI-assisted DWCF groups, suggesting that both modes can raise learners' attention to language precision. Moreover, the teacher-led DWCF was more successful than the AI-assisted DWCF in enhancing writing performance. The study argues that AI feedback is more mechanical and superficial than teacher-led corrections. The social interaction in teacher-led DWCF is more likely to ensure comprehensible input, promote deeper thinking, and foster proficiency development. Hence, a combination of teacher-led and AI-assisted DWCF can be the optimal mode in an authentic teaching environment.

Keywords: Teacher-led DWCF; AI-assisted DWCF; Self-correction awareness; Language proficiency

Online publication: August 13, 2026

1. Introduction

L2 English writing learners commonly encounter problems with linguistic accuracy and persistent errors [1]. Teachers' error corrections usually exclude the necessary interaction between feedback givers and recipients. This may lead to low proficiency in its uptake. Hence, optimization of corrective feedback has been a critical pedagogical issue. under this context, dynamic written corrective feedback (DWCF) was proposed to cater to individual learner differences [2]. Characterized by meaningfulness, timeliness, constancy, and manageability, DWCF highlights the intensive practice and meaning negotiation between the language instructor and the learners. The interaction in DWCF is believed to have significantly contributed to raising learners' self-correction awareness and avoiding error recurrence, leading to learning autonomy in the end [1–3].

However, manual error correction is not free from limitations. In China, most regions adopt large-class teaching so that error corrections become a heavy workload for most teachers. Thus, the consistency of feedback quality can hardly be ensured [4]. The emergence of Generative artificial intelligence (AI) could be a resort to reduce teachers' workload in the practical teaching environment, because it is advantageous in providing immediate and detailed metalinguistic explanations and tailored feedback aligned with user needs [5]. This allows educators to prioritize organization and content development in writing correction.

AI has shown great potential in addressing practical constraints. However, studies integrating AI with the framework of DWCF remain limited. Hence, this study tries to use AI to improve teachers' working efficacy, and compare teacher-led with AI-assisted DWCF on self-correction awareness and writing outcomes.

2. Literature review

2.1. Dynamic written corrective feedback

DWCF focuses on how to enhance L2 writing accuracy. Rather than a one-size-fits-all feature, DWCF caters to individual learner differences [2][6]. The merit of DWCF lies in its dynamic adaptability to adjust the content and format of feedback in real time. Meaningfulness, timeliness, constancy, and manageability are four elements to ensure its effectiveness [7]. Specifically, meaningfulness dictates the use of error location, or coded feedback, to prompt cognitive engagement from language learners. Timeliness and constancy govern the immediacy and frequency of feedback delivery to invite learners' lasting attention. Manageability addresses the practical needs of both instructors and learners [2]. DWCF usually limits writing tasks to brief ten-minute paragraphs to reduce teachers' workload and learners' cognitive load.

Six iterative steps are adopted to perform DWCF in real teaching settings [2]. Learners, firstly, compose a brief paragraph within 10 minutes, and instructors collect the manuscripts and locate the errors with codes. In the third step, the manuscripts are returned to the language learners for further revision and error logging. Again, instructors then provide feedback on these revised drafts. Different from the previous round, the feedback is more explicit with metalinguistic explanations. This feedback-revision process continues until the essay is grammatically accurate. During this dynamic learning process, learners internalize the rules governing linguistic features to improve their capacity for error identification and correction.

2.2. Input Hypothesis and dynamic written corrective feedback

The Input Hypothesis (IH) claims that the key factor in language acquisition is understandable input "i + 1" [8]. "i" refers to the learners' present level of ability, and "1" to linguistic structures just above this level. Input over the barrier of "i+n" is effortless in language acquisition. Also, being meaningful and contextualized is another factor leading to acquisition. For instance, modified interactions, like caretaker speech or teacher speak in L2 classrooms, naturally modify grammatical complexity and utilize paralinguistic cues to ensure comprehensibility. These interactions help cognitive processing and linguistic intake by matching input to the learners' "i + 1" level [8].

IH gives good theoretical insight into the implementation of DWCF. The metalinguistic explanations given by DWCF are explicit scaffolds to supply "i + 1" input for the unique learner needs. Teacher-led feedback delivers precise coding and contextual explanations [3], whereas AI-assisted DWCF is a combination of real-time processing and regular and consistent exposure to the “i + 1”.

2.3. Noticing Hypothesis and dynamic written corrective feedback

The Noticing Hypothesis (NH) illustrates how input is transformed into intake. Conscious attention is believed to be a precondition for acquisition [9]. It is an active reception rather than a passive one, as learners' response to linguistic stimuli mediates the transformation from input to intake [10]. Conscious attention is depicted in a tripartite hierarchy: perception, noticing, and comprehension [9]. To make it more specific, perception is a foundational unconscious awareness of the input, but not a key factor in driving language acquisition. Noticing, as the core of the hypothesis, indicates the learners' conscious learning. For example, learners can notice a specific grammatical pattern when they find that some words do not match in a context. This observation is essential for L2 learners moving from specific uses of the linguistic pattern to the rule governing this linguistic feature. The last term, comprehension, involves analytical thought and the inductive generalization of linguistic rules to generate systematic cognitive schemas [9]. Learners at this level are able to use the knowledge in new situations.

NH implies that DWCF can bridge the gap between inaccurate output and target-like forms. Learners can move from simply observing mistakes to actively noticing them through the metalinguistic explanations in DWCF. For AI-assisted DWCF, it is advantageous in sustaining learners' attention because of its real-time processing and precise illustration of the errors in the essay. Therefore, both manual and AI-assisted feedback create a cognitive foundation for the development of self-correction awareness [9].

2.4. Self-correction awareness

Self-correction awareness is a fundamental feature of human learning. It refers to learners' ability to identify and revise their individual language mistakes while writing [11]. It involves four steps: identifying the error, understanding the reasons behind the error, performing a correction, and reflecting on the result. In language classes, to guide learners through this process, teachers give them a checklist to proofread their vocabulary, grammar, text structure, and formatting in their essays. In the meantime, asking learners to record their common errors and revisions is also a vital practice, helping them move from passively receiving feedback to actively avoiding mistakes.

2.5. Previous studies

DWCF is a core instructional model in L2 writing. Extensive studies have laid a strong foundation for its practical application. Evans et al. (2010), as the first study introducing this method, put forward four key standards (meaningfulness, timeliness, constancy, and manageability) to define its dynamic nature and ensure its authentic operation in a real teaching environment [2]. A closed-loop framework involving feedback, revision, and resubmission is also demonstrated. However, this study is preliminary research that excludes a control group in the experiment. With a similar experimental design, Evans et al. (2011) reported significant gains in accuracy for ESL students using DWCF compared to traditional instruction [12]. Kurzer (2018) expanded the sample of the study from mere advanced learners to students with different proficiency levels and found that DWCF can effectively improve learners' self-editing ability [3].

Though various empirical studies have demonstrated the effectiveness of DWCF, the labor-intensive nature when teacher-led has prompted scholars to seek technological solutions, as Artificial Intelligence (AI) has emerged as a new hotspot; it offers significant technical advantages in error correction in L2 writing. Su

(2024) pointed out that AI can provide precise corrections, customize suggestions to cater to learners' needs, and reduce workload for practitioners [13]. Yang and Chen (2026) reported that ChatGPT can generate more accurate and detailed feedback when evaluating AI-assisted error correction [14]. Crosthwaite and Sun (2026) supported their point in their scoping review. Generative AI, with proper prompts, can improve language accuracy more than manual feedback, but the best model is to combine AI with human force to create a collaborative model. However, there is also another voice [15]. Lin and Crosthwaite (2024) proposed that ChatGPT feedback is redundant and sometimes low in relevance, and that students frequently receive only surface corrections, posing a challenge for deep content [16].

In summary, a multitude of studies have proved the effectiveness of DWCF in language accuracy and its promising prospects in L2 writing. Yet, in reality, teachers still suffer from an overwhelming workload. Integrating AI with the DWCF framework can effectively solve the practical problem because it can bring efficacy to mechanical error correction. Therefore, this study attempts to compare teacher-led and AI-assisted DWCF to answer the following two questions:

(1) Which group exhibits higher self-correction awareness in English writing: participants receiving teacher-led DWCF or those receiving AI-assisted DWCF?

(2) Is AI-assisted DWCF more effective in improving L2 learners' writing proficiency than teacher-led DWCF?

3. Methodology

3.1. Research design and procedure

The research questions listed in the previous section were answered quantitatively by a pretest-posttest-designed quasi-experiment. The participants were English Education majors, a group trained in combined expertise in language proficiency and English-teaching skills. Students were introduced to such courses as Comprehensive English, English Writing, Second Language Acquisition, Linguistics, and Educational Psychology. They were more likely to be engaged in the language learning process and revise their essay according to the feedback they received. The participants were allocated to either a teacher-led DWCF group or an AI-assisted DWCF group. Prior to the experiment, all participants knew of their right to withdraw at any time, and that withdrawal would not affect their academic grading. Besides, they were required to finish a writing pretest. The six-week intervention was integrated into an English writing course, which focused on syntactic variety, grammatical accuracy, self-correction strategies, paragraph construction, etc. Right after the intervention, the participants finished a writing posttest. Participants receiving AI-assisted feedback would be endowed with a chance to be exposed to the teacher-led DWCF after the intervention as compensation.

3.2. Subjects

The participants of this study were 101 first-year English majors (aged 18–21) from eight intact classes. According to the description of China's Standards of English Language Ability (2024 Edition), participants landed at level 6, corresponding to an IELTS band score of 5.5-6.0. They were randomly assigned to either a teacher-led DWCF group (N = 51) or an AI-assisted DWCF group (N = 50) using Microsoft Excel to remove influence from different teachers. Pretest results confirmed no significant proficiency differences between groups.

3.3. Instruments and measures

A self-correction awareness questionnaire was designed based on the concept of self-correction awareness. More specifically, the questionnaire covered five aspects: error recognition, error correction behavior, error reflection and attribution, feedback acceptance and learning willingness, and self-efficacy. A 5-point Likert scale was utilized to measure participants' perceptions.

Writing Tasks were another tool used in this study to assess L2 writing performance at the pretest and posttest stages. To limit possible practice effects, the writing topics were selected from the official Test of English as a Foreign Language (TOEFL) iBT writing pool, because the target participants normally attend the International English Language Testing System (IELTS) for postgraduate education. The topics were quite comparable to real university experiences in order to produce realistic writing samples. All participants were asked to write academic argumentative essays to ensure consistency in measurement. That is, learners first discuss a specific issue, express their opinions, and then support them with evidence. Writing performance was measured by an objective multi-dimensional grading rubric of IELTS [6]. Two raters with over seven years of experience in language education were selected to grade the tasks (ICC =.89).

3.4. Intervention

This intervention lasted for six weeks. Students wrote on one topic each week and received three rounds of error correction by either the teacher or AI.

In the teacher-led DWCF group, the instructor offered three rounds of unfocused feedback [6]. The first round located errors without supplying corrections. Students were expected to diagnose and re-edit the errors independently within a ten-minute revision. In the meantime, they established a personal profile to record error types and frequencies of each writing task. This was to enhance their attention to linguistic inaccuracies. The second round of feedback remained implicit, but metalinguistic explanations were provided for unresolved errors. Students were expected to process those errors independently in a ten-minute revision period. The last round of feedback was the most explicit. The instructor corrected all residual errors, and students then completed a final ten-minute revision to consolidate the correct forms. The dynamic model provides a complete learning cycle to engage the students during the entire process. To balance the learning time at the revision phase, each student was allowed 30 minutes regardless of the feedback rounds they had received in the process.

The implementation of AI-assisted DWCF is similar to that of teacher-led DWCF, but it diverges significantly from teacher-led DWCF in execution. AI-assisted DWCF was also administered in three rounds of stepwise error corrections. However, the teacher was a supervisor instead of a feedback provider, because the process was completed with the assistance of AI and the Automatic Writing Evaluation system. In the meantime, the error types and frequency were reported by the automatic feedback system.

4. Data analysis and discussions

The descriptive statistics are shown in **Table 1**.

Table 1. Descriptive statistics of self-correction awareness and writing proficiency

Group	Questionnaire		Pretest		Posttest	
	M	***SD***	***M***	***SD***	***M***	***SD***
Teacher-led DWCF (*N* = 51)	32.39	4.21	17.22	3.45	23.67	2.92
AI-assisted DWCF (*N* = 50)	31.58	3.98	18.28	2.76	21.54	4.18

4.1. Dynamic written corrective feedback and self-correction awareness

The Levene's test indicated the homogeneity of variances, F(1, 98) = .111, p = .740. To address the first research question, an independent samples t-test was performed to compare students' self-correction awareness across the two groups. The analysis revealed no statistically significant difference between the teacher-led DWCF group and the AI-assisted DWCF group (t(98) = .996, p = .321, d = .20).

The comparability of self-correction awareness in the two groups is probably due to the common structural logic of the interventions. Both the teacher-led and AI-assisted models adopted the iterative correction protocol of DWCF. Participants were repeatedly exposed to the linguistic errors across rounds. The meaning negotiation provided in the framework of DWCF raised students' attention and led them to a pattern of self-examination after feedback [9]. This can facilitate learners to establish a habit of checking language accuracy. This may be a possible explanation that mitigated any potential differences between the two groups.

4.2. Dynamic written corrective feedback and writing proficiency

To answer the second research question, an independent samples *t*-test was executed to analyze the pretest and posttest of the writing tasks between groups, and paired samples *t*-tests to examine changes in writing scores before and after intervention within each group. Levene's test indicated unequal variances, $F(1, 97) = 4.12$, $p = .045$, so the Welch-corrected *t* statistic is reported. When comparing the teacher-led DWCF group with the AI-assisted DWCF group ($t = -1.713$, $df = 95.093$, $p > .05$), no significant difference was found in the pretest. The paired-sample *t* tests were performed respectively in the teacher-led and AI-assisted DWCF groups. A significant improvement was reported in the teacher-led DWCF group ($t = -12.752$, $df = 50$, $p < .001$, $d = 1.79$) and in the AI-assisted DWCF group ($t = -5.435$, $df = 49$, $p < .001$, $d = -.769$). To control the influence of the pretest, the gain between pretest and posttest across groups was used as a dependent variable in the independent samples *t*-test. The results showed a significant difference between the two groups ($t = 2.972$, $df = 99$, $p < .05$, $d = .59$).

This disparity may stem primarily from the structural limitations of AI to reproduce the cognitive depth built into teacher-led instruction. Interpersonal interaction provided in teacher-led DWCF brings students' awareness of language errors in writing [9]. At the same time, when students made their first revision, they were required to record their error types and frequencies. The practice also helps students focus on their own language mistakes and prompts the development of students' meta-language cognition. Most importantly, teacher-led instruction can ensure the comprehensibility of the input through meaning negotiation in teacher-student interaction. However, the AI-assisted model is still limited to surface-level modifications [16]. In other words, automatic feedback provides standardized sentence-level corrections, but it does not explain the underlying logic of errors. Additionally, the unified output of AI can hardly satisfy the principle of IH to ensure its comprehensibility [8]. Therefore, although AI is effective in identifying mechanical errors, it cannot replace teacher-led DWCF to cultivate continuous and in-depth writing skills.

4.3. Pedagogical implications

The empirical data showed that although awareness of self-correction was similar in both groups, this change in awareness was not directly related to improvements in writing ability. The problem lies in the quality of the feedback model, but not the existence of consciousness. AI should be placed as an assisting tool to quickly and comprehensively screen out superficial language errors, such as spelling, grammar, and collocations. The practice can release teachers from the mechanical burden of grading and bring them back to their central role in fostering deep learning. Teachers can solve the problems that artificial intelligence cannot solve. Besides, teachers could lead students to figure out the roots of errors and provide personalized heuristic guidance through classroom discussions and individual communication.

As AI-assisted feedback continues to advance, it will need to bridge the gap between simply spotting errors and grasping the underlying logic. Today, more often than not, AI leaves language learners without a clear path to cognitive transformation. But it can be utilized as a tool not only for proofreading but also as a scaffolding for positive reflection.

5. Conclusion

This study examined the differences in self-correction awareness between AI-assisted DWCF and teacher-led DWCF and compared their impact on L2 learners' writing proficiency. Regarding the first research question, no significant difference was found in self-correction awareness between teacher-led and AI-assisted error correction. The data suggested that both modes can draw learners' attention to language accuracy. In response to the second research question, teacher-led DWCF is more beneficial than AI-assisted DWCF for enhancing L2 learners' writing proficiency because teacher-student interaction can ensure that participants understand the meaning of the input. Besides, AI-assisted feedback is more mechanical and superficial, while teacher-led error corrections are more likely to lead to deeper thinking and promote proficiency development.

This paper's contribution to the field is the integration of artificial intelligence into the DWCF framework, which is a major gap in the existing literature. This study rejects the direct relation between self-correction awareness and language proficiency and verifies the mediating effect of the DWCF framework. Theoretically, it exposes the metacognitive nature of DWCF in that the efficiency of DWCF relies on both error correction and the guidance of reflection and attribution.

Limitations of this study include its single-institution sample, limiting generalizability. Future research should recruit students with different language proficiency across institutions. Additionally, the brief six-week intervention may overlook the delayed effects of the approach, and future research should extend to a longer time period to assess sustained writing growth and literacy transfer.

Funding

Lingnan Normal University 2025 Special Project for High-level Talents (Project No.: ZW2513); Lingnan Normal University 2026 Reform Project of Teacher Education Innovation Mechanisms (Project No.: 2026JSJY61)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Hartshorn KJ, Evans NW, Merrill PF, et al., 2010, Effects of Dynamic Corrective Feedback on ESL Writing Accuracy. TESOL Quarterly, 44(1): 84–109.

[2] Evans NW, Hartshorn KJ, McCollum RM, et al., 2010, Contextualizing Corrective Feedback in Second Language Writing Pedagogy. Language Teaching Research, 14(4): 445–463.

[3] Kurzer K, 2018, Dynamic Written Corrective Feedback in Developmental Multilingual Writing Classes. TESOL Quarterly, 52(1): 5–33.

[4] Lee I, 2019, Teacher Written Corrective Feedback: Less Is More. Language Teaching, 52(4): 524–536.

[5] Barrot JS, 2023, Using Automated Written Corrective Feedback in the Writing Classrooms: Effects on L2 Writing Accuracy. Computer Assisted Language Learning, 36(4): 584–607.

[6] Xian L, 2024, Written Corrective Feedback and Individual Differences on Second Language Writing Development. Doctoral dissertation, The Hong Kong Polytechnic University.

[7] Hartshorn KJ, Evans NW, 2012, The Differential Effects of Comprehensive Corrective Feedback on L2 Writing Accuracy. Journal of Linguistics and Language Teaching, 3: 16–46.

[8] Krashen SD, 1985, The Input Hypothesis: Issues and Implications. Longman, New York.

[9] Schmidt RW, 1990, The Role of Consciousness in Second Language Learning. Applied Linguistics, 11(2): 129–158.

[10] Li ZQ, Li YZ, 2019, Research on Second Language Acquisition Input from the Perspective of the Noticing Hypothesis. Journal of Xi'an International Studies University, 27(1): 63–67.

[11] Makino TY, 1993, Learner Self-Correction in EFL Written Compositions. ELT Journal, 47(4): 337–341.

[12] Evans NW, Hartshorn KJ, Strong-Krause D, 2011, The Efficacy of Dynamic Written Corrective Feedback for University-Matriculated ESL Learners. System, 39(2): 229–239.

[13] Su CC, 2024, Role of Artificial Intelligence in Enhancing University English Writing Instruction. Educational Innovations and Emerging Technologies, 4(4): 12–17.

[14] Yang CTY, Chen HHJ, 2026, ChatGPT and L2 Chinese Writing: Evaluating the Impact of Model Version and Prompt Language on Automated Corrective Feedback. Computer Assisted Language Learning, 39(4): 1119–1147.

[15] Crosthwaite P, Sun S, 2026, Generative AI and L2 Written Feedback Studies: A Scoping Review. RELC Journal, 57(1): 207–219.

[16] Lin S, Crosthwaite P, 2024, The Grass Is Not Always Greener: Teacher vs. GPT-Assisted Written Corrective Feedback. System, 127: 103529.