



THE INFLUENCE OF AI-POWERED FEEDBACK ON ENGLISH LANGUAGE LEARNERS' WRITING DEVELOPMENT

Ibodat Obidkhujayeva

Graduate student of Master's degree of Webster University

E-mail: imirxojiyeva@gmail.com

Abstract. *Feedback is one of the most influential factors in second language (L2) writing development, yet providing timely and individualised feedback remains a major challenge for teachers of large classes. The emergence of automated writing evaluation (AWE) systems and, more recently, generative artificial intelligence (AI) tools such as ChatGPT has transformed the ways in which English language learners receive feedback on their texts. This article reviews theoretical foundations and empirical research on AI-powered feedback in order to examine its influence on learners' writing development. The analysis shows that AI feedback has a moderate positive effect on writing performance, especially on lexical and surface-level accuracy, and can increase revision activity and learner engagement. At the same time, its effectiveness depends on the accuracy of the feedback, learners' feedback literacy and the mediating role of the teacher. The article proposes an integrated human-AI feedback model for English writing instruction.*

Keywords: *AI-powered feedback; automated writing evaluation; ChatGPT; L2 writing; written corrective feedback; learner engagement; feedback literacy.*

INTRODUCTION

Writing is widely regarded as one of the most demanding skills for learners of English as a foreign or second language. It requires the simultaneous control of grammar, vocabulary, text organisation, genre conventions and audience awareness. Feedback plays a central role in the development of these abilities. In their influential synthesis, Hattie and Timperley identified feedback as one of the most powerful influences on learning and reported an average effect size of 0.79, while also emphasising that its impact varies considerably depending on the type of feedback and the way it is received (Hattie & Timperley, 2007:83–84). In L2 writing research, feedback has been treated as a key element of process-oriented instruction, and teachers, peers and computers have all been examined as possible sources of feedback (Hyland & Hyland, 2006:83).

Despite this consensus on its importance, the effectiveness of written corrective feedback has long been debated. Truscott argued that grammar correction in L2 writing classes is ineffective and may even be harmful because it diverts time and attention from more productive activities (Truscott, 1996:328–329). Ferris responded that this conclusion was premature and that well-designed, selective and clear correction can help learners improve accuracy (Ferris, 1999:2–4). Subsequent research has largely supported a more nuanced view: feedback is beneficial when it is accurate, focused and actively processed by the learner (Bitchener & Ferris, 2012:4).

A persistent practical problem is that high-quality individual feedback is time-consuming. In university classes with many students, teachers are often unable to provide detailed comments on multiple drafts. Automated writing evaluation systems were developed partly to address this problem, and generative AI tools have now expanded the possibilities even further. Unlike earlier systems, large language models can comment not only on grammar and spelling but also on argumentation, coherence and style, and they can respond to learners' follow-up questions (Barrot, 2023:2–3). These developments raise an important question: to what extent, and under what conditions, does AI-powered feedback actually contribute to English language learners' writing development?

The aim of this article is to answer this question through an integrative review of theoretical and empirical literature.

MAIN PART

Several theoretical perspectives explain why feedback can promote L2 writing development. From a cognitive perspective, Schmidt's noticing hypothesis holds that learners must consciously notice features of the language for input to become intake (Schmidt, 1990:129). Feedback on writing draws learners' attention to gaps between their own production and the target language. Swain's output hypothesis complements this view by proposing that producing language serves a noticing function, a hypothesis-testing function and a metalinguistic, reflective function (Swain, 1995:128–132). When a learner writes a sentence, receives a comment and revises the sentence, all three functions are activated.

Cognitive models of writing further explain where feedback can intervene. Flower and Hayes described composing as a set of recursive processes, namely planning, translating ideas into text and reviewing, all coordinated by a monitor (Flower &

Hayes, 1981:369–370). Feedback can support each stage: it can help learners generate and organise ideas, choose appropriate linguistic forms and evaluate their drafts. From a sociocultural perspective, feedback is effective when it operates within the learner's zone of proximal development, defined by Vygotsky as the distance between what a learner can do independently and what he or she can achieve with guidance (Vygotsky, 1978:86). Adaptive AI feedback, which can be adjusted in detail and explicitness, is potentially well suited to this kind of scaffolding.

The form of feedback also matters. Ellis proposed a typology of written corrective feedback that distinguishes, among others, direct and indirect correction, metalinguistic explanation, focused and unfocused feedback, electronic feedback and reformulation (Ellis, 2009:98–100). Contemporary AI tools can provide almost all of these types, sometimes within a single response. Finally, research in higher education highlights the learner's role. Carless and Boud define student feedback literacy as the understanding, capacities and dispositions needed to make sense of feedback and use it to improve work, and they describe its components as appreciating feedback, making judgements, managing affect and taking action (Carless & Boud, 2018:1316–1319). This concept is crucial for AI feedback, because learners must decide which suggestions to accept, question or reject.

The first generation of AI-powered feedback was represented by AWE systems such as Criterion, MY Access! and later Grammarly and Pigai, which combine natural language processing with scoring algorithms. Early classroom studies showed mixed results. Warschauer and Grimes found that AWE software encouraged students to revise more, but that revisions were mostly limited to surface-level editing rather than substantial changes in content and organisation (Warschauer & Grimes, 2008:32–34). Chen and Cheng reported that the perceived effectiveness of AWE in Taiwanese EFL classes depended strongly on how teachers integrated it into instruction, and that students found its feedback less useful for content and organisation (Chen & Cheng, 2008:104–106).

Reviews of this early research were cautious. Stevenson and Phakiti concluded that there was only modest evidence that AWE feedback improves the quality of texts written with the software and little evidence that improvements transfer to writing proficiency more generally (Stevenson & Phakiti, 2014:62–63). Comparative studies also revealed accuracy problems: Dikli and Bleyle found that automated feedback

on L2 essays sometimes missed errors or flagged correct forms, whereas instructor feedback was more precise (Dikli & Bley, 2014:14–15). Ranalli showed that students could use automated corrective feedback successfully in many cases, but that generic and inaccurate comments made correction considerably more difficult (Ranalli, 2018:666–668).

The introduction of generative AI represents a qualitative shift. Barrot notes that ChatGPT can provide explanations, examples and alternative formulations, adapt its feedback to the learner's request and address higher-order aspects of writing, although it may also produce inaccurate claims and encourage over-reliance (Barrot, 2023:3–5). Su, Lin and Lai found that ChatGPT can assist students in argumentative writing by supporting outlining, content revision and proofreading, but they stress the need for careful pedagogical design (Su, Lin & Lai, 2023:5–7). Comparative studies have begun to examine the quality of such feedback. Guo and Wang compared ChatGPT-generated and teacher feedback on EFL essays and found that ChatGPT produced a considerably larger amount of feedback that addressed content, organisation and language, while teachers' views on its usefulness were mixed (Guo & Wang, 2024:8446–8452). Steiss and colleagues reported that feedback from well-trained human evaluators was of higher quality on most criteria, but that ChatGPT feedback was not far behind and was produced much more quickly (Steiss et al., 2024:9–10).

Meta-analytic evidence provides the most reliable picture of overall effectiveness. Fleckenstein, Liebenow and Meyer synthesised 20 experimental and quasi-experimental studies and found a medium positive effect of automated feedback on students' writing performance ($g = 0.55$), together with significant heterogeneity between studies (Fleckenstein, Liebenow & Meyer, 2023:1). Focusing specifically on EFL and ESL learners, Ngo, Chen and Lai reported a medium between-group effect ($g = 0.59$) and a large within-group effect ($g = 0.98$) of AWE on writing performance (Ngo, Chen & Lai, 2024:727). Their moderator analysis showed that AWE was more effective for vocabulary use than for grammar, that medium and long periods of use produced larger effects than short interventions, and that undergraduate, EFL and intermediate-level learners benefited particularly (Ngo, Chen & Lai, 2024:736–738). These results are illustrated in Figure 1.

Beyond overall scores, AI feedback appears to influence several dimensions of writing development. The first is linguistic accuracy and lexical range. Because AI tools identify errors instantly and suggest alternatives, learners receive far more language-

related input than a teacher can provide. Li, Link and Hegelheimer found that the use of Criterion was associated with an increase in the number of revisions and improvements in linguistic accuracy, and that the way the instructor framed the tool influenced students' use of it (Li, Link & Hegelheimer, 2015:13–15).

The second dimension concerns revision behaviour and the division of labour between teachers and technology. Wilson and Czik showed that when automated feedback was combined with teacher feedback, teachers were able to devote a greater share of their comments to higher-level concerns such as ideas and organisation (Wilson & Czik, 2016:104–105). Link, Mehrzad and Rahimi similarly observed that AWE allowed teachers to shift attention away from surface errors, which influenced both students' revisions and their writing improvement (Link, Mehrzad & Rahimi, 2022:605–606). This suggests that AI feedback does not simply replace teacher feedback but can reshape it productively.

The third dimension is learner engagement. Zhang and Hyland examined students' affective, behavioural and cognitive engagement with teacher and automated feedback and found that the more engaged learner made greater progress, which indicates that the benefits of feedback depend on how deeply it is processed (Zhang & Hyland, 2018:99–100). In a case study of Grammarly, Koltovskaia found that some students engaged critically with automated suggestions, whereas others accepted them uncritically without verifying their correctness (Koltovskaia, 2020:8–10).

The fourth dimension is self-regulation. Nicol and Macfarlane-Dick argue that good feedback practice helps learners clarify goals, assess their own performance and close the gap between current and desired performance (Nicol & Macfarlane-Dick, 2006:205). When learners are asked to formulate questions to an AI tool, compare its suggestions with their own intentions and justify their revision decisions, AI feedback can strengthen these self-regulatory processes. Escalante, Pack and Barrett found no significant difference in writing outcomes between learners who received AI-generated feedback and those who received feedback from human tutors, while learners' preferences for the two sources were divided (Escalante, Pack & Barrett, 2023:10–12). This finding suggests that AI feedback can be a viable complement to human feedback when it is embedded in a well-designed learning cycle, as shown in Figure 2.

The positive findings must be balanced against serious limitations. First, AI feedback is not always accurate. Earlier AWE systems produced false alarms and missed errors

(Dikli & Bleyle, 2014:14), and generative models can produce fluent but misleading explanations or invented rules (Kasneci et al., 2023:5). Learners with lower proficiency are especially vulnerable because they may lack the knowledge to detect such errors. Second, excessive dependence on AI can reduce learners' own cognitive effort. If a tool rewrites whole paragraphs, the learner may produce a better text without developing the underlying competence, which recalls the long-standing concern that correction does not automatically lead to acquisition (Truscott, 1996:341).

Third, there are ethical and institutional concerns. Yan observed that students in an L2 writing practicum appreciated the efficiency of ChatGPT but also expressed concerns about academic honesty and educational equity (Yan, 2023:13943–13944). Questions of data privacy and the homogenisation of learners' written voice also remain unresolved. Fourth, the evidence base is still developing: the heterogeneity reported in meta-analyses (Fleckenstein, Liebenow & Meyer, 2023:12) and the variation across tools and learner groups (Ngo, Chen & Lai, 2024:738) indicate that AI feedback is not a single, uniform intervention and that its effects depend heavily on implementation.

The reviewed research suggests that AI-powered feedback is most effective when integrated into a structured cycle that keeps the teacher and learner in control. Teachers should first develop learners' feedback literacy by explaining the strengths and weaknesses of AI tools and by modelling how to evaluate suggestions critically (Carless & Boud, 2018:1320). Second, AI can be assigned primarily to lower-level concerns, such as grammar, spelling and word choice, while teachers focus their limited time on ideas, argumentation and organisation (Wilson & Czik, 2016:105). Third, learners should be required to document the revision process, for example through revision logs in which they record which suggestions they accepted or rejected and why. Such documentation promotes cognitive engagement and supports academic integrity.

Fourth, prompts for generative AI should be designed pedagogically. Instead of asking a tool to "correct" a text, learners can request indirect feedback, explanations of specific errors or questions about the clarity of their argument, which leaves the actual revision work to the learner (Guo & Wang, 2024:8455). Finally, the duration of use matters: because longer interventions produce larger effects (Ngo, Chen & Lai, 2024:737), AI feedback should be embedded throughout a writing course rather than used occasionally.

CONCLUSION

This article has examined the influence of AI-powered feedback on English language learners' writing development. The theoretical review showed that feedback promotes writing development by drawing attention to linguistic gaps, supporting the recursive processes of planning, drafting and reviewing, and providing scaffolding within the learner's zone of proximal development.

The empirical evidence indicates that AI-powered feedback has a moderate positive effect on writing performance, with especially clear benefits for vocabulary use and surface-level accuracy, as well as for revision activity and learner engagement. Generative AI has extended these possibilities to content and organisation, and its feedback quality is approaching that of trained human evaluators. However, the effects are not automatic. They depend on the accuracy of the feedback, the duration and design of its integration, learners' proficiency and feedback literacy, and, above all, the mediating role of the teacher.

Accordingly, AI-powered feedback should be viewed not as a replacement for teacher feedback but as a component of an integrated human-AI model in which technology handles frequent, form-focused responses while teachers provide higher-order guidance and cultivate learners' critical engagement. Future research should conduct longitudinal studies of generative AI feedback, investigate its effects on learners at different proficiency levels and in different educational contexts, including Central Asian universities, and examine how explicit training in feedback literacy influences learners' uptake of AI-generated comments.

REFERENCES

1. Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
2. Hyland, K. & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83–101. <https://doi.org/10.1017/S0261444806003399>
3. Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327–369. <https://doi.org/10.1111/j.1467-1770.1996.tb01238.x>
4. Ferris, D. (1999). The case for grammar correction in L2 writing classes: a response to Truscott (1996). *Journal of Second Language Writing*, 8(1), 1–11. [https://doi.org/10.1016/S1060-3743\(99\)80110-6](https://doi.org/10.1016/S1060-3743(99)80110-6)
5. Bitchener, J. & Ferris, D. R. (2012). *Written Corrective Feedback in Second Language Acquisition and Writing*. New York: Routledge.
5. Barrot, J. S. (2023). Using ChatGPT for second language writing: pitfalls and potentials. *Assessing Writing*, 57, 100745. <https://doi.org/10.1016/j.asw.2023.100745>

6. Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
7. Swain, M. (1995). Three functions of output in second language learning. In: G. Cook & B. Seidlhofer (eds.), *Principle and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson*, pp. 125–144. Oxford: Oxford University Press.
8. Flower, L. & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365–387. <https://doi.org/10.2307/356600>
9. Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
10. Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal*, 63(2), 97–107. <https://doi.org/10.1093/elt/ccn023>
11. Carless, D. & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
12. Warschauer, M. & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, 3(1), 22–36.
13. Chen, C.-F. E. & Cheng, W.-Y. E. (2008). Beyond the design of automated writing evaluation: pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12(2), 94–112.
14. Stevenson, M. & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19, 51–65.
15. Dikli, S. & Bley, S. (2014). Automated essay scoring feedback for second language writers: how does it compare to instructor feedback? *Assessing Writing*, 22, 1–17.
16. Ranalli, J. (2018). Automated written corrective feedback: how well can students make use of it? *Computer Assisted Language Learning*, 31(7), 653–674.
17. Su, Y., Lin, Y. & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, 100752. <https://doi.org/10.1016/j.asw.2023.100752>
18. Guo, K. & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435–8463. <https://doi.org/10.1007/s10639-023-12146-0>
19. Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J. et al. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
20. Fleckenstein, J., Liebenow, L. W. & Meyer, J. (2023). Automated feedback and writing: a multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, 6, 1162454.
21. Ngo, T. T.-N., Chen, H. H.-J. & Lai, K. K.-W. (2024). The effectiveness of automated writing evaluation in EFL/ESL writing: a three-level meta-analysis. *Interactive Learning Environments*, 32(2), 727–744. <https://doi.org/10.1080/10494820.2022.2096642>
22. Li, J., Link, S. & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, 27, 1–18.
23. Wilson, J. & Czika, A. (2016). Automated essay evaluation software in English Language Arts classrooms: effects on teacher feedback, student motivation, and writing quality. *Computers & Education*, 100, 94–109.
24. Link, S., Mehrzad, M. & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 35(4), 605–634.

- 
25. Zhang, Z. V. & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90–102.
26. Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: a multiple case study. *Assessing Writing*, 44, 100450. <https://doi.org/10.1016/j.asw.2020.100450>
27. Nicol, D. J. & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
28. Escalante, J., Pack, A. & Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, Article 57.
29. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F. et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
30. Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: an exploratory investigation. *Education and Information Technologies*, 28(11), 13943–13967.