Feedback Research in Second Language

EUROKD

2026 (4), 6–15
https://doi.org/10.32038/frsl.2026.04.02

AI-Mediated Written Corrective Feedback in L2 Writing: What Changes, What Doesn't, and What We Still Don't Know

Konstantinos M. Pitychoutis*
Liberal Arts Department, American University of the Middle East, Egaila 54200, Kuwait
Filomachi Spathopoulou
Liberal Arts Department, American University of the Middle East, Egaila 54200, Kuwait

Correspondence
Email: Konstantinos-p@aum.edu.kw

Abstract
This reflective commentary revisits key insights from research on written corrective feedback (WCF) within the context of feedback facilitated by large language models (LLMs) in higher-education L2 writing. The CAF framework shows that accuracy is most consistently improved through targeted, well-designed feedback. In contrast, claims about complexity and fluency should be approached cautiously, as they depend on factors like task design, sequencing, and learner engagement. The paper distinguishes between traditional automated writing assessments and the generative, dialogic features of LLMs. It proposes a straightforward approach for teachers to provide feedback, including specific prompts, metalinguistic explanations, learner reformulation, genre confirmation, and reflection periods. The paper emphasises that the educational value of LLM-based WCF relies more on thoughtful mediation, ethical considerations, and alignment with assessment goals than on speed or novelty. It concludes by highlighting priority areas for future research, such as the durability of learning gains, transfer from assisted to independent writing, the interaction of dose, timing, and scope, maintaining learner engagement, and ensuring equitable access to AI-mediated feedback across proficiency levels.

ARTICLE HISTORY
Received: 11 August 2025
Revised: 25 March 2026
Accepted: 10 May 2026

KEYWORDS
Written Corrective Feedback, Large Language Models (LLMs),Complexity-Accuracy-Fluency (CAF), Feedback Literacy, Second-Language Writing, Assessment Integrity

How to cite this article (APA 7th Edition):
Pitychoutis, K. M., & Spathopoulou, F. (2026). AI-mediated written corrective feedback in L2 writing: What changes, what doesn't, and what we still don't know. *Feedback Research in Second Language, 4,* 6–15. https://doi.org/10.32038/frsl.2026.04.02

Introduction

As university educators with extensive experience in higher education across the Arab Gulf and Europe, we reflect on the implications of the rapid adoption of large language models such as ChatGPT, Claude, Gemini, and Perplexity. These tools revive longstanding debates about the role of written corrective feedback (WCF) in second-language (L2) writing. Initially, they appear to build on early automated writing support by offering faster, more personalised, and interactive responses. However, the core pedagogical question is not just whether LLMs can efficiently identify and correct errors. More critically, we must ask whether their feedback facilitates the noticing, interpretation, and revision processes linked to language development in WCF research (Ma et al., 2024). Our background, let alone our age, places us at the crossroads of traditional language teaching and cutting-edge educational technology, raising the key question of this essay: Is AI-driven feedback a true solution for language learning or just an unrealised promise?

This question matters because WCF has always been more than just mechanically fixing surface errors. Hyland and Hyland (2006) noted that feedback is relational, context-specific, and influenced by rhetorical factors; it encompasses elements such as stance, mitigation, timing, genre, and the learner's evolving sense of authorship. Therefore, feedback from LLMs should not be viewed as a neutral technical tool. Instead, it operates within a complex educational environment in which correction, explanation, and the learner's identity are deeply interconnected.

The most reliable result in WCF research is accuracy. Targeted, clear feedback, particularly when accompanied by metalinguistic explanations, has demonstrated efficacy in improving accuracy in certain linguistic domains, with findings suggesting that these enhancements may persist beyond immediate corrections (Bitchener & Knoch, 2010; Shintani et al., 2013). This collection of studies establishes an essential foundation for contemporary discourse on LLMs. The main question is not if AI can make writing better in general, but if it can help with specific, well-thought-out, practical, and long-lasting changes.

Moreover, LLMs change the context in which feedback is delivered. Unlike earlier automated writing evaluation tools, they can sustain dialogue, offer explanations on demand, and produce multiple reformulations in real time. While these capabilities may encourage more practice and quicker revisions, they do not eliminate the need for pedagogical guidance. In fact, the adaptability of LLM outputs increases the importance of teacher oversight. Issues related to construct validity, disciplinary appropriateness, register, and assessment integrity remain crucial, even with a more conversational interface.

This reflection addresses three related questions: Which principles from the WCF literature remain relevant in AI-driven settings? What aspects of LLM-based feedback indicate a genuine shift rather than just extending previous AWE concepts? And what

does current evidence say about claims regarding complexity, accuracy, and fluency? The paper advocates a cautious yet pragmatic approach: view LLMs as educational tools integrated into well-structured feedback systems, not as direct replacements for teachers' judgement.

Literature Review

Framing WCF in the LLM Era

Any discussion of LLM-mediated written corrective feedback (WCF) should begin with a broader debate about its value, scope, and supporting evidence in second-language (L2) writing. The debate between Truscott (1996, 1999) and Ferris (1999, 2004) did more than polarise opinions; it shifted the discussion from whether grammar correction “works” to more specific issues, including the types of feedback, the linguistic areas targeted, implementation conditions, and evidence of learning. Later research has viewed WCF as a range of pedagogical approaches rather than a single method, varying in directness, scope, timing, and explicitness (Ellis, 2009). More recently, Liu (2025), in a bibliometric analysis of 321 studies indexed between 2014 and 2024, traced how the field’s dominant themes have shifted away from the binary efficacy debate towards more granular concerns: student engagement with feedback, the theoretical foundations underpinning WCF, and the situated factors shaping teachers’ practices. This evolution suggests a maturation of the discipline that should inform, rather than be displaced by, current discussions of generative AI. The strongest evidence in the current literature mainly emphasises increased accuracy rather than overall writing quality. Focused, clear feedback, especially with metalinguistic explanations, has been linked to improvements in specific language skills. However, the durability and scope of these improvements depend on design and learner factors (Bitchener & Knoch, 2010; Brown et al., 2026). This topic is particularly important in current debates about AI. Despite new delivery methods, LLM feedback still needs to be interpreted within the framework established by WCF research. In essence, the advent of LLMs does not diminish the need for theoretical clarity; rather, it underscores its continued importance.

From AWE to LLMs: Continuity and Inflexion

LLMs should be seen as an evolution of automated writing support, not a complete break from earlier feedback tools. Before generative AI, systems such as Criterion, WritingRoadmap, Pigai, and Grammarly provided scalable, rapid, and partially automated feedback, yet they had limitations, including shallow diagnostic capabilities, inconsistent sensitivity to genre, and reliance on teachers for meaningful use (Cotos, 2023; Karatay & Karatay, 2024; Koltovskaia, 2020). Unlike earlier tools, LLMs enable ongoing, context-aware, multi-turn interactions. Students can ask follow-up questions, clarify points, and make corrections, receiving immediate responses that mimic a dialogue with an expert teacher. This development does not replace teacher involvement but offers new opportunities for organised, scaffolded support (Pitychoutis, 2024; Teng, 2024). The change happens gradually; fundamental pedagogical principles of effective feedback still apply, with the environment adapting

to uphold them (Crosthwaite & Sun, 2026). Ultimately, the main concern is not whether LLMs surpass earlier tools in complexity, but whether their conversational skills can be integrated into feedback processes without compromising construct alignment, genre awareness, learner agency, or assessment validity. Their value extends beyond just automation, facilitating well-managed instructional feedback routines (Pitychoutis et al., 2025).

What Recent AI-WCF Studies Suggest (2023–2025)

Recent research on AI-supported WCF shows a cautiously optimistic yet still emerging evidence base. Lin and Crosthwaite (2024) found that teacher feedback and GPT-assisted feedback are neither simply better nor worse; they differ in tone, detail, and directness. This suggests that hybrid methods might be more effective than replacing one with the other. Yan and Zhang (2024) noted that, from the learner's perspective, just accepting feedback does not fully reflect learning progress. Engagement with ChatGPT-mediated feedback depends on the learner's prompting skills, metacognitive strategies, and willingness to critically evaluate suggestions rather than passively accept them.

Recent studies highlight feedback design factors. ElEbyary et al. (2024) suggested that the mode and timing of AI-driven WCF influence error detection and writing quality. Kamelabad et al. (2026) compared immediate and delayed LLM-generated corrective feedback, finding similar motivational and learning effects, which implies that interaction design may matter more than timing alone. Meyer et al. (2024) argue that AI feedback can boost revision, motivation, and positive emotions when it is timely, clear, and integrated into meaningful pedagogical tasks.

These findings do not support the idea that AI can fully replace teacher feedback or ensure growth in all writing skills. Instead, a balanced perspective suggests that LLM-supported WCF works best when it is targeted, interactive, and integrated into teaching routines that include explanation, oversight, and critical reflection. Overall, recent research aligns with earlier WCF studies: improving accuracy offers immediate benefits, learner engagement is vital, and teacher involvement remains crucial.

A CAF lens: what we can and cannot infer

The CAF (complexity, accuracy, and fluency) framework offers a useful method for analysing AI-supported WCF. However, the reliability of its components varies: accuracy, which correlates with traditional corrective feedback, is easiest to assess using error-correction and rule-based techniques. In contrast, measuring complexity and fluency is more challenging in AI-supported environments. Gains in syntactic complexity or fluency may stem from model-assisted rewriting, repeated practice, or prompt adjustments rather than true progress in the learner's interlanguage.

Caution is advised. While broader WCF research recommends careful interpretation of accuracy, it does not imply that higher accuracy automatically enhances complexity,

fluency, or overall writing quality (Brown et al., 2026). When working with large language models (LLMs), it is important to differentiate between performance improvements due to assistance and actual learning progress. A learner might produce more fluent or complex revisions with AI support without demonstrating similar gains in unaided writing. The current consensus suggests that LLM-based WCF can effectively boost accuracy-focused revision. However, any improvements in complexity and fluency are more dependent on design choices, are less certain, and need long-term evidence. Table 1 summarises what changes, what remains unchanged, and which questions are still unanswered with AI-supported WCF.

Table 1

AI-Mediated WCF at a Glance: What Changes, what Stays, and what We still do not Know

Category	Key points
What changes	Feedback becomes faster, more scalable, and more dialogic; explanations can be generated on demand, and revision can unfold across multiple turns.
What stays	Effective WCF still requires focus, explicitness, teacher mediation, genre sensitivity, and attention to ethics and assessment integrity.
What we still don't know	Transfer to unaided writing, durability of gains, interactions among timing, dose, and scope, proficiency-related differences, and the validity of broader claims about complexity and fluency.

Design Principles for AI-Mediated WCF

For LLM-mediated written corrective feedback to remain pedagogically justified, it should be grounded in principles from the WCF literature rather than in technological convenience. The key concern is not just whether LLMs can produce corrections swiftly, but how this feedback can be structured to enhance noticing, maintain construct validity, and uphold the teacher's interpretive role.

Keep WCF focused and explicit

The most compelling evidence in WCF literature highlights feedback that is targeted, explicitly phrased, and focused on specific error types. In AI-supported settings, this entails using precise and intentionally clear prompts. Instead of requesting broad revisions, teachers can direct the model to focus on one particular aspect, explain the relevant rule, and suggest a single correction. This approach minimises overcorrection, aligns feedback with educational goals, and increases the likelihood that revisions genuinely improve accuracy rather than resulting in aimless rewriting.

Prioritise noticing and learner agency

Correction on its own does not equal learning. What matters is whether learners actively engage with feedback, interpret it carefully, and apply it to their writing. Therefore, AI-assisted WCF should foster active participation rather than passive listening. A helpful approach is to have learners rephrase a rule, justify a suggested change, or apply the same correction to a different sentence. These tasks of explaining and transferring knowledge ensure that feedback leads to noticing and comprehension, rather than just superficial improvements.

Guard against construct drift

Because LLMs produce fluent, often rhetorically persuasive writing, they can unintentionally alter the register, stance, genre, or rhetorical structure of learner writing. This raises validity concerns, particularly in assessed writing, where the overall construct extends beyond grammatical accuracy. Therefore, AI-assisted WCF should align with the task's specific goals and rubric criteria. In practice, teachers need to guide the model towards specific linguistic objectives, request explanations for suggested edits, and review a sample of revisions to ensure standards for discipline and genre are upheld. The purpose of this oversight is not to overly restrict feedback, but to prevent superficial improvements from compromising the core construct.

Stage hybrid feedback

Recent research shows that teacher feedback and LLM-driven feedback work best as complementary tools rather than as competitors. A practical approach begins with the learner drafting their work, then using the LLM for specific corrections, making targeted revisions, and finally submitting the revised version for a brief teacher review, task alignment, and reflection. This phased process preserves the teacher's role in interpretation and judgement while leveraging AI's speed and accessibility during initial revisions. It also reinforces the main point of this paper: LLMs are most effective when incorporated into well-structured pedagogical routines, rather than used as standalone evaluators.

Be explicit about ethics and assessment integrity

An account of AI-supported WCF must clearly address ethics and integrity in assessment. Students need explicit guidance on when AI-supported feedback is permitted, which forms of help are acceptable, how to disclose their use of AI, and how to handle data appropriately. This includes advising against sharing personal or sensitive information in prompts, using only approved institutional platforms where possible, and keeping a brief log of AI use. Although these measures do not eliminate all risks, they help define the boundaries of ethical and pedagogical use of AI. Transparency in these practices is vital for responsible implementation.

What the field still needs: A research agenda

Despite growing interest in AI-supported written corrective feedback (WCF), empirical evidence remains limited, making it difficult to definitively assess its developmental benefits. The key question is whether improvements observed during AI-assisted revisions persist in later writing without AI. Until long-term studies are conducted, claims that large language model (LLM) feedback improves learner accuracy should be treated cautiously.

Another issue concerns design variables. While timing, scope, and feedback volume are well researched in traditional WCF, these factors may not directly apply to LLM-based contexts. Generative systems differ from earlier tools not only in speed but also in responsiveness, dialogic interaction, and opportunities for iterative clarification.

More systematic experiments are needed to understand how dose, timing, and scope interact in LLM environments, rather than assuming they are similar to previous correction methods (ElEbyary et al., 2024; Kamelabad et al., 2026).

A thorough understanding of learner engagement is essential. While current research emphasises uptake and behaviours, it offers limited insight into how learners interpret, evaluate, and internalise AI feedback. Cognitive processes such as metalinguistic reflection or uncritical acceptance can produce similar revision outcomes. Methods such as keystroke logging, eye-tracking, screen recording, and stimulated recall can reveal levels of engagement during AI feedback sessions (Yan & Zhang, 2024). Reynolds (2023) makes a compelling case for triangulating eye-tracking with qualitative data to capture the moment-by-moment cognitive processing of WCF, an approach particularly well suited to AI-mediated contexts, where the rapidity and density of feedback risk obscuring the cognitive work learners are, or are not, doing. Additionally, equity issues need further exploration. AI-driven WCF may benefit learners with strong linguistic, digital, or prompting skills, potentially providing less support to those with fewer resources, thereby widening the gap across proficiency levels and contexts.

Future research should assess not only the effectiveness of AI-enabled WCF but also consider learners, their environments, and the intended outcomes. Moreover, the construct validity of AI-supported feedback remains underexplored. While most research centres on sentence-level corrections, writing assessments often require considerations of stance, organisation, genre awareness, and rhetorical skills. Enhancing grammar with AI does not automatically mean the work aligns with the task's construct or rubric. Future studies should investigate whether AI-mediated WCF truly fosters growth in fundamental writing skills or just produces polished texts.

Final Reflections

LLM-based feedback should not be dismissed as a passing trend nor treated as a complete solution to longstanding challenges in L2 writing education. Its main advantages are the speed, accessibility, and conversational versatility it offers for revision, explanations, and repeated practice. This means LLMs indeed alter the practical conditions for providing written corrective feedback. Nonetheless, the core pedagogical principles remain unchanged: effective feedback still depends on focus, clarity, learner engagement, alignment with learning objectives, and teacher guidance. AI mainly changes the environment in which these principles are applied, not the principles themselves.

Therefore, a cautious-optimism approach is best. LLMs can be particularly valuable for supporting accuracy-focused revision and enabling more frequent, timely interactions with language forms. However, claims of broader improvements in complexity, fluency, and overall writing development are less certain, especially when assistance is mistaken for deep learning. Moreover, the efficiency of AI-generated

feedback should not obscure ongoing concerns about trust, dependency, ethics, and assessment validity.

In this context, AI-mediated WCF is unlikely to replace teachers. Instead, it will foster disciplined hybrid practices in which LLMs serve as responsive teaching tools under human supervision. The discipline has already begun to recognise promising areas and the need for caution. It appears that feedback's immediacy and interactivity are changing, while the need for research-based, carefully guided practice remains. The unresolved question is whether this support produces lasting, transferable, and equitable growth in learners' writing.

ORCID

https://orcid.org/0000-0002-2761-5764

https://orcid.org/0000-0002-1826-7753

Publisher's Note

The claims, arguments, and counter-arguments made in this article are exclusively those of the contributing authors. Hence, they do not necessarily represent the viewpoints of the authors' affiliated institutions, or EUROKD as the publisher, the editors and the reviewers of the article.

Acknowledgements

We would like to thank the editors for their comments and support.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRediT Authorship Contribution Statement

Konstantinos M. Pitychoutis: Conceptualisation, Methodology, Writing Original Draft, Review & Editing, Supervision

Filomachi Spathopoulou: Conceptualisation, Writing Original Draft, Review & Editing

Generative AI Use Disclosure Statement

The authors used Grammarly Premium for proofreading and editing this manuscript. All substantive intellectual content, argumentation, and conclusions are the authors' own. The authors take full responsibility for the accuracy and integrity of the published work.

Ethics Declarations

World Medical Association (WMA) Declaration of Helsinki–Ethical Principles for Medical Research Involving Human Participants

This paper is a reflective theoretical commentary involving no primary research with human participants. The Declaration of Helsinki is therefore not applicable.

Competing Interests

The authors declare no competing interests.

Data Availability

No primary data were generated or analysed in the preparation of this manuscript.

References

Bitchener, J., & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing, 19*(4), 207–217. https://doi.org/10.1016/j.jslw.2010.10.002

Brown, D. M., Liu, Q., & Norouzian, R. (2026). Effectiveness of written corrective feedback in developing L2 accuracy: A Bayesian meta-analysis. *Language Teaching Research, 30*(3), 1357–1389. https://doi.org/10.1177/13621688221147374

Cotos, E. (2023). Automated feedback on writing. In O. Kruse, et al. (Eds.), *Digital writing technologies in higher education* (pp. 347–364). Springer, Cham. https://doi.org/10.1007/978-3-031-36033-6_22

Crosthwaite, P., & Sun, S. (2026). Generative AI and L2 written feedback studies: A scoping review. *RELC Journal*. Advance online publication. https://doi.org/10.1177/00336882251386530

ElEbyary, K., Shabara, R., & Boraie, D. (2024). The differential role of AI-operated WCF in L2 students' noticing of errors and its impact on writing scores. *Language Testing in Asia, 14*(1), 59. https://doi.org/10.1186/s40468-024-00312-1

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

Ferris, D. R. (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

Ferris, D. R. (2004). The "grammar correction" debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime?). *Journal of Second Language Writing, 13*(1), 49–62. https://doi.org/10.1016/j.jslw.2004.04.005

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

Kamelabad, A. M., Turano, B., Lundin, M., & Skantze, G. (2026). Personalized language learning with an LLM chatbot: Effects of immediate vs. delayed corrective feedback. *Frontiers in Education, 11*, Article 1703664. https://doi.org/10.3389/feduc.2026.1703664

Karatay, Y., & Karatay, L. (2024). Automated writing evaluation use in second language classrooms: A research synthesis. *System, 123*, Article 103332. https://doi.org/10.1016/j.system.2024.103332

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing, 44*, Article 100450. https://doi.org/10.1016/j.asw.2020.100450

Lin, S., & Crosthwaite, P. (2024). The grass is not always greener: Teacher vs. GPT-assisted written corrective feedback. *System, 127*, Article 103529. https://doi.org/10.1016/j.system.2024.103529

Liu, W. (2025). A bibliometric analysis of written corrective feedback in second language writing. *Language Teaching Research Quarterly, 49*, 133–150. https://doi.org/10.32038/ltrq.2025.49.07

Ma, Q., Crosthwaite, P., Sun, D., & Zou, D. (2024). Exploring ChatGPT literacy in language education: A global perspective and comprehensive approach. *Computers & Education: Artificial Intelligence, 7*, Article 100278. https://doi.org/10.1016/j.caeai.2024.100278

Meyer, J., Jansen, T., Schiller, R., et al. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers & Education: Artificial Intelligence, 6*, Article 100199. https://doi.org/10.1016/j.caeai.2023.100199

Pitychoutis, K. M. (2024). Harnessing AI chatbots for EFL essay writing: A paradigm shift in language pedagogy. *Arab World English Journal (AWEJ)* Special Issue on ChatGPT, 197-209. https://dx.doi.org/10.24093/awej/ChatGPT.13

Pitychoutis, K. M., Topalidou, S., & Spathopoulou, F. (2025). Empowering EFL exam preparation with AI: Practical tips for C2 classrooms. *TESOL Greece Journal,* 167, 20-26. https://dx.doi.org/10.2139/ssrn.5244479

Reynolds, B. L. (2023). Exploring learner attention and processing in second language writing: The role of eye-tracking and written corrective feedback. *Feedback Research in Second Language, 1*, 226–235. https://doi.org/10.32038/frsl.2023.01.12

Shintani, N., Ellis, R., & Suzuki, W. (2013). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning, 64*(1), 103-131. https://doi.org/10.1111/lang.12029

Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences of ChatGPT-mediated feedback. *Computers & Education: Artificial Intelligence, 7*, Article 100278. https://doi.org/10.1016/j.caeai.2024.100270

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

Truscott, J. (1999). The case for "the case against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing, 8*(2), 111–122. https://doi.org/10.1016/S1060-3743(99)80124-6

Yan, D., & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. *Humanities & Social Sciences Communications, 11*, 1086. https://doi.org/10.1057/s41599-024-03543-y