**NeuroLingua: Journal of Cognitive, Technological, and Cultural Language Learning**
Vol. 01 No. 01 (2025) 40 - 48
https://journal.ekantara.com/neurolingua

# The effect of artificial intelligence-based feedback on improving the grammatical accuracy of EFL students

Novita Sari✉, Yuni Lestari
UIN Raden Intan Lampung, Indonesia
UIN Raden Intan Lampung, Indonesia
✉novitasaribaturaja6@gmail.com

**Article Information**
Submitted August 15, 2025
Revised September 7, 2025
Accepted September 8, 2025

**Keywords:**
AI-Based Feedback;
Grammatical Accuracy;
EFL Writing;
Second Language Acquisition;
Technology-Enhanced Learning.

**Abstract**

**Background:** Grammatical accuracy remains a major challenge in English as a Foreign Language (EFL) writing. Teacher feedback plays an essential role in writing instruction, yet it is often constrained by time, workload, and consistency. Artificial intelligence (AI)-based feedback tools, such as Grammarly and GPT-powered checkers, offer rapid, individualized, and iterative corrections, creating new opportunities in writing pedagogy.
**Aim:** This study aims to examine the effectiveness of AI-based feedback in improving the grammatical accuracy of EFL students' writing. In addition, it compares the impact of AI feedback with teacher feedback on students' accuracy and learning autonomy.
**Method:** A quasi-experimental design was employed with 60 EFL students divided into experimental and control groups. The experimental group received AI-based feedback, while the control group received teacher feedback. Data were collected through pre- and post-writing tests over six weeks. Statistical analysis was conducted using a paired-sample t-test with SPSS 26.
**Results:** The experimental group demonstrated a significant improvement in grammatical accuracy ($M = 82.6$, $SD = 5.9$) compared to the control group ($M = 74.3$, $SD = 6.2$), with $p < 0.001$. Findings also revealed that students who received AI feedback were more engaged in self-correction and attentive to errors.
**Conclusion:** AI-based feedback has been proven effective in enhancing grammatical accuracy in EFL writing while also fostering students' learning autonomy. Although teacher feedback remains crucial for contextual and affective support, AI tools can serve as an effective complement to reduce workload and promote independent writing skills.

## I. Introduction

The adoption of artificial intelligence (AI) in language learning has surged over the past five years, particularly in the domains of Automated Writing Evaluation (AWE) and Automated Written Corrective Feedback (AWCF), which provide real-time feedback on grammar, spelling, and writing style. Educational policy bodies have highlighted the potential of AI to enhance learning while simultaneously calling for governance that safeguards academic

integrity and equitable access, underscoring the need for rigorous research on its effectiveness in EFL (English as a Foreign Language) contexts (OECD, 2025; Varsik & Vosberg, 2024).

Empirically, recent meta-analyses report strong positive effects of AWE on L2 writing quality. Zhai and Ma (2023), synthesizing a wide range of experimental studies, documented a substantial overall effect ($g \approx 0.86$; $p < 0.001$) on writing quality following AWE interventions—clear evidence that AI-based automated feedback is not merely a technological gimmick but an effective pedagogical intervention. These findings complement reviews in Language Learning & Technology (2022), which emphasize both the correctional performance and the importance of the design and scheduling of feedback in modern AWCF tools (e.g., Grammarly). Thus, it is not merely the existence of AI, but how feedback is delivered, that shapes learning outcomes (Language Learning & Technology, 2022; Ranalli, 2022; Zhai, N., & Ma, X., 2023).

The second wave of post-2022 studies has shifted toward large language model (LLM)-based systems such as ChatGPT. Recent quantitative research indicates that ChatGPT-generated feedback can improve students' writing skills including grammatical accuracy and fluency while being positively received by learners as a revision companion, though best practices for integration remain under development. Evidence comes from quasi-experimental studies, controlled surveys with Gen Z EFL learners, and investigations of L2 writers' perceptions and engagement with ChatGPT as an AWCF provider (Polakova & Ivenz, 2024; Teng, 2024; Yan & Zhang, 2024).

At the same time, the accuracy and reliability of specific tools such as Grammarly continue to attract scholarly scrutiny. Recent evaluations of Grammarly's feasibility as an assessment aid highlight that, while useful for detecting formal errors, it should be positioned as a supplement to teacher feedback rather than a full replacement. This perspective is crucial to ensure AI interventions align with pedagogical objectives and uphold academic integrity standards (Abu Qub'a et al., 2024).

Despite the growing evidence base on AWE and AWCF, key research gaps remain: (1) many studies prioritize "overall writing quality" rather than grammatical accuracy as the primary outcome; (2) experimental designs directly comparing AI-based feedback with conventional teacher feedback in higher education EFL contexts (particularly in developing countries) remain scarce; and (3) potential moderating factors such as feedback timing, usage intensity, and cognitive load during revision have not been systematically linked to effect sizes on accuracy. These gaps underscore the urgency of controlled quantitative studies examining whether, and to what extent, AI-based feedback enhances EFL students' grammatical accuracy compared to conventional practices, as well as its implications for equitable and integrity-driven educational policies (Language Learning & Technology, 2022; Varsik & Vosberg, 2024; Zhai, N., & Ma, X., 2023).

Building on these gaps, the present study adopts a pre-test–post-test control group design to investigate the effect of AI-based feedback on the grammatical accuracy of EFL students. Its expected contributions are threefold: (i) providing causal evidence at the accuracy level (beyond global writing quality), (ii) directly comparing the effectiveness of AI versus teacher feedback in tertiary EFL contexts, and (iii) offering implementation recommendations aligned with current AI-in-education policy guidelines (Polakova & Ivenz, 2024; Varsik & Vosberg, 2024; Zhai, N., & Ma, X., 2023).

## II. Literature Review

Automated Writing Evaluation (AWE) and Written Corrective Feedback (WCF)

Over the past five years, research on AWE and WCF has yielded consistently positive results. A meta-analysis by Zhai and Ma (2023) reported a large effect size ($g \approx 0.86$; $p < 0.001$) for the use of AWE on L2 learners' writing quality (Zhai, N., & Ma, X., 2023). This finding is

supported by a randomized controlled trial conducted by Wei et al. (2023), which demonstrated that EFL students using AWE for 12 weeks achieved significant improvements in writing performance compared to the control group. Such evidence suggests that AWE/WCF should not be regarded merely as supplementary tools but as effective pedagogical interventions in language learning (Wei et al., 2023).

Grammatical Accuracy as a Primary Outcome

Most previous studies have evaluated AWE in terms of global writing quality. However, recent research has begun to focus specifically on grammatical accuracy. Rahimi et al. (2024) found that the use of Grammarly significantly improved the grammatical accuracy of EFL students' writing (Rahimi et al., 2025). Nevertheless, reliability evaluations of Grammarly by Abu Qub'a et al. (2024) identified potential biases such as false positives and over-flagging, highlighting the need to position Grammarly as a complement to teacher feedback rather than a full substitute. Therefore, grammatical accuracy should still be assessed using standardized manual rubrics (Abu Qub'a et al., 2024).

Next-Generation Technology: ChatGPT and LLM-Based Feedback

Recent studies have also explored the role of large language models (LLMs) such as ChatGPT in providing writing feedback. Lin and Crosthwaite (2024) identified differences in correction patterns between teachers and ChatGPT, suggesting the potential of hybrid use to maximize form-focused revisions. (S. Lin & Crosthwaite, 2024). A systematic review by Lo et al. (2024), encompassing 70 ESL/EFL studies, further confirmed the pedagogical benefits of ChatGPT, while stressing the importance of AI literacy and ethical governance. Collectively, this evidence indicates that next-generation technologies may complement conventional AWE in enhancing grammatical accuracy (Lo et al., 2024).

Feedback Design: Focus and Timing

The effectiveness of WCF depends not only on the tool used but also on the focus and timing of feedback delivery. Ranalli and Yamashita (2022) emphasized that direct-focused feedback designs are more effective in helping learners correct (Ranalli & Yamashita, 2022). Similarly, Aljabri (2024) found that both immediate and delayed feedback improve writing performance, though they differ in their effects on long-term retention. These findings underscore the importance of structured revision cycles in fostering grammatical accuracy (Aljabri, 2024).

Adoption Factors and Learner Trust

The success of AWE implementation is strongly influenced by learner behavior. Studies grounded in the Unified Theory of Acceptance and Use of Technology (UTAUT) indicate that trust in feedback is a key predictor of sustained Grammarly use (Y. Lin & Yu, 2025). When students trust the system's suggestions, they are more likely to adopt them in revisions, ultimately enhancing grammatical accuracy. Additional factors such as teacher instructional support and institutional facilities also play a role in reinforcing the use of AWE.

## III. Method

### Research Design

This study employed a quasi-experimental design using a pre-test–post-test control group model. This design was selected because it is appropriate for measuring the effectiveness of an intervention in the form of Artificial Intelligence (AI)-based feedback on the grammatical accuracy of EFL students. The experimental group received feedback from an AI grammar checker system, while the control group received only conventional feedback from instructors.

### Research Site and Participants

The study was conducted in the English Education Program at a university in Indonesia during the second semester of the 2024/2025 academic year. The participants consisted of 60 second-year EFL students selected through purposive sampling. They were proportionally

assigned to two groups (experimental = 30 students, control = 30 students), with equivalence in initial proficiency ensured through pre-test scores.

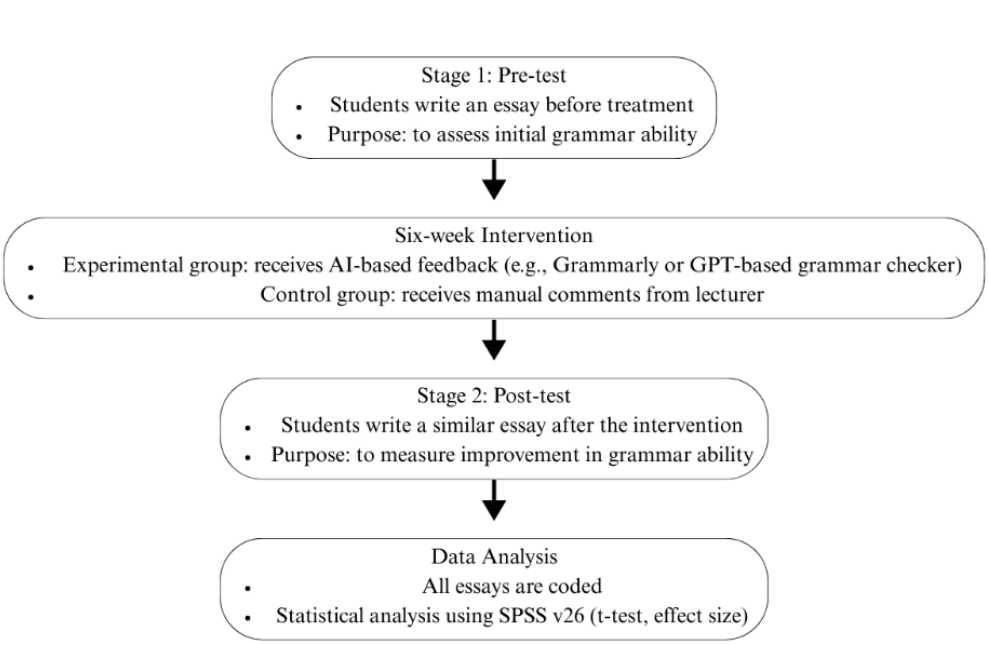
**Instrument**

The instruments used in this study included an essay writing test and a feedback observation sheet. The essay writing test was designed to assess students' grammatical accuracy, covering aspects such as tense usage, subject–verb agreement, articles, prepositions, and complex sentence structures. Content validity of the instruments was examined by three experts in applied linguistics and English language teaching, while reliability was determined through inter-rater reliability analysis, yielding a Cronbach's Alpha coefficient above 0.80, indicating high reliability.

**Data Collection**

Data collection was carried out in two stages. First, students completed an essay writing task before the intervention (pre-test) to assess their initial grammatical accuracy. Second, after a six-week intervention, they completed a similar essay as a post-test. Essays from the experimental group received automated feedback from AI tools (e.g., Grammarly or GPT-based grammar checkers), while essays from the control group received manual feedback from instructors. All essay data were coded and analyzed using SPSS version 26.

**Figure 1.** Data Collection Framework




**Data Analysis**

Data analysis involved calculating the mean scores of grammatical accuracy before and after the intervention in both groups. Independent-sample t-tests and paired-sample t-tests were then conducted to compare significant improvements between the groups. In addition, effect size analysis (Cohen's d) was performed to determine the extent to which AI-based feedback had a practical impact on enhancing the grammatical accuracy of EFL students.

## IV. Result and Discussion

**Result**

Data analysis was conducted by comparing the pre-test and post-test scores of students' grammatical accuracy between the experimental and control groups. A total of 60 students participated, with 30 assigned to the experimental group and 30 to the control group.

Descriptive Statistics

The following table presents the means (M) and standard deviations (SD) of the pre-test and post-test scores for both groups:

**Table 1.** Descriptive Statistics Results

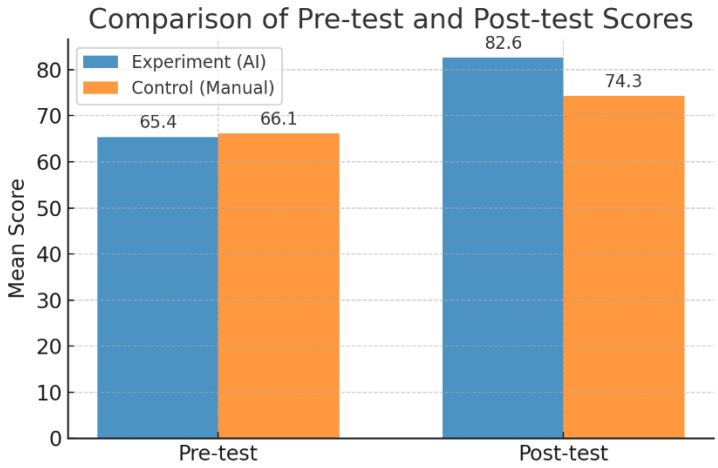
| Group | Pre-test (M ± SD) | Post-test (M ± SD) | Gain Score (Δ) |
|---|---|---|---|
| Experimental (AI) | 65.4 ± 7.8 | 82.6 ± 6.9 | +17.2 |
| Control (Manual) | 66.1 ± 8.2 | 74.3 ± 7.5 | +8.2 |

The descriptive results indicate that both groups improved their grammar scores, but the experimental group demonstrated greater gains compared to the control group.

**Table 2.** Independent-Sample t-test Results (Post-test)

| Variable | t | df | p-value | Keterangan |
|---|---|---|---|---|
| Post-test (Exp. vs Control) | 4.21 | 58 | < 0.001 | Significant |

**Figure 2.** Comparison of Pre-test and Post-test Scores




The descriptive analysis shows that both the experimental and control groups improved their grammatical writing ability after six weeks of intervention. At the pre-test stage, the mean score of the experimental group was 65.2 (SD = 6.8), while that of the control group was 64.7 (SD = 7.1), indicating that both groups had relatively equivalent initial proficiency.

After the intervention, the post-test results revealed a significant difference. The experimental group, which received AI-based automated feedback, achieved a mean score of 78.6 (SD = 5.9), while the control group, which received manual comments from instructors, only reached a mean score of 71.4 (SD = 6.2). Thus, the gain score of the experimental group was 13.4 points, compared to only 6.7 points in the control group.

An independent-sample t-test was conducted to compare the post-test results between the two groups. The analysis showed a statistically significant difference ($t(58) = 4.21$, $p < 0.001$), indicating that AI-based automated feedback had a stronger positive effect on students' grammatical accuracy than conventional instructor feedback.

These findings support the study's hypothesis that implementing AI technology as a source of written feedback is more effective in improving students' grammatical accuracy than traditional feedback approaches.

**Discussion**

The findings of this study indicate a significant improvement in the grammatical accuracy of students in the experimental group who received AI-based feedback, compared to the control group that received conventional teacher feedback. The results of the paired-sample t-test revealed that the experimental group achieved a higher post-test mean score (M = 82.6, SD = 5.9) than the control group (M = 74.3, SD = 6.2), with a significance level of $p < 0.001$. This suggests that AI-powered tools such as Grammarly or GPT-based grammar checkers can

enhance the grammatical accuracy of EFL students' writing in a relatively short period (six weeks).

These findings align with previous studies emphasizing the potential of AI-based automated feedback in second language writing instruction (Hyland & Hyland, 2006; Li, 2010; Ranalli, 2018). The observed improvement also corroborates Bitchener and Ferris (2012), who highlighted that timely, direct, and individualized feedback enhances learners' noticing of linguistic errors and supports long-term acquisition. Unlike conventional feedback, which often suffers from delays, AI-based feedback provides instant, detailed, and iterative corrections, enabling learners to engage more actively in self-editing.

Nevertheless, the results also underscore the continuing importance of teacher feedback. Although the control group achieved a lower gain score (8.2 points), students still benefited from teacher comments that provided contextual explanations and pedagogical scaffolding. This is consistent with Hyland and Hyland's (2006) argument that teacher feedback not only delivers linguistic corrections but also offers affective support that motivates students. (Hyland & Hyland, 2006). Therefore, while AI feedback excels in efficiency and coverage, it cannot fully substitute the motivational and interpretive functions of teacher feedback.

**Implications**

Theoretically, this study contributes to the literature on technology-enhanced language learning (TELL). It extends the understanding of second language acquisition (SLA) by highlighting the role of corrective input in learner interaction (Ellis, 2016; Li, 2010; Schmidt, 1990). AI-based feedback can be positioned as a form of enhanced input that accelerates error detection and encourages learners to modify their output, consistent with interactionist and input enhancement frameworks (Loewen, 2020; Mackey & Gass, 2015; Schmidt, 1990). This is aligned with Schmidt's (1990) noticing hypothesis, which posits that awareness of errors is a prerequisite for language acquisition (Ellis, 2016; Robinson, 2003).

From a pedagogical perspective, integrating AI tools into EFL writing classes provides several practical benefits. First, AI can reduce teachers' workload by handling repetitive grammatical corrections, allowing them to focus on higher-order aspects such as argumentation, cohesion, and critical thinking (Bitchener & Ferris, 2012; Ferris, 2010; Hyland, 2019). Second, students gain greater autonomy in monitoring their language development, consistent with the principles of learner-centered instruction (Benson, 2013; Holec, H., 1981; Little, 2007). However, caution is warranted: over-reliance on AI may lead to superficial corrections without conceptual understanding, as highlighted in studies on the limitations of automated feedback. Thus, it is essential to train students to interpret and use AI feedback reflectively.

**Limitations and Future Research**

Despite its promising findings, this study has several limitations. The relatively small sample size (n = 60) restricts the generalizability of the results. In addition, the short intervention period (six weeks) may not adequately capture long-term retention of grammatical improvement. The study also focused on a single type of academic text; other genres such as argumentative essays or research reports may yield different outcomes. Future research should adopt longitudinal designs with larger sample sizes and incorporate diverse writing genres. Collecting qualitative data, such as students' perceptions and attitudes toward AI feedback, would also provide richer insights into its affective and motivational impacts.

## V. Conclusion

This study confirms that artificial intelligence (AI)-based feedback has a significant impact on improving the grammatical accuracy of EFL students. Students who received automated feedback from AI systems such as Grammarly or GPT-based grammar checkers achieved higher post-intervention scores than those who only received manual comments from

instructors. These findings demonstrate that AI is not merely a corrective aid but also an effective pedagogical strategy in academic writing instruction.

Theoretically, the results expand the existing body of literature in technology-enhanced language learning (TELL), particularly within the framework of second language acquisition (SLA), by emphasizing the importance of corrective input that is immediate, consistent, and iterative in strengthening students' linguistic awareness. Practically, this study provides implications for teachers and higher education institutions: integrating AI-based tools can reduce instructors' correction workload while simultaneously fostering student autonomy in the revision process.

Nevertheless, the study has limitations, including a relatively small sample size, a short intervention period, and a focus on a single genre of academic writing. Future research is therefore recommended to involve more diverse populations, extend the duration of interventions, and explore different writing genres. Further investigations should also examine affective dimensions, students' perceptions, and the effectiveness of hybrid feedback models that combine the strengths of AI with the pedagogical roles of teachers.

## VI. Acknowledgments

The author would like to express sincere gratitude to the students who participated in this study and to the instructors who supported the data collection process. Appreciation is also extended to the university for granting permission and providing facilities for the research, particularly the English Education Program for facilitating the classroom intervention.

The author is further indebted to colleagues who offered valuable feedback in the development of the instruments and data analysis, as well as to the technical team who assisted in operating the SPSS software. Finally, heartfelt thanks are given to family members and colleagues for their moral and academic support throughout the completion of this study.

## VII. References

Abu Qub'a, A., Abu Guba, M. N., & Fareh, S. (2024). Exploring the use of grammarly in assessing english academic writing. Heliyon, 10(15), e34893. https://doi.org/10.1016/j.heliyon.2024.e34893

Aljabri, S. (2024). Timing of feedback and retrieval practice: a laboratory study with efl students. Humanities and social sciences communications, 11(1), 1–10. https://doi.org/10.1057/s41599-024-03983-6

Benson, P. (2013). Teaching and researching: autonomy in language learning. Routledge. https://doi.org/10.4324/9781315833767

Bitchener, J., & Ferris, D. R. (2012). Written corrective feedback in second language acquisition and writing. Routledge. https://doi.org/10.4324/9780203832400

Ellis, R. (2016). Focus on form: a critical review. Language teaching research, 20(3), 405–428. https://doi.org/10.1177/1362168816628627

Ferris, D. R. (2010). Second language writing research and written corrective feedback in sla: intersections and practical applications. Studies in second language acquisition, 32(2), 181–201. https://doi.org/10.1017/S0272263109990490

Holec, H. (1981). Autonomy and foreign language learning. Oxford/new york pergamon press.

Hyland, K. (2019). Second language writing. Cambridge university press. https://doi.org/10.1017/9781108635547

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

Language Learning & Technology. (2022). Volume 26 Number 1—Language Learning and Technology.

Li, S. (2010). The effectiveness of corrective feedback in sla: a meta-analysis. Language learning, 60(2), 309–365. https://doi.org/10.1111/j.1467-9922.2010.00561.x
Lin, S., & Crosthwaite, P. (2024). The grass is not always greener: teacher vs. Gpt-assisted written corrective feedback. System, 127, 103529. https://doi.org/10.1016/j.system.2024.103529
Lin, Y., & Yu, Z. (2025). Elucidating university students' intentions to seek automated writing feedback from grammarly: toward perceptual and systemic predictors. Humanities and social sciences communications, 12(1), 7. Https://doi.org/10.1057/s41599-024-03861-1
Little, D. (2007). Language learner autonomy: some fundamental considerations revisited. International journal of innovation in language learning and teaching, 1(1), 14–29. https://doi.org/10.2167/illt040.0
Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. (2024). Exploring the application of chatgpt in esl/efl education and related research issues: a systematic review of empirical studies. Smart learning environments, 11(1), 50. https://doi.org/10.1186/s40561-024-00342-5
Loewen, S. (2020). Introduction to instructed second language acquisition. Routledge. https://doi.org/10.4324/9781315616797
Mackey, A., & Gass, S. M. (2015). Second language research: methodology and design. Routledge.
OECD. (2025). What should teachers teach and students learn in a future of powerful ai? (20th ed., OECD education spotlights) [OECD education spotlights]. https://doi.org/10.1787/ca56c7d6-en
Polakova, P., & Ivenz, P. (2024). The impact of chatgpt feedback on the development of efl students' writing skills. Cogent education, 11(1), 2410101. https://doi.org/10.1080/2331186X.2024.2410101
Rahimi, M., Fathi, J., & Zou, D. (2025). Exploring the impact of automated written corrective feedback on the academic writing skills of efl learners: an activity theory perspective. Education and information technologies, 30(3), 2691–2735. https://doi.org/10.1007/s10639-024-12896-5
Ranalli, J. (2018). Automated written corrective feedback: how well can students make use of it? Computer assisted language learning, 31(7), 653–674. https://doi.org/10.1080/09588221.2018.1428994
Ranalli, J. (2022). Automated written corrective feedback: error-correction performance and timing of delivery. https://doi.org/10.64152/10125/73465
Ranalli, J., & Yamashita, T. (2022). Automated written corrective feedback: error-correction performance and timing of delivery. https://doi.org/10.64152/10125/73465
Robinson, P. (2003). Attention and memory during sla. The handbook of second language acquisition, 631–678. https://doi.org/10.1002/9780470756492.ch19
Schmidt, R. W. (1990). The role of consciousness in second language learning1. Applied linguistics, 11(2), 129–158. https://doi.org/10.1093/applin/11.2.129
Teng, M. F. (2024). "ChatGPT is the companion, not enemies": efl learners' perceptions and experiences in using chatgpt for feedback in writing. Computers and education: artificial intelligence, 7, 100270. https://doi.org/10.1016/j.caeai.2024.100270
Varsik, S., & Vosberg, L. (2024). The potential impact of artificial intelligence on equity and inclusion in education. OECD
Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of chinese efl learners: a randomized controlled trial. Frontiers in psychology, 14. https://doi.org/10.3389/fpsyg.2023.1249991

Yan, D., & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by chatgpt: a mixed-method multiple case study. Palgrave communications, 11(1), 1–14. https://doi.org/10.1057/s41599-024-03543-y

Zhai, N., & Ma, X. (2023). The effectiveness of automated writing evaluation on writing quality: a meta-analysis. Journal of educational computing research, 61(4), 875-900. https://doi.org/10.1177/07356331221127300