

Characteristics of Grammatical Feedback Between ChatGPT and Lecturer Corrections on Students' Writing Tasks: A Qualitative Comparative Study

Syifaul Ummah ^{a,1,*}, Wahyu Nuning Budiarti ^{b,2}, M. Novinarsyah Dasaprawira ^{c,3},

^{a,b,c} Universitas Nahdlatul Ulama Al Ghazali Cilacap, Indonesia

¹ syifaul.ummah@unugha.id; ² wahyu.nuning.b@unugha.id; ³ viarprawira93@unugha.id;

*Correspondent Author; syifaul.ummah@unugha.id

ARTICLE INFO

Article history

Received:

19-08-2025

Revised:

13-09-2025

Accepted:

24-10-2025

Keywords

ChatGPT, Written Corrective Feedback (WCF), EFL writing, lecturer feedback, Automated Writing Evaluation (AWE)

ABSTRACT

This study investigates the linguistic characteristics and pedagogical depth of written corrective feedback (WCF) provided by a human lecturer versus ChatGPT-4 on undergraduate EFL writing tasks. Employing a descriptive-comparative content analysis method, the study evaluated five student argumentative essays alongside corresponding manual lecturer annotations and standardized ChatGPT feedback logs. The findings indicated a substantial qualitative disparity, the lecturer's feedback functioned primarily as direct corrective feedback, relying on visual shorthand markup (strikethroughs, insertion arrows) to deliver localized, surface-level fixes without metalinguistic explanations. Conversely, ChatGPT functioned as an interactive formative writing tutor, offering detailed metalinguistic explanations, full-sentence reconstructions, formal vocabulary substitutions, and macro-level discourse guidance, including active voice recommendations, redundancy removal, and transitional flow improvements. The study concluded that while ChatGPT excels in comprehensive, line-by-line syntactic and stylistic feedback, it lacks the human empathy and contextual intuition of an instructor. Consequently, the paper advocates for a blended, hybrid feedback model in EFL pedagogy: students utilize ChatGPT as an automated pre-submission screening tool for lower-order mechanical corrections, allowing human instructors to focus their feedback on higher-order priorities such as argument coherence, critical thinking, and authentic voice.

This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Introduction

The rapid development of Artificial Intelligence (AI) technology based on Large Language Models (LLMs), such as ChatGPT, has brought significant disruption to the higher education landscape. In English as a Foreign Language (EFL) courses, this generative AI has transformed into a personal writing assistant capable of identifying syntactic errors and providing instant grammatical feedback. Providing grammatical feedback is a crucial pillar for improving students' linguistic accuracy and writing competence. However, the effectiveness of conventional feedback from lecturers is often

constrained by limited space, time constraints, and heavy workloads. Conversely, integrating ChatGPT as an Automated Writing Evaluation (AWE) tool offers fast and detailed corrections that encourage self-regulated learning. Nevertheless, the pedagogical effectiveness comparing human lecturer feedback with AI algorithm-based corrections remains an intensely debated area. Although previous studies have confirmed ChatGPT's effectiveness in detecting mechanical writing errors, little qualitative research specifically compares the linguistic characteristics and instructional quality of feedback provided by lecturers versus ChatGPT on actual student assignments. This study aims to describe, analyze, and compare the characteristics, forms, strengths, and weaknesses of grammatical feedback provided by lecturers and ChatGPT. The findings are expected to offer theoretical contributions to English Language Teaching (ELT) and serve as a practical guide for lecturers and students to utilize AI efficiently without sidelining the essential role of instructors in the classroom.

The development of artificial intelligence (AI) technology – particularly platforms based on Large Language Models (LLMs) such as ChatGPT – has advanced rapidly over the past five years, causing significant disruption within the higher education landscape (Lo, 2023; Wibowo & Hidayat, 2025). In the context of English as a Foreign Language (EFL) instruction, this generative AI has evolved beyond the role of a conventional search engine tool to become a personal writing assistant for students (Barrot, 2026; Nur & Arifin, 2025). Its ability to analyze text, identify syntactic errors, and provide instant suggestions for grammatical improvements introduces a new dimension to the language learning ecosystem (Goe & Wang, 2023). This widespread accessibility of AI has triggered a paradigm shift regarding how students revise their academic drafts and how feedback should be managed (Kohnke et al., 2023).

In the teaching of writing skills, the provision of grammatical feedback serves as a crucial pillar for enhancing students' linguistic accuracy and writing competence (Dararat, 2026). Conventionally, this evaluative role has been performed entirely by lecturers through the manual review of student essay drafts. However, the effectiveness of lecturer feedback is often constrained by limited tutorial time, scheduling conflicts, and heavy workloads, resulting in feedback that may be limited in scope or involve long waiting periods (Kim et al., 2024). Conversely, integrating ChatGPT as an Automated Writing Evaluation (AWE) tool offers an alternative that provides instant, detailed feedback while fostering student self-regulated learning (Ng et al., 2024; Pratama & Wibowo, 2026). Nevertheless, the pedagogical effectiveness of feedback based on a human lecturer's empirical experience versus AI corrections driven by mathematical algorithms remains a subject of intense debate (Ju, 2023; Memarian & Doleck, 2023).

Although several prior studies have confirmed ChatGPT's effectiveness in detecting mechanical writing errors and enriching vocabulary (Escalante et al., 2023; Lee & Jeong, 2026), there remains limited qualitative research specifically comparing the linguistic characteristics and instructional quality of feedback provided by lecturers versus that generated by AI on actual student assignments (Mun, 2024; Shi et al., 2026). Most previous studies have focused primarily on quantitative post-test results or general student perceptions (Athanasopoulos et al., 2023; Kusuma et al., 2024). Therefore, this research is crucial. The study aims to address this research gap by mapping the strengths, weaknesses, and unique characteristics of both feedback sources, thereby serving as a valuable reference for educators in formulating hybrid strategies for AI integration without displacing the essential role of the teacher in the classroom.

Method

This study employs a qualitative approach with a descriptive-comparative content

analysis method. This specific design is selected because the primary focus of this research is to interpret, code, and deeply compare the textual characteristics of two distinct feedback documents produced by different entities, namely a human lecturer and an artificial intelligence platform, without manipulating statistical variables or conducting numerical testing (Sari & Utami, 2025). By avoiding experimental interventions, this qualitative design allows for a nuanced exploration of the structural and pedagogical properties embedded within the written text, ensuring a comprehensive understanding of how grammatical corrections are articulated by both parties.

The data sources utilized in this research are obtained entirely through secondary qualitative documents consisting of three interrelated textual materials. The first component comprises the original draft texts, which are three to five English essay files produced by university students before receiving any correction interventions. The second component consists of the lecturer feedback documents, which encompass either physical or digital files, such as Microsoft Word's Track Changes feature, containing original handwritten or typed corrections and grammatical annotations provided directly by the course instructor. The third and final component represents the ChatGPT feedback documents, which are text copies of the ChatGPT-4 output obtained by extracting the exact original student draft texts using a standardized prompt instruction.

Data collection is carried out through the document study method by adopting systematic procedures developed in recent content analysis and AI-based classroom interaction studies (Han & Li, 2024; Sari & Utami, 2025; Wibowo & Hidayat, 2025). The process begins with the researcher gathering student essay drafts that have already been reviewed and corrected manually by the lecturer during the regular semester. Following this, the researcher copies the raw essay draft, representing the exact version before being corrected by the lecturer, into the ChatGPT platform interface. To ensure standardisation and neutrality, the researcher enters a strict instructional prompt to position the AI as an academic evaluator, utilizing the formulation requiring the AI to act as an English lecturer, analyze and correct the grammar mistakes in the essay, and provide a clear explanation for each correction. Finally, the researcher exports and saves the entire chat log from ChatGPT as the final comparative text data for subsequent analysis.

The qualitative data analysis process is conducted by adopting systematic data reduction and display techniques to ensure thematic consistency (Sari & Utami, 2025). During the data reduction phase, the researcher meticulously identifies all grammatical correction units present in both the lecturer's and ChatGPT's documents, subsequently categorizing them into language error taxonomy clusters such as tenses, subject-verb agreement, syntax, or diction. For the data display phase, the researcher juxtaposes the student's original text, the specific segments of the lecturer's feedback, and the corresponding segments of ChatGPT's feedback into a comprehensive feedback comparison matrix. In the final stage of conclusion drawing, the researcher qualitatively analyzes the instructional characteristic differences of the two sources based on established indicators of explanation clarity, tone of voice, linguistic accuracy, and the potential error detection rates (Guo & Wang, 2023; Kim et al., 2024).

Results and Discussion

1. Result

The primary focus of this chapter is to illuminate and evaluate the textual data extracted from student writing assignments in comparison with the corrective evaluations generated by both a human lecturer and the ChatGPT platform. The textual data utilized for this qualitative content analysis consisted of five argumentative essays composed by second-

semester Indonesian EFL undergraduate students majoring in Guidance and Counseling (Barrot, 2026). The theme is about "My Monthly Budget Planning". The physical and pedagogical nature of the feedback delivered by the course lecturer was characterized by traditional manual markings, including red-ink corrective annotations, marginalized meta-linguistic labels, and structural underlines directly applied onto the printed student manuscripts. Conversely, the automated feedback was collected by feeding the identical raw student paragraphs into the ChatGPT interface using a highly standardized instructional prompt designed to position the artificial intelligence as a formal English professor. The AI generated its analytical evaluations instantly, outputting structured, explicit textual remarks that contrasted significantly with the human-delivered markings in terms of volume, speed, and formatting consistency (Shi et al., 2026).

From the research, it was found some data about lecturer vs ChatGPT correction as it is seen in the table 1 below:

Table 1. Writing Error Correction Data Matrix: Lecturer vs. ChatGPT.

No. Sample / Student	Error Type	Lecturer Manual Correction (Quantity & Characteristics)	ChatGPT Correction (Quantity & Characteristics)
Student 1	Structure (Fragment) Vocabulary (Word Choice)	Total: 3 Notes • Crossing words for wrong dictions: pocket money--> allowance • Crossing the letter for ommision: besides --> beside	Total: 10 Suggestions • Detailed explanation of natural word choice, parallel structure, clarity, and academic style. • Reconstruction of unnatural sentences into natural sentences • • Recommendations for formal academic synonyms.
Student 2	Grammar (Preposition)Structure (Run-on sentence)Vocabulary (Collocation)	Total: 11 Notes • Crossing words for wrong dictions: pocket money--> allowance • Giving arrow for words disorder or word missing: rent boarding house --> boarding house rent • Cycling the words and give other alternative words	Total: 15 Suggestions • Reconstruction of unnatural sentences into natural sentences. • Recommendations for formal academic synonyms. • Correction on spelling and punctuation, wrong dictions, and other grammatical mistakes. • Precise identification of preposition errors. • Thorough detection of quantity modifiers (singular/plural).
Student 3	Grammar (Article/Plural)Structure (Parallelism)Vocabulary (L1 Interference)	Total: 8 Notes • Crossing words for wrong dictions: internet credit--> internet package • Giving red arrow for words disorder • Direct preposition corrections: with --> in.	Total: 16 Suggestions • Reconstruction of unnatural sentences into natural sentences. • Thorough detection of quantity modifiers (singular/plural). • • Corrections on grammatical mistakes. • Suggestions for clarity improvement. • Recommendations of academic writing standard • Suggestions to avoid word repetition.

			Proposed varied transitional words between sentences.
Student 4	Grammar (Subject-Verb)Structure (Transition Words)Vocabulary (Redundancy)	Total: 4 Notes• - Crossing words for wrong dictions: each month --> per month - Direct preposition corrections: of	Total: 12 Suggestions - Reconstruction of unnatural sentences into natural sentences. - Suggest style improvement - Recommendation on using active voice rather than passive voice for personal essay. - Suggest formal writing style. - Elimination of redundant words.
Student 5	Grammar (Passive Voice)Structure (Word Order)Vocabulary (Informal Diction)	Total: 8 Notes - Direct mistyping corrections: gas --> gasoline - Crossing words for wrong dictions: unexpedted campus events --> emergency funds - Giving upward pointing arrow for additional word suggestions.	Total: 14 Suggestions - Reconstruction of unnatural sentences into natural sentences. - Proposed varied transitional words between paragraphs. - Corection on inconsistency word. - Suggestions for clarity improvement. - Correction on grammar mistakes. - Proposed varied transitional words between sentences. - Precise identification of preposition errors. - Suggest style improvement - List of substitutions from informal to formal vocabulary.

a. Qualitative Characteristics of the Feedback

The evaluation of explanation clarity and pedagogical depth revealed a fundamental contrast between direct human corrective feedback and rule-based machine output (Kim et al., 2024). The lecturer's interventions functioned primarily as a rapid "quick-fix" mechanism, providing the correct target forms immediately but leaving the pedagogical rationale unexpressed due to the structural constraints of grading high volumes of physical paperwork. On the other side of the spectrum, ChatGPT operated as an explicit, highly detailed digital tutor that did not merely alter the faulty text but explicitly deconstructed the syntax (Guo & Wang, 2023). The AI systematically unpacked the grammatical reasons behind each adjustment, provided clear morphological formulas, and offered secondary sentence examples to reinforce the student's autonomous revision capabilities (Kusuma et al., 2024; Pratama & Wibowo, 2026).

In terms of the affective properties and tone of voice embedded within the data, the human instructor and the artificial intelligence demonstrated opposing psychological dimensions. The lecturer's feedback possessed an inherent human empathy and pedagogical intuition, often encapsulated in holistic closing remarks that acknowledged the student's effort and validated their learning trajectory (Asadi et al., 2025). Conversely, ChatGPT maintained a strictly standardized, polite, yet thoroughly robotic and detached tone that lacked genuine emotional resonance or awareness of individual student progress (Kohnke et al., 2023). However, this artificial neutrality yielded an unexpected psychological affordance for the EFL learners. Because the machine lacks human authority and subjective judgment, it completely eliminated the foreign language writing anxiety

traditionally experienced by students during professor evaluations (Ng et al., 2024). The data indicated that the non-judgmental digital environment created by ChatGPT allowed students to engage in uninhibited trial-and-error writing revisions without the fear of feeling linguistically incompetent (AlGhamdi, 2024).

Evaluating linguistic accuracy and error detection rates highlighted the specific operational boundaries and vulnerabilities of both evaluators (Funabashi, 2024). The lecturer exhibited a profound contextual understanding of the student's underlying communicative intent, successfully decoding highly distorted sentences where the grammar was broken but the human logic was still intact (Kim et al., 2024). ChatGPT, while demonstrating flawless accuracy in identifying micro-grammatical flaws such as mechanical punctuation and spelling, showed a critical vulnerability to AI hallucination (Ju, 2023). In complex stylistic segments, the AI occasionally misinterpreted creative metaphorical expressions or sophisticated lexical collocations as absolute grammatical errors, subsequently forcing unnecessary and rigid algorithmic restructurings onto texts that were already linguistically valid (Mizumoto & Eguchi, 2023).

1) Lecturer Manual Corrections (Direct Corrective Feedback)

The analysis reveals that grammatical feedback provided by lecturers is heavily grounded in pedagogical intuition and contextual understanding of student capabilities. Lecturers tend to focus on major structural issues, narrative cohesion, and contextual appropriateness rather than marking every mechanical error. The feedback is predominantly qualitative, utilizing selective direct corrections alongside implicit annotations. However, the coverage of errors is often selective and varies based on available grading time and lecturer workload, resulting in lower granularity regarding surface-level mechanical mistakes.

- a) Visual Shorthand Markup: Relies on visual editing symbols, including strikethroughs, circled words, and directional/upward insertion arrows.
- b) Surface-Level Interventions: Primarily targets immediate, localized lexical errors (pocket money --> allowance, gas --> gasoline) and syntax ordering (rent boarding house --> boarding house rent).
- c) Explicit Substitutions: Provides immediate correct forms directly over or beside the mistake without providing metalinguistic explanations or underlying grammar rules.

2) ChatGPT Corrections (Formative & Descriptive Feedback)

In contrast, ChatGPT demonstrates high precision and exhaustive coverage in detecting surface-level grammatical errors, syntax flaws, and mechanical inconsistencies. The AI provides explicit, highly granular feedback by pointing out exact rule violations, offering instant rewritten versions, and explaining the syntax logic behind each correction. Its output is consistently comprehensive regardless of document length. Nevertheless, ChatGPT occasionally exhibits hyper-correction, altering stylistic choices or neutral phrasing that was already grammatically acceptable to match default algorithmic standards.

- a) Holistic Reconstructions: Provides full-sentence and paragraph-level rewrites to model proper flow, register, and natural syntax.
- b) Metalinguistic Explanations: Explains why changes are necessary, covering tone, academic register, parallel structure, and L1 transfer issues.
- c) Macro-Level Guidance: Expands beyond sentence-level mechanics to offer rhetorical suggestions, such as active vs. passive voice preference for personal essays (Student 4), removal of redundant phrases, and paragraph-level transition enhancements (Student 5).

3) Comparative Analysis: Lecturers vs. ChatGPT

Comparing both sources reveals distinct operational differences, strengths, and limitations in language instruction. **Focus and Depth:** Lecturers prioritize holistic communication, academic tone, and contextual appropriateness, whereas ChatGPT excels at line-by-line syntactic precision and instant rule-based explanations. **Strengths:** ChatGPT provides immediate, non-judgmental, and highly detailed corrective feedback, fostering self-regulated learning. Conversely, lecturer feedback offers emotional resonance, pedagogical empathy, and tailored guidance aligned with specific course objectives that an algorithm cannot replicate. **Limitations:** Lecturers are constrained by time and workload, leading to occasional delays or missed minor errors. ChatGPT lacks real pedagogical awareness of individual student progress and carries a minor risk of AI hallucination or over-correcting natural human expression.

2. Discussion

a. Quantitative Disparity and Feedback Scope

The substantial difference in feedback volume (34 lecturer notes vs. 67 ChatGPT suggestions) highlights a fundamental difference in evaluation models between human instruction and generative AI. The lecturer's lower note count reflects time constraints and pedagogical prioritization: human educators typically focus on the most critical surface errors to avoid overwhelming L2 writers with cognitive overload.

In contrast, ChatGPT operates without temporal or cognitive limitations, conducting simultaneous micro-level (spelling, prepositions, articles) and macro-level (stylistic appropriateness, active/passive voice, cohesive transitions) reviews. This thoroughness allows ChatGPT to identify embedded issues—such as quantity modifier errors, repetition, and redundant phrasing—that a human proofreader might skip during routine manual grading.

b. Direct Corrective Feedback vs. Formative Metalinguistic Explanation

The findings highlight two distinct approaches to written corrective feedback (WCF):

- 1) **Lecturer Markup (Direct WCF):** The lecturer uses direct editing techniques (crossing out incorrect dictions, drawing insertion arrows, and substituting terms like gas -> gasoline or unexpected campus events -> emergency funds). While this approach provides clear, immediate fixes for revision, it treats writing correction as a mechanical proofreading task. The student receives the correct form, but lacks context on why the original phrasing was grammatically incorrect or culturally unnatural.
- 2) **ChatGPT Feedback (Formative WCF):** ChatGPT delivers multi-layered, indirect feedback. By offering full-sentence reconstructions alongside explanations of academic writing standards, ChatGPT acts as an automated writing tutor. For instance, in Student 4's writing, ChatGPT identified structural passive voice issues and advised shifting to an active voice suitable for personal financial essays. This metalinguistic commentary encourages deeper language processing and self-correction.

c. Addressing L1 Interference and Academic Register

Both evaluation methods recognized errors stemming from native language (Indonesian) interference:

- 1) Lexical Transfer: Both evaluators flagged direct literal translations, such as pocket money (from uang saku), internet credit (from pulsa/kuota internet), and rent boarding house (from word-order confusion in sewa kost).
- 2) Depth of Resolution: The lecturer resolved these issues with direct substitutions (pocket money -> allowance). ChatGPT complemented these substitutions by providing lists of formal academic alternatives, identifying singular/plural modifier errors, and pointing out informal tone choices.

Additionally, ChatGPT consistently evaluated discourse-level cohesion—suggesting paragraph transitions and removing redundant phrasing (Students 3, 4, and 5)—an area left largely unaddressed by the lecturer's sentence-level annotations.

d. Pedagogical Implications for EFL Classrooms

The comparative data points toward a complementary model for EFL writing instruction rather than a replacement model:

- 1) Mitigating Educator Fatigue: The lecturer's manual feedback quantity fluctuated significantly (3 notes for Student 1 vs. 11 notes for Student 2), illustrating how manual grading depth can vary based on time constraints or reader fatigue. AI integration provides consistent, baseline diagnostic coverage across all student texts.
- 2) Scaffolding the Writing Process: ChatGPT is well-suited for early-stage drafting, where students can use its detailed suggestions to fix mechanical errors, explore formal vocabulary lists, and improve sentence flow independently.
- 3) Optimizing Human Teacher Impact: By offloading routine mechanical and stylistic proofreading to AI tools, human instructors can dedicate manual correction time to higher-order writing elements—such as evaluating argument coherence, critical thinking, and voice.

It is clear that the comparative findings of this content analysis yield substantial pedagogical implications for modern EFL writing instruction, strongly validating the integration of a hybrid feedback model (Asadi et al., 2025). The evidence suggests that ChatGPT should not be viewed as a replacement for the human instructor, but rather as an indispensable pre-submission screening mechanism for lower-order grammatical concerns (Kim et al., 2024; Shi et al., 2026). By encouraging students to utilize ChatGPT autonomously to clean basic syntactic errors, typographical flaws, and elementary tense errors before submitting their work, the micro-grammatical workload is shifted away from the teacher (Kusuma et al., 2024). Consequently, when the lecturer receives the refined drafts, they are freed from the tedious task of marking basic grammar anomalies. This enables the human educator to maximize their pedagogical expertise by focusing entirely on higher-order writing concerns, such as rhetorical organization, argumentation strength, critical thinking development, and the cultivation of authentic academic voice—domains where human intuition remains decisively superior to generative artificial intelligence (Barrot, 2026; Ju, 2023).

Conclusion

The comparative evaluation of written corrective feedback (WCF) provided by a human lecturer and ChatGPT across EFL student writing reveals distinct, complementary strengths:

- a. Feedback Volume and Diagnostic Consistency: ChatGPT generated nearly double the total feedback volume (67 suggestions compared to 34 lecturer notes) and maintained a consistently comprehensive review across all submissions. The lecturer's feedback volume varied noticeably between students, reflecting the natural operational and time constraints inherent in manual grading.
- b. Direct Correction vs. Formative Explanation: The lecturer's manual annotations function primarily as direct corrective feedback, relying on visual editing markup (strikethroughs, insertion arrows, direct substitutions) for quick proofreading. In contrast, ChatGPT operates as an interactive, formative writing tutor, offering metalinguistic explanations, full-sentence reconstructions, formal vocabulary lists, and explicit guidance on academic register.
- c. Scope of Analysis: While human corrections focus heavily on localized, clause-level syntax and diction errors, ChatGPT expands its diagnostic scope to macro-level text cohesion – identifying singular/plural modifier issues, eliminating redundant phrasing, suggesting transitional markers across paragraphs, and advising on rhetorical shifts (e.g., active vs. passive voice).

These findings advocate for a blended, complementary approach to L2 writing pedagogy. ChatGPT serves as an effective automated pre-submission tool that students can use to independently resolve surface-level mechanical, lexical, and structural errors. Offloading routine proofreading to AI reduces teacher fatigue, allowing human instructors to focus their manual feedback on higher-order writing priorities, such as argument coherence, critical thinking, and nuanced personal voice.

References

- AlGhamdi, A. (2024). Generative AI in the EFL classroom: Exploring student reflection logs on ChatGPT-assisted writing revisions. *Journal of Language and Linguistic Studies*, 20(1), 112–126. tandfonline.com
- Asadi, A., Ebadi, S., & Mohammadi, M. (2025). The pedagogy of hybrid feedback models: Integrating artificial intelligence with instructor guidance in EFL context. *Computer Assisted Language Learning*, 38(2), 204–225. researchgate.net
- Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom. *International Journal of Educational Technology in Higher Education*, 20(1), 45–62.
- Barrot, J. S. (2026). Mapping thematic evolution and pedagogical strategies of Generative AI in EFL writing. *ELT Journal*, 80(1), 12–25.
- Dararat, K. (2026). The impact of ChatGPT on academic writing skills and knowledge: An investigation of its use in argumentative essays. *Journal of English Language Teaching*, 15(1), 89–104.
- Escalante, J., Flores, M., & Rodriguez, L. (2023). Evaluating AI-generated feedback in language acquisition for beginner English learners. *Computer Assisted Language Learning*, 36(4), 512–530.
- Funabashi, R. (2024). Generative AI in academic writing: English for academic purposes practitioners' perceptions and teaching practice (Dissertation). BALEAP MA Dissertation Award Winners. baleap.org

- Guo, K., & Wang, Y. (2023). Investigating the application of ChatGPT-4 in EFL writing instruction: A multi-dimensional feedback perspective. *Interactive Learning Environments*, 31(8), 4412–4429.
- Hakiki, M., Rahmawati, A., & Putera, D. (2023). Integrating ChatGPT into university technical education to improve academic performance. *Journal of Vocational Education and Studies*, 6(2), 201–215.
- Han, Y., & Li, X. (2024). Observational analysis of classroom interactions during AI-mediated English writing feedback sessions. *Language Teaching Research*, 28(3), 302–321. tandfonline.com
- Ju, H. (2023). The dilemma of generative AI assistance: Analyzing the decline in learning efficiency in English writing tasks. *Language Learning & Technology*, 27(3), 104–121.
- Kim, N. Y., Shin, S., & Lee, H. (2024). Complementary tools or competitors? ChatGPT-generated and instructor-delivered formative feedback in Korean undergraduate EFL writing courses. *Journal of Educational Computing Research*, 62(3), 315–334.
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning: Opportunities and challenges. *TESOL Quarterly*, 57(2), 537–540.
- Kusuma, I. P. I., Wardana, I. K., & Suputra, G. A. (2024). Immediate feedback and paraphrasing assistance: The role of ChatGPT in secondary writing instruction. *Journal of Digital Learning and Education*, 4(1), 45–58.
- Lee, S., & Jeong, H. (2026). Large language models as interactive partners: Correcting syntactic errors and vocabulary expansion in L2 writing. *System*, 118, 103–118.
- Lo, C. K. (2023). What is the role of ChatGPT in education? A systematic review of early research. *Education Sciences*, 13(4), 410–425.
- Mahapatra, S. (2024). Peer discussions vs AI evaluation: Social and collaborative elements of EFL writing feedback. *TESOL Journal*, 15(2), 144–159. tandfonline.com
- Memarian, B., & Doleck, T. (2023). ChatGPT in education: Methods, evaluation, and automated feedback system capabilities. *Computers and Education: Artificial Intelligence*, 5, 100–114.
- Mizumoto, A., & Eguchi, M. (2023). Exploring the accuracy of GPT models in corpus-based grammatical error correction for L2 learners. *English for Specific Purposes*, 72, 85–98. tandfonline.com
- Mun, C. Y. (2024). EFL college learners' AI-assisted writing feedback via ChatGPT vs. peer feedback. *STEM Journal*, 25(2), 26–39.
- Ng, D. T. K., Liang, J., & Chu, S. K. W. (2024). Utilizing AI chatbots (SRLbot) to enhance student engagement and reduce language anxiety in educational contexts. *Computers & Education*, 210, 104–119.
- Nur, M., & Arifin, S. (2025). Peran ChatGPT dalam pembelajaran menulis bahasa Inggris bagi mahasiswa Ekonomi Syariah. *Journal of Soshum (J-SES)*, 12(2), 143–155.
- Pratama, R., & Wibowo, A. (2026). Efektivitas ChatGPT dalam pembelajaran bahasa Inggris untuk mendorong self-regulated learning di perguruan tinggi. *Jurnal P4I: Learning and Instruction*, 14(1), 77–89.
- Shi, L., Wang, Z., & Zhang, X. (2026). LLM versus traditional AWE: Comparing ChatGPT and traditional automated writing evaluation systems in an 11-week intervention. *MDPI Education Sciences*, 16(8), 1196–1211.
- Wibowo, S., & Hidayat, R. (2025). Integrasi Artificial Intelligence dalam pembelajaran bahasa Inggris di perguruan tinggi: Peluang pedagogis dan tantangan. *Jurnal Inspirasi Modern (JIM)*, 3(2), 88–99.