

Beyond Accuracy: How EFL Students Navigate GPT-Generated Feedback in Argumentative Writing across Proficiency Levels

Moh. Shofi Zuhri¹, Moh. Kavin Lidinillah²

*English Department, Faculty of Tarbiyah and Teacher Training, Universitas Tarbiyatut Tholabah
Lamongan, Indonesia*

Article Info

Keywords:

Keyword 1; Automated
Grammatical Error
Correction
Keyword 2; GPT
Keyword 3; EFL writing

Article History:

Received : 31/04/2026
Revised : 31/05/2026
Accepted : 31/06/2026
Available Online:
31/07/2026

Abstract

Essay-writing proficiency is an important indicator of English language competence among English as a Foreign Language (EFL) learners. However, providing written corrective feedback manually can place considerable demands on instructors' time and workload, particularly in large classes. This study evaluates the performance of a Generative Pre-trained Transformer (GPT) model in detecting and correcting grammatical and lexical errors in EFL students' argumentative essays, while also exploring how students perceive and respond to the feedback it provides. An exploratory descriptive qualitative design was employed with 15 English Language Education students from three semester levels (1, 3, and 5). Essay drafts and semi-structured interview transcripts were analyzed using the Error Analysis framework (Corder, 1981; Ellis, 2008) and Thematic Analysis (Braun & Clarke, 2006). The findings show that GPT achieved its highest correction accuracy in Semester 1 students' essays (78%), whereas the rate of over-correction increased to 20% in Semester 5 as sentence structures became more complex. The qualitative findings reveal a clear difference in students' metacognitive responses. Semester 1 students tended to exhibit automation bias by accepting GPT-generated corrections with little critical evaluation, while Semester 3 and 5 students demonstrated greater learner agency by evaluating, filtering, and negotiating the suggested corrections. Overall, the findings suggest that GPT can serve as a useful form of scaffolding for formative feedback, but its integration into EFL writing pedagogy should be accompanied by critical AI literacy to help students maintain their metalinguistic awareness and authorial voice.

Introduction

The rapid adoption of generative artificial intelligence in educational settings has brought an important question to the fore: can students safely trust the language corrections produced by AI? The Pew Research Center reported that the proportion of U.S. teenagers

using ChatGPT for schoolwork doubled within a single year, rising from 13% in 2023 to 26% in 2024 (Sidoti, Park, & Gottfried, 2025). Yet the growing presence of generative AI in classrooms does not necessarily mean that every correction produced by these systems is accurate or safe to adopt without scrutiny. Interestingly, the present study reveals a pattern that runs counter to common expectations: the more proficient the EFL writer, the greater the possibility that their writing may be unnecessarily “corrected” by AI. In the essays produced by the most proficient participants (Semester 5), one in five GPT-generated corrections (20%) constituted over-correction, altering structures or word choices that were already grammatically acceptable while potentially weakening the writer’s authentic voice. This finding suggests that the accuracy of automated feedback is not uniform across proficiency levels and that uncritical trust in AI-generated output, however appealing it may be, carries genuine pedagogical risks.

In the context of English as a Foreign Language (EFL) learning, essay-writing proficiency is widely regarded as an important indicator of advanced language competence. Effective academic writing requires not only an understanding of coherent discourse structures but also grammatical accuracy and appropriate lexical choices (Hyland, 2003; Ferris, 2002). Nevertheless, EFL learners often continue to struggle with recurring grammatical errors and inappropriate vocabulary use, partly as a result of first-language (L1) transfer and an incompletely developed interlanguage system (Rozovskaya & Roth, 2014). At the same time, providing written corrective feedback manually places substantial demands on instructors’ time and cognitive workload, particularly in large classes. As a result, feedback may be delayed, inconsistent, or less effective in supporting students’ independent revision processes (Ferris, 2002).

In response to these limitations, the fields of Computer-Assisted Language Learning (CALL) and Natural Language Processing (NLP) have increasingly moved toward the development of Automated Grammatical Error Correction (GEC) systems. Advances in deep-learning-based artificial intelligence, particularly the Transformer architecture introduced by Vaswani et al. (2017), have significantly reshaped the way textual information can be processed. Through self-attention mechanisms and parallel sequence processing, Transformer-based models can capture long-range contextual dependencies with considerable efficiency. These capabilities allow them to process sentence-level context more deeply than earlier rule-based and statistical systems, including n-gram models (Bryant et al., 2019; Zhang et al., 2020).

Although Transformer-based models such as BERT, T5, and BART have demonstrated substantial improvements on major GEC benchmarks, including CoNLL-2014 and BEA-2019, much of the existing research has focused on native English data or L2 datasets characterized by particular linguistic patterns (Bryant et al., 2019). A research gap therefore remains concerning how effectively Transformer-based models respond to the specific error patterns found in EFL learners' writing, particularly complex lexico-grammatical errors shaped by local-language interference. Existing GEC systems tend to perform reliably when dealing with typographical errors and basic syntactic problems, but they may still produce over-corrections or struggle with contextual interpretation when addressing lexical choices and collocations that are highly dependent on the EFL context (Rozovskaya & Roth, 2014).

The novelty of the present study lies in bringing together three dimensions that have rarely been examined simultaneously in Transformer-based GEC research. First, rather than evaluating a model at a single proficiency point, this study deliberately compares GPT performance across argumentative essays produced by EFL students at three stages of interlanguage development—emerging, intermediate, and upper-intermediate. This approach makes it possible to examine patterns of accuracy and over-correction in relation to the linguistic complexity of the writers, rather than reducing model performance to a single accuracy score. Second, the study goes beyond a purely technical evaluation of GPT output by connecting model performance with students' metacognitive responses. In doing so, it explores the contrast between automation bias and learner agency, two dimensions that have seldom been examined alongside evidence of a model's corrective performance. Third, by evaluating the freely available version of GPT, the study reflects a more realistic learning context for many EFL students in developing countries, rather than relying on idealized conditions involving premium AI models that often dominate international GEC benchmarks (Bryant et al., 2019; Zhang et al., 2020).

The integration of Transformer-based models as automated feedback tools may offer meaningful benefits for EFL writing pedagogy. From the learner's perspective, real-time contextual feedback can support metalinguistic awareness and shorten the cycle between writing and revision (Nagata & Nakatani, 2010). From the instructor's perspective, such systems can reduce the administrative burden associated with language correction, allowing teachers to devote greater attention to the substance of students' arguments, their ideas, and the rhetorical organization of their essays. At the same time, effective integration requires a

careful examination of what these systems can and cannot do, how accurately they perform across different types of errors, and which forms of linguistic difficulty remain particularly challenging for AI-generated feedback.

Against this background, and in response to the identified gap in the literature, this study explores and evaluates the use of a Generative Pre-trained Transformer (GPT) model to detect and correct grammatical and lexical errors in EFL students' argumentative essays. More specifically, the study addresses three main questions

1. How accurate and contextually appropriate are the grammatical and lexical corrections generated by GPT for EFL students' argumentative essays across three proficiency levels (Semesters 1, 3, and 5)?
2. What types of grammatical and lexical errors are most effectively detected by GPT, and which types are more likely to result in over-correction or miscorrection?
3. How do EFL students perceive, respond to, and experience GPT-generated automated feedback during the process of independently revising their argumentative essays?

Theoretically, this study contributes to the growing body of CALL and NLP research by offering a closer understanding of how Transformer-based models interact with the linguistic characteristics of EFL student writing. Practically, the findings may inform AI-assisted language education and EFL writing pedagogy by highlighting the need for writing-support tools that are not only accurate and adaptive but also sensitive to learners' linguistic development and individual authorial choices.

Method

This study employed an exploratory qualitative design to examine GPT-generated feedback on EFL students' argumentative writing and explore how students responded to the feedback during revision. The participants were 15 English Language Education students selected through purposive sampling and divided equally across three semester levels: Semester 1 (emerging), Semester 3 (intermediate), and Semester 5 (upper-intermediate). All participants provided informed consent, and their identities were protected through participant codes.

The data consisted of 15 argumentative essays of approximately 300–500 words and semi-structured interview transcripts. All participants wrote on the same topic without using translation tools or automated grammar checkers to ensure that the essays represented their initial writing abilities.

Each essay was processed using a standardized GPT prompt designed to identify and correct grammatical and lexical errors, including syntactic, morphological, lexical, and collocational problems, without changing the students’ intended ideas. After receiving the AI-generated feedback, participants independently revised their essays and subsequently took part in 20–30-minute semi-structured interviews concerning the clarity, accuracy, usefulness, and limitations of the feedback.

GPT performance was analyzed using the Error Analysis framework (Corder, 1981; Ellis, 2008) and classified into four categories: Correct Detection & Correction, Over-correction, Mis correction, and Missed Error. Interview data were analyzed thematically following Braun and Clarke (2006), with particular attention to students’ perceptions of usability, cognitive load, trust in AI feedback, and their responses during revision.

To enhance credibility, GPT outputs were independently coded by the researcher and a language expert, with disagreements resolved through discussion and consensus. Data-source triangulation and peer debriefing were also used to strengthen the credibility of the findings. The study received ethical approval from the relevant institution, and all participants provided informed consent

Findings and Discussion

1. GPT Performance in Detecting and Correcting Errors in Student Essays

Based on the analysis of 15 EFL student essays across three semester levels, the freely available version of GPT demonstrated a generally responsive performance in identifying and correcting language errors. The GPT-generated corrections were classified into four main categories: Correct Detection & Correction, Over-correction, Mis correction, and Missed Error, as presented in Table 1.

Table 1. Distribution of GPT Correction Performance Categories across Student Essays.

GPT Correction Category	Operational Description	Semester 1	Semester 3	Semester 5	Dominant Characteristics of AI Output
Correct Detection & Correction	Successfully identifies an error and provides a grammatically and contextually appropriate correction.	78%	72%	65%	Highly accurate in subject–verb agreement, tense, and pluralization.

GPT Correction Category	Operational Description	Semester 1	Semester 3	Semester 5	Dominant Characteristics of AI Output
Over-correction	Changes a structure or word choice that is already grammatically acceptable.	5%	12%	20%	A tendency toward stylistic formalization and the erosion of authorial voice.
Miscorrection	Identifies an error but provides a contextually inappropriate correction.	10%	6%	5%	Difficulty interpreting meaning in sentences strongly affected by L1 transfer.
Missed Error	Fails to identify an existing grammatical or lexical error in the original text.	7%	10%	10%	Subtle collocational errors and cross-paragraph cohesion problems may go undetected.

The data in Table 1 reveal an inverse pattern between Correct Detection & Correction and Over-correction as students move to higher semester levels. In the Semester 1 essays, GPT achieved its highest correction accuracy, at 78%. This pattern appears to be associated with the predominance of rule-based errors, such as noun-phrase agreement problems and relatively simple sentence structures.

By contrast, the proportion of over-correction increased substantially in the Semester 5 essays, reaching 20%. This occurred as the more advanced students began to employ complex-compound sentences and more varied rhetorical structures. GPT, which is probabilistically trained to optimize standard fluency and readability, tended to simplify or revise these complex sentences into more stereotypical forms, even when the students’ original structures did not actually violate grammatical conventions.

2. Linguistic Error Typology and the Adaptability of GPT Corrections across Semester Levels

The analysis also revealed differences in the linguistic error patterns produced by students across three stages of interlanguage development (Selinker, 1972).

a. Semester 1 Students: Emerging Level

Errors in this group were largely characterized by literal L1 transfer and limited understanding of basic grammatical structures. Common examples included omission of the copula or *to be* (“They zero knowing about this topic”), subject–verb agreement errors (“Every students need to learn”), and noun–phrase constructions influenced by Indonesian word order (“Era modern this day”).

GPT was highly responsive to these types of errors. The model was able not only to identify the location of the errors accurately but also to restructure the sentences into more natural forms, such as “In this modern era, every student needs to learn...”.

b. Semester 3 Students: Intermediate Level

At this level, the errors shifted from basic morphosyntactic problems toward clause coordination and lexical selection. Frequently observed problems included redundant transitional conjunctions (“Although technology is good, but it has bad sides”) and inappropriate verb–preposition collocations (“Discuss about the problem”).

GPT generally recognized these forms of redundancy effectively. However, in some cases involving local lexical interference, such as the use of *join* to convey the meaning of participating in an activity (“I joined the event yesterday,” when *attended* was intended), GPT occasionally failed to identify the problem. This represents a Missed Error, as the sentence was syntactically acceptable even though the intended lexical meaning differed from the context.

c. Semester 5 Students: Upper-Intermediate Level

The Semester 5 essays demonstrated relatively mature grammatical control but presented greater challenges in discourse cohesion and semantic nuance. The errors identified at this level were generally related to parallel structure and adherence to academic register.

At this stage, the limitations of GPT became more apparent. When students adopted a more exploratory writing style, GPT frequently produced over-corrections

by replacing nuanced lexical choices with more generic vocabulary. Such revisions could weaken the emphasis of the writers' original arguments.

3. Metacognitive Responses and Pedagogical Dynamics: Accessibility of Scaffolding versus Automation Bias

Qualitative findings from the in-depth interviews revealed a marked contrast in the metacognitive responses of lower- and higher-level students to GPT-generated automated feedback. Semester 1 students, who demonstrated relatively limited linguistic awareness, tended to exhibit automation bias by accepting corrections without independently verifying them. In contrast, Semester 3 and Semester 5 students, whose linguistic awareness appeared more developed, engaged in greater critical filtering: they examined the reasoning behind the suggested corrections, rejected over-corrections, and sought to preserve their authorial voice.

a. Clear GPT Explanations as Scaffolding for Reducing Cognitive Load

Participants across all three semester levels reported that the format of GPT feedback was highly helpful compared with conventional instructor feedback. GPT did not merely identify the location of an error; it consistently provided three components of feedback: (1) the original form, (2) the suggested correction, and (3) a communicative explanation of the relevant grammatical rule.

"Usually, when the lecturer just crosses something out or marks it with 'G' or 'T,' we do not know what is wrong. But ChatGPT explains the reason. For example, it says that students should be singular because it comes after every. The explanation is easy to understand, so I immediately get it without having to open a grammar book again."

(Student S1-B, Interview, 2026)

This response suggests that GPT explanations can function as form-focused feedback that reduces the cognitive load students experience when processing corrective feedback.

b. "Just Following Along": Automation Bias among Semester 1 Students

Although the clarity of GPT explanations provided an evident benefit, the interviews also revealed a potentially problematic pedagogical implication among Semester 1 students. Influenced by GPT's confident and authoritative tone, these students tended to adopt its suggestions unquestioningly. They often accepted and copied the suggested revisions without critically evaluating whether the changes accurately reflected the meaning they intended to express.

“I followed all of ChatGPT’s suggestions. I changed everything exactly as it recommended. When it comes to grammar, AI is smarter and knows better English than I do because I am only in my first semester, so I just trusted what it told me.”

(Student S1-D, Interview, 2026)

This tendency can be understood theoretically as a form of **automation bias** (Parasuraman & Manzey, 2010), referring to the tendency to rely excessively on automated systems when making decisions while overlooking one’s own capacity for independent judgment. The difficulty experienced by Semester 1 students in detecting GPT’s miscorrections or over-corrections appears to be related to their relatively limited metalinguistic awareness, which may lead them to treat AI-generated output as an unquestionable source of truth.

c. Learner Agency and the Negotiation of Authorial Voice among Semester 3 and 5 Students

A different pattern emerged among Semester 3 students, particularly those in Semester 5. With a stronger linguistic foundation, these students tended to treat GPT not as an authoritative instructor but rather as a kind of “peer editor.” When GPT produced over-corrections that altered their writing style or weakened the force of their arguments, the more advanced students actively rejected or negotiated the suggested revisions.

“There were some sentences that ChatGPT made much more polished, but the ideas somehow felt bland and lost the distinctive character of my writing. I only accepted the corrections to punctuation or tense, while I kept my own sentence structures and word choices.”

(Student S5-A, Interview, 2026)

This critical stance reflects **learner agency** (Mercer, 2011), whereby students draw on their metalinguistic understanding to retain their authorial voice rather than allowing AI-generated revisions to determine the final form of their writing.

4. Discussion and Theoretical Synthesis

The integration of GPT performance data with qualitative interview findings gives rise to several important theoretical insights within the fields of Computer-Assisted Language Learning (CALL) and AI-supported language pedagogy.

First, the findings support the potential of Transformer-based models such as GPT to function as a source of formative corrective feedback, consistent with the Interaction Hypothesis (Long, 1996) and the concept of the Zone of Proximal Development (ZPD)

(Vygotsky, 1978). By providing immediate, context-sensitive feedback, GPT can help learners *notice the gap* between the interlanguage forms they produce and the target-language forms expected in English. The reasoning-based explanations provided by GPT may also help bridge learners' ZPD by making grammatical patterns more accessible during revision. In this sense, AI-generated feedback can support the internalization of grammatical rules in ways that complement more traditional forms of corrective feedback (Ferris, 2002). This interpretation is consistent with recent research suggesting that ChatGPT can function as a provider of automated written corrective feedback and encourage learners to engage actively in revising their writing, even though the system may occasionally produce contextually inappropriate corrections (Long, H. S., 2024; Yan & Zhang, 2024).

Second, the findings highlight a pedagogical paradox associated with AI-assisted language learning. On the one hand, GPT's highly accessible and persuasive explanations can make feedback easier to understand and may encourage greater learner engagement. On the other hand, the same clarity, particularly when delivered in an authoritative tone, may create an *illusion of correctness* that contributes to automation bias, especially among beginning learners such as the Semester 1 participants. This finding extends the concerns raised by Rozovskaya and Roth (2014) regarding the challenges of automated grammatical error correction. The pedagogical risk of integrating AI into EFL writing therefore extends beyond technical inaccuracies, such as miscorrections and over-corrections. It also involves the possibility of cognitive offloading, whereby students may increasingly rely on AI-generated judgments at the expense of their own critical thinking and metalinguistic awareness.

This perspective is also consistent with recent discussions of automation bias in human–AI collaboration, which suggest that clear system explanations alone may not be sufficient to prevent overreliance. What appears to matter more is whether users are encouraged to engage critically with AI-generated output and independently verify the suggestions they receive (Romeo & Conti, 2025).

The theoretical and pedagogical implications of these findings suggest that the integration of generative AI into EFL writing instruction should not be treated as an entirely unmediated process. Regardless of how sophisticated an AI system may be, it cannot replace the pedagogical role of the teacher. Instead, EFL teachers can serve as mediators who help transform students' interactions with AI from passive acceptance to

critical and reflective engagement. Students, particularly those at lower proficiency levels, need opportunities to learn how to question, verify, and reject AI-generated suggestions when necessary. Such practices can help preserve their independence as writers while allowing them to benefit from the affordances of AI-assisted writing support

Conclusions

This study shows that the freely available version of GPT can be relatively effective in detecting and correcting surface-level language errors, particularly subject-verb agreement, tense forms, and simple sentence structures. The highest correction accuracy was observed in the Semester 1 students' writing, at 78%. However, the rate of over-correction increased to 20% among Semester 5 students as their rhetorical structures and sentence patterns became more complex. These findings suggest that although GPT can be useful for addressing basic grammatical problems, its use with more advanced writing may also encourage stylistic normalization that can affect students' authorial voice.

The patterns of errors further indicate different stages of interlanguage development across the three semester levels. The Semester 1 essays were characterized primarily by L1 transfer, while Semester 3 students showed greater difficulty with clause coordination and collocation. At the Semester 5 level, the challenges shifted toward lexical nuance and academic register. From a metacognitive perspective, GPT's structured and accessible feedback appeared to reduce students' cognitive load, but it also contributed to automation bias among beginning learners. In contrast, intermediate and upper-intermediate students demonstrated greater learner agency by evaluating, filtering, and rejecting corrections that they perceived as inconsistent with their intended meaning or ideas.

Several limitations should be considered when interpreting these findings. First, the relatively small sample of 15 students from a single academic program limits the extent to which the findings can be generalized to the broader EFL population. Second, the study focused exclusively on the freely available version of GPT; therefore, the performance of paid versions or other AI-based writing and language-support systems, such as GPT-4, Claude, or Grammarly Premium, is not represented. Third, the cross-sectional design does not capture how automation bias or learner agency may develop over time.

Future research could address these limitations by involving larger and more institutionally diverse samples, comparing the performance of several contemporary Transformer-based models, and adopting longitudinal designs to examine changes in students' critical AI literacy over time. Further studies could also develop and empirically

evaluate critical AI literacy training modules for beginning EFL learners, particularly those designed to reduce the risk of automation bias while supporting productive and reflective use of AI in writing

References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Bryant, C., Felice, M., Andersen, O. E., & Briscoe, T. (2019). The BEA-2019 shared task on grammatical error correction. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 52–75). Association for Computational Linguistics. <https://doi.org/10.18653/v1/W19-4406>
- Corder, S. P. (1981). *Error analysis and interlanguage*. Oxford University Press.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford University Press.
- Ferris, D. R. (2002). *Treatment of error in second language student writing*. University of Michigan Press. <https://doi.org/10.3998/mpub.9059>
- Hyland, K. (2003). *Second language writing*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511667251>
- Long, H. S. (2024). Exploring the use of ChatGPT as a tool for written corrective feedback in an EFL classroom. *The Journal of AsiaTEFL*, 21(2), 397–412. <https://doi.org/10.18823/asiatefl.2024.21.2.8.397>
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press. <https://doi.org/10.1016/B978-012588825-7/50015-3>
- Mercer, S. (2011). Understanding learner agency as a complex dynamic system. *System*, 39(4), 427–436. <https://doi.org/10.1016/j.system.2011.08.001>

- Nagata, N., & Nakatani, K. (2010). The effectiveness of intelligent computer-assisted language learning in learning Japanese grammar. *Computer Assisted Language Learning*, 23(4), 291–302. <https://doi.org/10.1080/09588221.2010.502018>
- Nurjati, N., Syahria, N., & Khan, A. K. B. S. (2025). ChatGPT-based reflective feedback to improve EFL graduate students' academic writing for publication. *JOLLT Journal of Languages and Language Teaching*, 13(4), 1802–1816. <https://doi.org/10.33394/jollt.v13i4.14858>
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human interaction with automated systems: An integrated review. *Human Factors*, 52(3), 381–410. <https://doi.org/10.1177/0018720810376055>
- Prasetya, R., Syarif, A., & Mayof, A. (2026). Comparative analysis of ChatGPT and Grammarly in supporting grammar correction and writing development in EFL students. *Journal of English and Education*, 12(1), 83–101. <https://doi.org/10.20885/jee.v12i1.44160>
- Puteri, C. G., Cahyono, B. Y., Widiati, U., & Suryati, N. (2026). Using ChatGPT as a writing assistant: A study on essay quality development among Indonesian EFL university students. *Indonesian Journal of Applied Linguistics*, 15(2), 332–346. <https://doi.org/10.17509/3ebmy158>
- Romeo, G., & Conti, D. (2025). Exploring automation bias in human–AI collaboration: A review and implications for explainable AI. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-025-02422-7>
- Rozovskaya, A., & Roth, D. (2014). Building a state-of-the-art grammatical error correction system. *Transactions of the Association for Computational Linguistics*, 2, 419–434. https://doi.org/10.1162/tacl_a_00192
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(1–4), 209–232. <https://doi.org/10.1515/iral.1972.10.1-4.209>
- Sidoti, O., Park, E., & Gottfried, J. (2025). About a quarter of U.S. teens have used ChatGPT for schoolwork – double the share in 2023. Pew Research Center. <https://www.pewresearch.org/short-reads/2025/01/15/about-a-quarter-of-us-teens-have-used-chatgpt-for-schoolwork-double-the-share-in-2023/>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (pp. 5998–6008). Curran Associates, Inc.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

- Yan, D., & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. *Humanities and Social Sciences Communications*, 11, 1086. <https://doi.org/10.1057/s41599-024-03543-y>
- Zhang, Y., Zong, C., & Li, J. (2020). Neural grammatical error correction: A survey. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, 2744–2760. <https://doi.org/10.1109/TASLP.2020.3023023>