



Artificial Intelligence and Natural Language Processing in EFL Writing Assessment: Opportunities and Constraints

Rima SADEK^{1*}

¹University of Algiers 3, (Algeria)

sadek.rima@univ-alger3.dz

Aymen GOMRA²

²University of Algiers 2(Algeria)

aymengomra@gmail.com

Published:31/12/2025

Abstract

This article examines the role of Artificial Intelligence (AI) and Natural Language Processing (NLP) in the assessment of English as a Foreign Language (EFL) writing. As AI-driven tools increasingly shape educational practices, traditional approaches to writing assessment face new methodological and ethical challenges. This study adopts a qualitative secondary research design, critically analyzing three empirical studies that explore the integration of NLP-based systems in EFL contexts. The analysis focuses on key dimensions, including reliability, feedback quality, learner engagement, and issues of academic integrity. The findings indicate that while AI-powered tools enhance efficiency and provide immediate, scalable feedback, they also raise concerns regarding fairness, transparency, and over-reliance on automated evaluation. Moreover, the pedagogical value of such tools remains contingent on their alignment with learner-centered approaches and genre-sensitive writing instruction. The article argues that AI should be positioned as a complementary, rather than substitutive, component of writing assessment. It concludes by highlighting the need for critical integration frameworks that balance technological innovation with pedagogical and ethical considerations.

Keywords:artificial intelligence, EFL writing assessment, natural language processing, pedagogical innovation, secondary research.

*Corresponding author sadek.rima@univ-alger3.dz

Introduction

The rapid advancement of artificial intelligence (AI) has significantly transformed contemporary educational practices, particularly in the domain of language learning and assessment. Within English as a Foreign Language (EFL) context, the integration of natural language processing (NLP) technologies has introduced new possibilities for evaluating writing, offering automated feedback, real-time assessment, and scalable solutions to long-standing pedagogical challenges. These developments are especially relevant in higher education, where increasing student populations and limited instructional time often constrain the effectiveness of traditional writing assessment methods.

Despite these advancements, the adoption of AI-driven assessment tools remains a contested issue. Existing research has predominantly emphasized the efficiency and technical performance of NLP-based systems, particularly their capacity to process large volumes of student writing and generate immediate feedback. However, this technocentric focus has often come at the expense of a deeper examination of pedagogical and ethical implications, including issues of fairness, transparency, and the potential impact on the development of higher-order writing skills. In addition, the literature remains fragmented, with studies frequently addressing isolated implementations of AI tools rather than offering a synthesized understanding of their role within EFL writing assessment.

Addressing this gap, the present study offers a critical analysis of selected empirical research on AI and NLP in EFL writing assessment. It aims to evaluate not only the effectiveness of these tools but also their limitations within pedagogical practice. More specifically, the study investigates how AI-based systems influence feedback quality, learner engagement, and academic integrity, while also examining the extent to which they align with constructivist and genre-based approaches to writing instruction.

This study argues that although AI and NLP technologies offer significant advantages in terms of efficiency and accessibility, their current application in EFL writing assessment remains pedagogically constrained and ethically complex. Rather than replacing human evaluation, these tools should be integrated as complementary mechanisms that support, rather than redefine, the assessment process. To develop this argument, the paper is organized into four main parts. It first outlines the theoretical foundations of AI in language education. It then presents the methodological approach. This is followed by a critical analysis of selected studies. Finally, it discusses the main implications for research and practice.

1.Theoretical Background

This section outlines the key theoretical foundations underpinning the use of Artificial Intelligence and Natural Language Processing in EFL writing assessment. It focuses on the pedagogical and analytical frameworks most relevant to understanding the role, potential, and limitations of automated assessment systems. It also situates these technologies within broader debates on writing as a socially constructed and context-dependent practice, highlighting their implications for language teaching and assessment.

1.1 Artificial Intelligence and Natural Language Processing

Artificial intelligence has increasingly been integrated into educational contexts, particularly through NLP-driven systems capable of analyzing and evaluating written language. In EFL writing assessment, these technologies enable automated scoring, error detection, and instant feedback, thereby addressing challenges related to time constraints and large student populations (Luckin et al., 2016; Zawacki-Richter et al., 2019). NLP-based tools rely on computational models trained on large linguistic datasets to identify patterns in grammar, vocabulary, and discourse structure, offering scalable alternatives to traditional human assessment.

However, the reliability and validity of such systems remain contested. While automated writing evaluation (AWE) tools demonstrate consistency in scoring, they often privilege surface-level features over deeper aspects of writing such as coherence, argumentation, and critical thinking (Perelman, 2014). This limitation raises concerns about construct validity, as the criteria used by AI systems may not fully align with pedagogical objectives in EFL contexts. Consequently, the integration of AI in writing assessment requires careful consideration of both its technical capabilities and its pedagogical implications.

Beyond issues of validity and reliability, the integration of AI in language assessment also raises important concerns related to transparency and interpretability. Many NLP-based systems, particularly those relying on complex machine learning models, operate in ways that are not easily interpretable by users, making it difficult for educators and learners to understand how specific evaluative decisions are produced. As Burrell (2016) argues, algorithmic systems often lack transparency due to the opacity of their internal processes, which can limit their accountability in decision-making contexts. In educational settings, this lack of interpretability may reduce the pedagogical value of automated feedback, as learners receive corrections without clear explanations.

In addition, concerns about algorithmic bias have been increasingly emphasized, particularly when training data do not adequately represent the linguistic diversity of EFL learners (Zawacki-Richter et al., 2019). Such limitations may result in the misinterpretation of non-native language patterns and reinforce standardized norms that do not reflect varied communicative contexts. These issues highlight the need for a critical and responsible integration of AI technologies, where transparency, fairness, and pedagogical alignment are carefully considered.

1.2 Constructivist Perspectives on Writing and Feedback

The pedagogical use of AI in writing assessment can be better understood through a constructivist framework, which views learning as an active, learner-centered process. From this perspective, writing development is not merely the production of correct linguistic forms but the construction of meaning through interaction, reflection, and revision (Vygotsky, 1978; Bruner, 1996). Feedback plays a central role in this process, guiding learners toward higher levels of competence within their zone of proximal development.

AI-powered feedback systems align partially with constructivist principles by providing immediate and individualized responses. However, their effectiveness depends on the nature of the feedback delivered. Research suggests that while automated feedback can support lower-order concerns such as grammar and mechanics, it is less effective in addressing higher-order skills, including organization, argumentation, and audience awareness (Hyland & Hyland, 2006). As a result, AI tools should be viewed as supportive rather than autonomous agents in the learning process, complementing rather than replacing teacher-mediated feedback.

1.3 Genre-Based Approaches to EFL Writing

Genre theory offers a central interpretive framework for understanding writing assessment by foregrounding the social, rhetorical, and communicative purposes of texts. From a genre-based perspective, writing is not merely a linguistic artefact but a socially situated practice shaped by recurring communicative conventions, audience expectations, and disciplinary norms (Swales, 1990; Hyland, 2004). In this sense, effective assessment must extend beyond surface-level linguistic accuracy to account for how successfully a text fulfills its communicative intent within a specific discourse community. Genre pedagogy therefore reframes writing evaluation as an interpretive act that integrates form, function, and context, requiring evaluators to consider both textual features and their situated meanings.

Within this theoretical orientation, AI and natural language processing systems encounter significant operational constraints. Although recent advances in large language models have improved discourse-level processing and coherence detection, their capacity to reliably interpret rhetorical purpose, genre variation, and contextual appropriateness remains limited. Studies in AI-assisted writing evaluation consistently indicate that automated systems are more effective at identifying formal linguistic patterns than at capturing pragmatic and rhetorical dimensions of writing, particularly those that depend on cultural, disciplinary, and situational knowledge (Baker & Hegelheimer, 2023). This limitation reflects the inherent difficulty of encoding genre as a stable computational construct, given its dynamic and context-dependent nature.

This gap reveals a fundamental tension between the standardizing logic of automated assessment systems and the inherently flexible, socially embedded nature of genre-based writing. While AI tools operate through pattern recognition and statistical generalization, genre theory emphasizes variability, purpose-driven variation, and context-sensitive interpretation. As a result, automated evaluation risks reducing writing to decontextualized features that could be measured, potentially overlooking the rhetorical and communicative dimensions that define academic literacy. Consequently, the integration of AI in EFL writing assessment must be carefully aligned with genre-sensitive pedagogies in order to avoid reductive forms of evaluation and to preserve the interpretive depth essential to writing as a communicative act.

1.4 Toward a Critical Integration of AI in Writing Assessment

The integration of artificial intelligence and natural language processing into EFL writing assessment has significantly reshaped contemporary evaluation practices, particularly by enhancing efficiency, scalability, and consistency. Automated writing evaluation systems can process large volumes of student texts and generate immediate feedback, thereby addressing persistent challenges related to teacher workload and delayed feedback cycles. Recent research by Li et al. (2024) demonstrates that AI tools such as ChatGPT are capable of providing balanced feedback across language, content, and organization, highlighting their potential to support more comprehensive assessment practices in EFL classrooms. These developments position AI as a valuable resource for facilitating formative assessment and improving access to individualized feedback.

Despite these advantages, important pedagogical limitations remain. AI-based systems tend to prioritize measurable linguistic features such as grammatical accuracy and lexical complexity, often at the expense of higher-order writing skills, including argumentation, coherence, and rhetorical effectiveness. Chen et al. (2023) emphasize that many automated systems still struggle to evaluate content relevance and discourse-level organization, limiting their capacity to fully capture the multidimensional nature of writing. This creates a misalignment with constructivist and genre-based approaches, which conceptualize writing as a meaning-making process shaped by context, audience, and communicative purpose.

Learner engagement with AI-generated feedback further complicates its pedagogical value. Evidence suggests that students frequently focus on surface-level corrections rather than engaging in deeper revisions that involve critical thinking and restructuring of ideas. Yang et al. (2024) show that learners initially respond to automated feedback in a mechanical manner, prioritizing error correction while demonstrating limited engagement with more complex linguistic and rhetorical suggestions. Such patterns indicate that, although AI can facilitate revision, it does not inherently promote the reflective and dialogic processes central to effective writing development.

In light of these considerations, a balanced and pedagogically grounded approach to AI integration is essential. Rather than replacing human evaluation, AI should be incorporated as a complementary tool within a hybrid assessment model, where automated systems support lower-order feedback and human evaluators remain responsible for higher-order interpretation and judgment. This approach ensures alignment with educational objectives while preserving the complexity of writing as a socially situated practice. Ultimately, the effectiveness of AI in EFL writing assessment depends not on its technological sophistication alone, but on its integration within a coherent pedagogical framework that prioritizes learning, critical engagement, and communicative competence.

2. Previous Empirical Studies

This section critically examines three recent empirical studies investigating the integration of artificial intelligence and natural language processing tools in EFL writing assessment. The selected studies by Li et al. (2024), Chen et al. (2023), and Yang et al. (2024) were chosen for their methodological relevance and their shared focus on AI-supported feedback in second language writing contexts. Rather than summarizing each study independently, the analysis

foregrounds convergences and divergences in their findings in relation to feedback quality, writing development, and learner engagement.

2.1 Li et al. (2024): AI-Generated Feedback and Writing Improvement

Li et al. (2024) examine the effectiveness of ChatGPT-based feedback in improving learners' writing performance. The study was conducted in China with undergraduate EFL students enrolled in an English writing course. Data were collected through pre-test and post-test writing tasks, student written compositions, and revision logs generated during interaction with ChatGPT. The findings suggest that AI-generated feedback contributes positively to learners' grammatical accuracy and lexical development, particularly in revision-based writing tasks. The study highlights the immediacy and accessibility of AI feedback as key advantages, enabling learners to engage in repeated cycles of correction and refinement.

However, a critical reading of the study reveals limitations in its treatment of higher-order writing skills. While improvements in surface-level accuracy are clearly documented, less attention is given to the development of argumentative depth, coherence, and rhetorical awareness. This reflects a broader tendency in AI-supported assessment research to prioritize measurable linguistic gains over discursive and cognitive dimensions of writing. Furthermore, the study does not fully interrogate the extent to which learners internalize feedback or transfer it to new writing contexts, leaving questions regarding long-term pedagogical impact.

2.2 Yang et al. (2024): Learner Engagement with AI Feedback

Yang et al. (2024) investigate how EFL learners interact with AI-generated feedback, particularly in the context of ChatGPT-assisted writing tasks. The study was conducted in a university in South Korea with EFL undergraduate students enrolled in an academic writing course. Data were collected using learner questionnaires, semi-structured interviews, and analysis of students' writing revisions produced with ChatGPT assistance. Their outcomes indicate that students generally perceive AI feedback as useful, accessible, and non-threatening, which enhances their willingness to revise their writing. The immediacy of feedback is identified as a motivational factor that supports iterative writing practices.

Nevertheless, the study also reveals a pattern of superficial engagement. Learners tend to focus primarily on correcting grammatical errors, often neglecting deeper aspects of writing such as argument development and textual organization. In addition, there is evidence of uncritical acceptance of AI suggestions, raising concerns about reduced learner autonomy and over-reliance on automated systems. These findings suggest that while AI can facilitate engagement in revision activities, it does not automatically foster reflective or meaning-oriented writing processes.

2.3 Chen et al. (2023): Systematic Perspectives on AI in Writing Instruction

Chen et al. (2023) provide a systematic review of AI-assisted writing tools in EFL contexts, synthesizing findings from multiple empirical studies. This study is a systematic literature review covering empirical research conducted across multiple international contexts, including Asia, Europe, and North America. The data sources consisted of peer-reviewed journal articles retrieved from Scopus, Web of Science, and ERIC databases, analyzed

through thematic synthesis. Their analysis confirms that AI technologies generally enhance writing accuracy and provide efficient feedback mechanisms, particularly in grammar correction and sentence-level improvements. The study also highlights the increasing integration of automated writing evaluation systems in higher education settings.

Despite its comprehensive scope, the review reveals a significant conceptual limitation: the predominance of studies focusing on technical performance rather than pedagogical depth. The authors note that many AI tools are evaluated primarily in terms of error detection accuracy, with insufficient attention to their role in fostering critical thinking, genre awareness, and communicative competence. This aligns with concerns raised in constructivist and genre-based literature, which emphasize writing as a socially situated and context-dependent practice. Consequently, while Chen et al. (2023) confirm the functional value of AI tools, they also implicitly expose their pedagogical reductionism.

3. Comparative Analysis

A cross-analysis of the three studies reveals a consistent and theoretically significant pattern in the pedagogical impact of AI and NLP tools on EFL writing development. Across all datasets, these technologies demonstrate strong effectiveness in supporting lower-order writing skills, particularly grammatical accuracy, lexical refinement, and sentence-level correction. This convergence suggests that automated systems are well-suited to handling rule-based linguistic operations, where feedback can be standardized, immediate, and computationally reliable. Consequently, automated writing evaluation contributes positively to revision practices by enabling faster error detection, reducing teacher workload, and increasing the frequency of student revisions, which collectively enhance procedural writing efficiency in EFL contexts.

Nonetheless, an analytically divergence emerges when examining higher-order writing competencies. None of the studies provide vigorous evidence that AI systems significantly improve argumentative development, rhetorical sophistication, or critical thinking abilities. This limitation is not merely technical but epistemological in nature, as it reflects the difficulty of encoding meaning-making, contextual interpretation, and discourse-level reasoning into algorithmic models. When interpreted through constructivist and genre-based frameworks, this gap becomes more pronounced: writing is conceptualized not as the accumulation of correct linguistic forms, but as a dynamic process of meaning construction shaped by audience, purpose, and social context. From this perspective, AI systems remain constrained by their inability to fully engage with the interpretive and situated dimensions of writing.

Another important point of convergence across the three studies concerns learner interaction with AI-generated feedback. Although students consistently report positive perceptions of AI tools; particularly regarding accessibility, immediacy, and perceived objectivity, their engagement with feedback tends to remain predominantly surface-oriented. Learners frequently prioritize grammatical correction while overlooking deeper revisions related to organization, coherence, and argumentation. This pattern indicates that the pedagogical affordances of AI are not automatically translated into meaningful learning gains. Instead,

there is a clear disjunction between the technological capabilities of AI systems and their actual pedagogical appropriation in classroom contexts. This disconnect underscores the necessity of mediated implementation, where teacher guidance plays a central role in scaffolding learners' interaction with automated feedback and ensuring that revision processes extend beyond mechanical correction toward reflective and discursive development.

4. Analytical Synthesis

The selected studies illustrate both the pedagogical potential and structural limitations of AI in EFL writing assessment. On one hand, AI-driven tools enhance efficiency, accessibility, and immediate feedback provision, making them valuable supports for writing instruction. On the other hand, their limited capacity to address higher-order cognitive and rhetorical dimensions of writing raises important concerns regarding their pedagogical sufficiency.

From a theoretical perspective, these findings reinforce the relevance of constructivist and genre-based approaches, which emphasize writing as a socially situated and meaning-making process. The evidence suggests that AI systems, while technologically advanced, remain fundamentally constrained in capturing the contextual and interpretive dimensions of writing.

Therefore, a purely automated approach to writing assessment appears insufficient. The studies collectively support the need for a hybrid model in which AI is used to complement, rather than replace, human judgment. Such a model ensures that efficiency gains do not come at the expense of pedagogical depth, interpretive richness, and educational integrity in EFL writing assessment.

5. Discussion

The findings emerging from the selected empirical studies, when interpreted through constructivist and genre-based perspectives, reveal a persistent tension between the technological capabilities of AI systems and the pedagogical complexity of EFL writing. While AI-driven tools demonstrate clear advantages in terms of efficiency, immediacy, and linguistic accuracy, their educational value remains constrained when examined against broader theories of writing development.

From a constructivist standpoint, writing is understood as an active process of meaning-making in which learners develop competence through reflection, interaction, and iterative revision (Vygotsky, 1978). In this regard, the three studies consistently show that AI systems, particularly those examined by Li et al. (2024) and Yang et al. (2024), promote revision activity through immediate feedback. However, this engagement is largely procedural rather than conceptual. Learners tend to focus on correcting surface-level errors, with limited evidence of deeper cognitive restructuring or reflective rewriting. This suggests that while AI can stimulate activity within the writing process, it does not automatically foster the dialogic and mediated learning processes central to constructivist theory.

Similarly, when viewed through a genre-based lens, writing is not merely a linguistic product but a socially situated act shaped by communicative purpose, audience expectations, and disciplinary conventions (Hyland, 2004). Chen et al. (2023) highlight that AI systems are

primarily effective in identifying formal linguistic patterns but remain limited in their ability to evaluate rhetorical appropriateness and genre-specific meaning. This creates a structural misalignment between automated evaluation systems and the contextualized nature of academic writing. Consequently, AI risks promoting a decontextualized model of writing that privileges form over function.

The convergence of findings across the three studies further reinforces this tension. All studies confirm the pedagogical efficiency of AI tools in enhancing grammatical accuracy and providing immediate feedback. However, they also consistently report limitations in supporting higher-order writing skills, such as argument development, coherence, and rhetorical adaptation. This pattern indicates that AI systems operate effectively within a narrow linguistic domain but struggle to engage with the interpretive and communicative dimensions emphasized in both constructivist and genre-based frameworks.

Another key issue emerging from this synthesis concerns learner dependence on automated feedback. Yang et al. (2024) demonstrate that students often accept AI-generated suggestions uncritically, which raises concerns about reduced learner autonomy and diminished critical engagement with writing processes. Rather than fostering independent problem-solving or metalinguistic awareness, AI systems may inadvertently reinforce a correction-oriented approach to writing, thereby limiting the development of deeper writing competence.

Taken together, these insights suggest that the pedagogical integration of AI in EFL writing assessment cannot be understood as a purely technological enhancement. Instead, it must be situated within a broader theoretical and instructional framework that accounts for the complexity of writing as both a cognitive and social practice. While AI tools contribute significantly to efficiency and accessibility, their pedagogical effectiveness is maximized only when embedded within constructivist and genre-sensitive teaching approaches that prioritize meaning-making, contextual awareness, and learner agency.

To sum up, the synthesis of theoretical perspectives and empirical evidence points toward a hybrid model of writing assessment. In such a model, AI functions as a supportive mechanism for addressing lower-order linguistic concerns, while human educators retain responsibility for interpreting meaning, evaluating rhetorical quality, and guiding higher-order writing development. This balanced integration ensures that technological innovation enhances rather than reduces the complexity of EFL writing pedagogy, preserving its fundamentally communicative and socially situated nature.

Conclusion

This paper examined the integration of artificial intelligence and natural language processing tools in EFL writing assessment by synthesizing theoretical perspectives and critically analyzing three recent empirical studies (Li et al., 2024; Yang et al., 2024; & Chen et al., 2023). The analysis was grounded in constructivist and genre-based approaches to writing, which emphasize writing as a meaning-making, socially situated, and context-dependent process rather than a purely linguistic product.

The findings indicate that AI-driven tools offer clear pedagogical advantages, particularly in enhancing efficiency, providing immediate feedback, and supporting the development of lower-order writing skills such as grammar and vocabulary accuracy. However, the studies also consistently demonstrate that these tools are limited in addressing higher-order dimensions of writing, including argumentation, coherence, rhetorical awareness, and genre sensitivity. This limitation reflects a structural misalignment between the operational logic of automated systems and the complex nature of academic writing as conceptualized in constructivist and genre theories.

The analysis further reveals that learner engagement with AI-generated feedback tends to remain surface-oriented, with students focusing primarily on error correction rather than deeper revision processes. This raises important concerns regarding learner autonomy and the risk of over-reliance on automated systems, particularly when AI feedback is not mediated by pedagogical guidance. In this sense, AI does not inherently guarantee improved writing development unless embedded within a structured instructional environment.

In light of these insights, the study concludes that AI should not be viewed as a replacement for human assessment but rather as a complementary tool within a hybrid pedagogical model. Such an approach allows educators to capitalize on the efficiency of AI systems while preserving the interpretive, contextual, and critical dimensions of writing assessment that remain essential in EFL contexts. Ultimately, the effective integration of AI in writing pedagogy depends on maintaining a balance between technological innovation and sound educational principles, ensuring that writing remains a cognitively rich and communicative practice.

References

- Baker, R. S., & Hegelheimer, V. (2023). Artificial intelligence in second language writing assessment: Opportunities and limitations for discourse-level evaluation. *ReCALL*, 35(2), 145–162. <https://doi.org/10.1017/S095834402200018X>
- Burrell, J. (2016) How the Machine “Thinks”: Understanding Opacity in Machine Learning Algorithms. *Big Data & Society*, 3, 1-12. <https://doi.org/10.1177/2053951715622512>
- Bruner, J. (1996). *The culture of education*. Harvard University Press.
- Chen, X., Zou, D., Xie, H., & Cheng, G. (2023). Exploring the effectiveness of AI-assisted writing tools in EFL contexts: A systematic review. *Education and Information Technologies*, 28(9), 11245–11267. <https://doi.org/10.1007/s10639-023-11786-6>
- Hyland, K. (2004). *Genre and second language writing*. University of Michigan Press.
- Hyland, K., & Hyland, F. (2006). Feedback on second language students’ writing. *Language Teaching*, 39(2), 83–101. <https://doi.org/10.1017/S0261444806003399>
- Li, Y., Wang, Z., & Wang, Y. (2024). Evaluating the effectiveness of ChatGPT in providing feedback on EFL writing: A comparative study. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2024.2301234>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education.
- Perelman, L. (2014). When “the state of the art” is counting words. *Assessing Writing*, 21, 104–111. <https://doi.org/10.1016/j.asw.2014.05.001>
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Yang, H., Zhang, L. J., & Dixon, L. Q. (2024). Student engagement with AI-generated feedback in second language writing: Insights from ChatGPT use. *System*, 121, 103196. <https://doi.org/10.1016/j.system.2024.103196>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16(1), 1–27. <https://doi.org/10.1186/s41239-019-0171-0>