International Journal of Education, Humanities and Social Sciences

Review

Artificial Intelligence and Teacher Feedback in Second Language Writing: A Comparative Review and Future Directions

Jiaxin Zhou[1,*]

[1] China West Normal University, Nanchong, China
* Correspondence: Jiaxin Zhou, China West Normal University, Nanchong, China

Abstract: The rapid integration of artificial intelligence, particularly large language models such as ChatGPT, has profoundly reshaped feedback practices in second language (L2) writing instruction. This paper presents a systematic and comprehensive review of contemporary research on AI-generated and teacher-provided feedback in L2 writing contexts, focusing on their respective functions, comparative effectiveness, and interactive relationships. The review clarifies three core thematic areas: the complementary value of dual feedback modes, student engagement with AI-mediated feedback, and optimized prompt design for AI-based writing tools. Findings indicate that AI feedback demonstrates notable advantages over traditional teacher feedback in terms of efficiency, coverage, and capacity for positive reinforcement. However, teacher feedback remains indispensable for delivering personalized guidance, addressing nuanced rhetorical issues, and fostering learner self-regulation and agency. Drawing on these insights, this paper advocates for a hybrid AI-teacher feedback framework that strategically integrates the strengths of both modalities to maximize instructional outcomes. Furthermore, the review identifies critical gaps in the existing literature and proposes several promising future research directions, including learner-centered perspectives, diversified feedback modalities encompassing multimodal and peer-assisted approaches, and the adoption of mixed-methods research designs. By bridging theoretical foundations with practical implications, this review aims to inform educators, researchers, and policymakers about the evolving landscape of feedback in L2 writing and to guide the development of more effective, equitable, and technologically informed pedagogical practices.

Keywords: artificial intelligence; teacher feedback; second language writing; automated writing evaluation; prompt engineering; feedback strategy

Received: 20 May 2026
Revised: 28 June 2026
Accepted: 13 July 2026
Published: 20 July 2026

1. Introduction

Feedback constitutes a foundational element in second language (L2) writing instruction, playing a pivotal role in supporting learners' development of linguistic accuracy, rhetorical awareness, and procedural competence in written expression. It functions not merely as corrective input but as a scaffolded mechanism that fosters metacognitive reflection, encourages revision behaviors, and cultivates sustained engagement with iterative writing practices. Historically, written feedback delivered by instructors has served as the primary mode of formative assessment in L2 writing classrooms. However, this conventional approach faces several persistent operational challenges: feedback tends to be limited in scope due to time and workload constraints, often prioritizing surface-level errors over deeper discourse or genre-related features; evaluation criteria may vary significantly across instructors or even within the same instructor's practice over time, undermining consistency and fairness; and delivery is frequently delayed—sometimes by days or weeks—reducing its immediacy and relevance to learners' cognitive processing during drafting and revision cycles. These limitations collectively diminish the pedagogical impact of feedback, particularly for

learners requiring timely, detailed, and differentiated support to consolidate writing skills progressively [1].

Recent advances in generative artificial intelligence—especially large language models capable of contextual understanding, multi-dimensional text analysis, and adaptive response generation—have introduced new possibilities for feedback provision in L2 writing contexts. Such systems can process full drafts rapidly, identify patterns across lexical, syntactic, discourse, and pragmatic dimensions, and generate tailored suggestions aligned with specific learning objectives or curriculum standards. Unlike static automated tools, contemporary AI models support dynamic interaction, enabling learners to ask clarifying questions, request alternative phrasings, or explore genre-specific conventions through iterative dialogue. This capacity for scalable, responsive, and multimodal feedback addresses longstanding bottlenecks associated with human-delivered feedback, particularly in large-enrollment or resource-constrained educational settings. Moreover, AI systems operate without fatigue or bias related to individual learner profiles, offering consistent baseline support while freeing instructors to focus on higher-order mentoring tasks such as goal-setting, strategic planning, and socio-affective guidance [2].

A substantial and expanding body of empirical work has examined comparative effectiveness, learner reception, and instructional design implications of AI-generated versus instructor-provided feedback in L2 writing environments [3, 4]. Studies have investigated differences in feedback specificity, actionable utility, perceived credibility, and long-term impact on writing development. Others have probed affective and behavioral responses—including motivation, self-efficacy, revision uptake, and dependence tendencies—when learners interact with AI feedback tools. Pedagogical integration research explores optimal implementation models, such as blended feedback workflows, scaffolding sequences, and instructor-AI collaboration protocols that preserve teacher agency while leveraging technological affordances. Despite this proliferation, methodological heterogeneity—including variations in participant demographics, task types, feedback modalities, outcome measures, and theoretical grounding—has resulted in inconsistent findings and limited generalizability. Consequently, there remains a pressing need for rigorous synthesis that maps conceptual alignments, identifies robust empirical trends, surfaces unresolved tensions—such as concerns about overreliance, transparency of AI reasoning, or equity in access—and articulates evidence-informed pathways for future inquiry, policy development, and classroom innovation.

2. Theoretical Foundations and Research Context

2.1. Evolution of L2 Writing Feedback Research

L2 writing feedback research has evolved systematically over the past three decades, reflecting broader shifts in applied linguistics and pedagogical theory. Early investigations primarily adopted qualitative, classroom-embedded approaches to document common feedback practices and teacher beliefs. As methodological rigor increased, researchers began employing controlled experimental designs and longitudinal case studies to examine how different feedback modalities—such as direct correction, metalinguistic explanation, and peer-mediated commentary—influence linguistic accuracy, rhetorical development, and learner engagement. A growing body of work now integrates cognitive, sociocultural, and ecological perspectives to explore feedback not merely as an instructional intervention but as a dynamic, context-sensitive process embedded within larger learning ecosystems. Domestically, Chinese scholarship on L2 writing feedback has similarly progressed through discernible stages: initial exploratory analysis laid conceptual groundwork by identifying prevalent patterns in teacher-student written interaction; subsequent phases emphasized pragmatic implementation, with educators adapting feedback frameworks to align with curriculum standards, class size constraints, and institutional assessment policies; more recent efforts prioritize multi-dimensional interactive research, incorporating learner agency, digital affordances, and

cross-modal feedback integration—including audio annotations, video conferencing, and multimodal response formats—to support holistic writing development [5, 6].

Although both domestic and international research communities share core thematic concerns—including automated writing evaluation systems, formative assessment principles, and teacher-student feedback dialogues—their theoretical grounding and practical orientation diverge meaningfully. International scholarship frequently situates feedback inquiry within foundational frameworks such as Vygotskian sociocultural theory, interactionist SLA models, and genre-based pedagogy, emphasizing long-term developmental trajectories and the co-construction of knowledge through feedback-mediated dialogue. In contrast, domestic research tends to foreground classroom feasibility, curriculum alignment, and measurable short-term outcomes—such as improvements in grammatical accuracy, task completion rates, or student self-efficacy following targeted feedback interventions. This divergence does not indicate methodological deficiency but rather reflects distinct educational priorities, systemic constraints, and epistemological traditions [7]. Increasingly, however, convergent trends are emerging, particularly in the adoption of mixed-methods designs, learner-centered feedback protocols, and technology-enhanced feedback scaffolding that bridges theoretical insight with pedagogical utility across diverse instructional contexts.

2.2. Practical Value and Limitations of Teacher Feedback

Teacher-written feedback continues to serve as a cornerstone of effective second-language writing instruction, particularly in contexts where learners require sustained, personalized guidance to develop rhetorical awareness and linguistic accuracy. Empirical investigations into classroom feedback practices have revealed persistent discrepancies between pedagogical intentions and actual implementation. For instance, while educators often articulate a principled preference for addressing higher-order concerns—such as argument coherence, logical development, and genre-appropriate organization—in initial drafts, and reserving attention to surface-level features—such as grammatical accuracy, punctuation, and lexical precision—for later revision stages, observational data consistently show that linguistic corrections dominate feedback across all phases of the writing process. This divergence suggests systemic challenges related to time constraints, assessment pressures, and limited professional development opportunities in feedback literacy [8]. Moreover, comparative analyses between teacher self-reports and student perceptions indicate that learners frequently perceive significantly more feedback than instructors believe they deliver, pointing to potential gaps in metacognitive awareness, communication clarity, or consistency in feedback delivery. Such findings underscore the need for structured training programs that enhance teachers' capacity to align feedback goals with instructional design, scaffold learner autonomy, and foster reflective practice through systematic observation and peer collaboration.

Beyond its diagnostic function, teacher feedback plays a distinctive and irreplaceable role in advancing learners' higher-order writing competencies, especially in academic and research-oriented contexts. Evidence indicates that carefully calibrated, task-specific feedback—particularly when integrated with explicit strategy instruction—can effectively mitigate common developmental challenges among advanced L2 writers, including overdependence on prefabricated phrases, formulaic sentence structures, and shallow paraphrasing behaviors that prioritize lexical substitution over conceptual reformulation [6, 9]. In graduate-level writing tasks requiring critical synthesis and original argumentation, such feedback supports the gradual internalization of disciplinary conventions, strengthens source integration skills, and cultivates rhetorical flexibility across genres. Importantly, this impact is not merely additive but transformative: it enables learners to move beyond mechanical error correction toward metacognitive engagement with their own writing processes, fostering greater ownership, strategic decision-making, and long-term transferability of skills. These outcomes highlight feedback not as a static evaluative tool but as a dynamic, dialogic, and pedagogically

mediated activity that bridges theory and practice, intention and execution, and individual cognition and sociocultural context.

2.3. Application and Dilemmas of AI Feedback in L2 Writing

Large language models, exemplified by widely adopted generative AI systems, have significantly transformed automated writing evaluation frameworks in second-language writing instruction [10, 11]. A comprehensive synthesis of recent developments identifies three principal functional domains for artificial intelligence in this educational context: the generation of immediate, rule-based or pattern-informed feedback on linguistic form; the provision of holistic and analytic scoring aligned with pedagogical rubrics; and the delivery of adaptive scaffolding that supports learners through iterative drafting processes. However, these applications are accompanied by persistent methodological and practical constraints. Output consistency remains a critical concern, as model responses may vary substantially across similar input prompts or repeated submissions. Contextual sensitivity is frequently limited—models often fail to interpret discipline-specific conventions, genre expectations, or culturally embedded rhetorical norms that shape effective L2 writing. Furthermore, ethical considerations arise regarding data privacy, transparency of algorithmic decision-making, potential reinforcement of linguistic biases present in training corpora, and the risk of overreliance undermining learner agency and metacognitive development.

Empirical investigations into learner experiences reveal nuanced perceptions of AI-generated feedback. While users consistently value its speed, scalability, and capacity to identify surface-level grammatical, lexical, and mechanical errors—such as subject-verb agreement, article usage, or punctuation—many express reservations about interpretive depth and pedagogical validity. Learners report frequent mismatches between AI suggestions and intended meaning, especially in cases involving idiomatic expression, pragmatic appropriateness, or register-sensitive phrasing. Crucially, students demonstrate clear epistemic awareness: they recognize AI tools as highly competent for diagnostic correction of formal accuracy but fundamentally inadequate for higher-order writing development, including argumentation coherence, conceptual elaboration, audience adaptation, and stylistic refinement. This functional boundary shapes strategic tool use—learners tend to deploy AI selectively, primarily during early drafting stages for error screening, while reserving human interaction for substantive revision and reflective feedback integration. Consequently, attitudes toward AI assistance reflect a pragmatic duality: appreciation for its operational utility coexists with skepticism about its pedagogical sufficiency, resulting in cautious, context-dependent adoption rather than wholesale substitution of instructor-led guidance.

3. Comparative Analysis of AI and Teacher Feedback

3.1. Differences in Feedback Attitude and Expression Strategies

A recent empirical investigation involving fifty student compositions identified notable distinctions between artificial intelligence–generated feedback and human teacher feedback, particularly concerning affective orientation and communicative framing [12]. Analysis revealed that AI systems consistently adopted a more affirming evaluative stance, especially when addressing macro-level features such as overall organizational coherence, paragraph sequencing, and rhetorical structure. This tendency toward positivity appears rooted in algorithmic design priorities that emphasize encouragement and confidence-building through reinforcement rather than critique. In contrast, human instructors demonstrated a pronounced emphasis on linguistic accuracy, grammatical conformity, lexical appropriateness, and syntactic precision—prioritizing diagnostic identification of deviations from standard academic conventions. Such pedagogical focus reflects longstanding traditions in writing instruction that view error awareness as foundational to language development and disciplinary literacy acquisition. The divergence is not merely stylistic but epistemological: AI feedback often functions as a normative benchmark, whereas teacher feedback operates as a scaffolded interpretive

dialogue grounded in contextualized understanding of individual learner trajectories, curricular objectives, and genre-specific expectations.

Regarding expression strategies, both modalities predominantly employed explicit feedback—clearly identifying issues and specifying required revisions—but diverged significantly in implementation logic and cognitive scaffolding intent. Artificial intelligence systems typically delivered prescriptive corrections by directly substituting erroneous constructions with linguistically standardized alternatives, thereby minimizing ambiguity and accelerating surface-level revision [13, 14]. While this approach enhances immediate task efficiency and reduces cognitive load during editing, sustained reliance may inadvertently diminish opportunities for metacognitive reflection, hypothesis testing, and self-regulated error analysis. Human teachers, by contrast, frequently utilized metalinguistic commentary—such as labeling clause types, naming grammatical categories, or prompting syntactic reanalysis—which invites learners to engage in active interpretation, rule inference, and conceptual integration. This heuristic method fosters deeper encoding of linguistic knowledge, strengthens procedural memory for grammatical operations, and cultivates transferable analytical competencies across diverse writing contexts. Empirical evidence suggests that such cognitively demanding feedback supports longer-term retention, adaptive application in novel tasks, and progressive internalization of writing conventions—outcomes essential for achieving advanced academic literacy and disciplinary communication proficiency.

3.2. Complementary Mechanisms of Dual Feedback Modes

Empirical evidence consistently demonstrates the complementary strengths of artificial intelligence–driven feedback and human teacher feedback in second language writing instruction. Artificial intelligence systems demonstrate particular efficacy in detecting, diagnosing, and correcting surface-level linguistic errors—including grammar, spelling, punctuation, and lexical collocation—with high consistency, broad coverage, and immediate response time. In contrast, human instructors possess irreplaceable capacity for interpreting rhetorical intent, evaluating argumentative coherence, assessing genre-appropriate discourse structures, and adapting guidance to learners' evolving cognitive development, affective states, and sociocultural backgrounds. This functional differentiation underscores a fundamental pedagogical principle: standardization and scalability—hallmarks of AI support—must be deliberately integrated with contextual sensitivity and developmental responsiveness—the core competencies of professional educators. Consequently, contemporary writing pedagogy increasingly emphasizes hybrid feedback architectures that strategically distribute tasks across technological and human agents, thereby reconciling efficiency demands with deep learning objectives.

Controlled experimental investigations have repeatedly shown that learners who receive systematically coordinated feedback from both AI tools and human teachers outperform peers relying exclusively on either modality alone. The integration yields synergistic benefits: AI augments the breadth and frequency of formative input by enabling continuous, low-stakes practice opportunities outside classroom hours; it also reduces instructor workload associated with routine error identification, freeing cognitive resources for higher-order instructional tasks. Meanwhile, teacher involvement remains indispensable—not merely as a quality assurance mechanism but as an active curatorial agent who interprets AI-generated suggestions, filters irrelevant or misleading outputs, contextualizes corrections within broader learning goals, refines feedback language for learner comprehension, and introduces metacognitive scaffolding that fosters self-regulated writing development. Such deliberate human mediation ensures that feedback remains pedagogically grounded, linguistically accurate, culturally appropriate, and aligned with curriculum standards and individualized learning trajectories.

Research further indicates that well-designed AI feedback significantly enhances learner engagement across multiple dimensions. When feedback is personalized, actionable, and embedded within meaningful writing tasks, it contributes to the strengthening of learners' ideal second language self-concept—a motivational construct

reflecting their envisioned future proficiency and identity as L2 users. This psychological shift manifests in measurable improvements in affective engagement (e.g., increased writing confidence and reduced anxiety), behavioral engagement (e.g., greater task persistence, revision frequency, and voluntary practice), and cognitive engagement (e.g., deeper processing of feedback, strategic use of linguistic resources, and transfer of learned features to new writing contexts). Moreover, sustained exposure to responsive, adaptive AI feedback cultivates learner agency and reflective practice, encouraging writers to view errors not as deficits but as diagnostic signals guiding iterative improvement. These cumulative effects collectively establish a robust foundation for long-term writing competence development, supporting progression from controlled accuracy toward fluent, purposeful, and socially situated L2 expression.

4. AI Feedback Optimization: Prompt Engineering and Strategy Improvement

4.1. The Bridge Prompt Optimization Model

The quality of AI-generated feedback is fundamentally contingent upon the structural integrity, clarity, and pedagogical intentionality embedded within the prompt design. Empirical observations in English as a second language writing instruction have revealed substantial variability in the coherence, relevance, and instructional utility of automated feedback—particularly in complex generative tasks such as text continuation, where linguistic fluency, logical progression, and rhetorical appropriateness must be simultaneously evaluated. To systematize prompt construction and enhance output reliability, a comprehensive six-dimensional framework has been formulated to guide educators and instructional designers. This framework explicitly integrates Background context to situate the task within authentic communicative scenarios; Role specification to define the AI's functional stance (e.g., expert writing tutor, peer reviewer, or linguistic analyst); precise Instructional directives that articulate step-by-step procedural expectations; clearly articulated Goal statements that foreground desired learning outcomes; curated Data inputs that include genre-appropriate source texts, learner-produced drafts, or error typologies; and concrete Examples or Expectation benchmarks that model ideal response formats, depth of analysis, and evaluative granularity. Crucially, the Goal and Instruction dimensions are designated as non-negotiable core components, as their omission consistently correlates with vague, off-task, or procedurally inconsistent outputs. Validation studies demonstrate measurable improvements across four interrelated performance metrics: enhanced internal content logic through improved causal and sequential reasoning; heightened grammatical accuracy via targeted syntactic and morphological correction; strengthened discourse coherence achieved through explicit signaling of thematic continuity and paragraph-level transitions; and expanded lexical and syntactic diversity reflected in richer collocational range, varied clause structures, and register-appropriate phrasing.

Beyond its technical efficacy, the framework serves as a scalable pedagogical scaffold that empowers educators to move beyond ad hoc prompting toward evidence-informed, repeatable, and classroom-integrated AI feedback practices. Its standardized architecture supports consistent implementation across diverse institutional contexts, proficiency levels, and curricular objectives. Importantly, the model underscores the critical function of constraint articulation: explicitly delineating both required analytical foci—such as identifying cohesion devices or evaluating argumentative validity—and prohibited elements—including speculative commentary, subjective value judgments, or prescriptive grammar rules unsupported by contemporary applied linguistics research—significantly sharpens feedback specificity, reduces interpretive ambiguity, and increases alignment with domain-specific assessment criteria. Furthermore, the framework facilitates iterative refinement, enabling teachers to diagnose prompt-related failure modes (e.g., overgeneralization, misalignment with rubric descriptors, or insufficient contextual grounding) and adjust input parameters accordingly. This operational

transparency not only strengthens the formative value of AI feedback but also cultivates teacher agency in human–AI collaborative assessment ecosystems.

4.2. Optimization of Feedback Focus and Diversified Strategies

Beyond prompt engineering, the strategic optimization of feedback focus and the deliberate integration of multiple feedback modalities constitute essential levers for enhancing the pedagogical efficacy and holistic quality of writing feedback in second language acquisition contexts. Historically, instructional feedback has tended to prioritize surface-level linguistic accuracy—such as grammar, spelling, and syntax—at the expense of deeper, more cognitively demanding dimensions like rhetorical organization, argumentative coherence, conceptual development, and genre-appropriate discourse features. This imbalance risks reinforcing a narrow, error-avoidance orientation that may inadvertently suppress students' willingness to experiment with complex syntactic structures or nuanced lexical choices. In contrast, contemporary AI-assisted feedback mechanisms demonstrate broader coverage across functional domains: they routinely address macro-level concerns—including paragraph cohesion, logical sequencing, thesis clarity, and audience adaptation—as well as micro-level linguistic features. Such multidimensional responsiveness helps redress longstanding structural gaps in conventional teacher-led feedback, thereby supporting more balanced skill development aligned with communicative competence frameworks. Moreover, AI-generated feedback often exhibits greater consistency in application and scalability across large cohorts, enabling timely responses without compromising breadth of coverage—though its capacity for contextual interpretation and socio-pragmatic nuance remains inherently constrained by current model architectures [6, 15].

However, reliance solely on automated, prescriptive correction—characterized by immediate, unmediated identification and rectification of perceived errors—carries significant pedagogical risks, particularly concerning learner agency and metacognitive growth. Unmoderated direct correction may foster dependency, reduce opportunities for self-regulation, and discourage reflective revision practices essential for long-term writing development. Consequently, an emerging consensus in applied linguistics and writing studies emphasizes the necessity of hybrid, multi-source feedback ecosystems. These systems intentionally combine complementary strengths: teacher feedback offers contextualized, formative guidance grounded in pedagogical expertise and interpersonal scaffolding; peer feedback cultivates collaborative learning, critical evaluation skills, and genre awareness through authentic interaction; self-feedback promotes metacognitive monitoring, goal-setting, and ownership of the writing process; while machine feedback delivers rapid, scalable, pattern-based analysis across large textual corpora. To design such integrated systems effectively, educators must rigorously evaluate each modality—not only for technical reliability and construct validity—but also for its alignment with learners' developmental readiness, affective responses, perceived usefulness, and actual uptake in subsequent revisions. Systematic calibration across these interdependent dimensions ensures feedback functions not merely as assessment, but as a dynamic, adaptive catalyst for sustained writing proficiency [4, 16].

5. Discussion: Construction of AI-Teacher Hybrid Feedback Model

A comprehensive synthesis of empirical findings from ten rigorously designed studies consistently indicates that artificial intelligence (AI) and human teacher feedback serve distinct yet mutually reinforcing functions within second language (L2) writing instruction [3, 17]. Rather than positioning AI as a replacement for pedagogical expertise, the evidence strongly supports an integrated, synergistic approach—where AI augments rather than supplants teacher agency. This hybrid feedback model emerges not merely as a pragmatic compromise but as the most educationally sound and scalable pathway forward, aligning with contemporary principles of differentiated instruction, formative assessment, and learner-centered pedagogy. Its feasibility is further reinforced by growing institutional capacity in digital infrastructure, teacher professional development

initiatives focused on AI literacy, and evolving frameworks for ethical AI integration in educational contexts.

First, AI tools demonstrate exceptional efficiency in generating high-volume, consistent feedback across large cohorts, thereby enabling pedagogical practices previously constrained by time and workload limitations [4]. For instance, advanced language models can analyze and respond to fifty student essays within one hour—facilitating iterative, multi-draft writing cycles that are foundational to process-oriented instruction. Second, AI-generated feedback exhibits notable strengths in breadth and affective balance: it systematically addresses linguistic accuracy, rhetorical structure, lexical range, and discourse coherence while deliberately highlighting strengths alongside areas for improvement. This dual-focus approach fosters greater learner self-efficacy, reduces anxiety associated with error-focused evaluation, and encourages sustained engagement with revision tasks. Third, teacher feedback retains irreplaceable value in delivering nuanced, context-sensitive guidance that transcends algorithmic capabilities. Educators uniquely interpret individual learning histories, diagnose underlying cognitive or linguistic challenges, provide metalinguistic explanations grounded in disciplinary knowledge, and scaffold the development of sophisticated rhetorical judgment and genre awareness—capacities no current AI system can authentically replicate. Moreover, teachers play a decisive role in mitigating overreliance on automated tools by modeling critical evaluation of AI output and cultivating learner autonomy through reflective practice and strategic feedback literacy. Fourth, educators function as essential architects and quality stewards of the hybrid model: AI feedback does not operate autonomously but requires deliberate instructional design—including purposeful prompt engineering, contextual calibration, manual verification, and pedagogical refinement. Teachers integrate AI-generated insights into lesson planning, adapt outputs to curriculum objectives and learner profiles, and mediate between the standardization inherent in algorithmic processing and the flexibility demanded by diverse classroom needs—thereby transforming technological capability into meaningful, adaptive writing instruction innovation.

6. Conclusion and Future Research Directions

This paper presents a systematic synthesis of recent empirical and conceptual developments in second language (L2) writing feedback research, with particular attention to the evolving roles of artificial intelligence–driven tools and human instructors. The analysis reveals that AI-mediated feedback and teacher-delivered feedback operate along distinct yet interdependent dimensions of pedagogical functionality. Artificial intelligence systems demonstrate notable strengths in delivering timely, scalable, and consistent feedback across multiple linguistic and rhetorical dimensions—such as grammar, lexical accuracy, surface-level coherence, and affective reinforcement—thereby supporting foundational skill consolidation and learner motivation. In contrast, teacher feedback remains indispensable for addressing higher-order writing competencies, including rhetorical stance, genre awareness, argumentative depth, and discourse-level cohesion. Moreover, teachers uniquely provide metalinguistic explanations grounded in contextual understanding, scaffolded revision strategies tailored to individual developmental trajectories, and responsive interpersonal mediation that fosters writer identity and agency. These complementary affordances underscore that neither modality supplants the other; rather, their integration yields synergistic potential when deliberately coordinated within coherent instructional frameworks.

Building on this integrative perspective, five priority directions for future inquiry are advanced. First, rigorous learner-centered investigations are needed to examine how affective, cognitive, and sociocultural variables—including self-efficacy, feedback orientation, linguistic confidence, and prior digital learning experiences—shape learners' interpretation, uptake, and implementation of feedback from diverse sources. Second, longitudinal and comparative studies should assess the cumulative impact of hybrid feedback configurations—such as sequential, parallel, or adaptive combinations of AI and

teacher input—on sustained writing development across proficiency levels and academic genres. Third, methodological pluralism must be strengthened through purposeful integration of qualitative, quantitative, and multimodal approaches—including think-aloud protocols, interactional analytics, corpus-informed writing trajectory mapping, and classroom ethnography—to triangulate findings and enhance validity. Fourth, prompt engineering frameworks require iterative refinement and context-sensitive adaptation—not only across educational settings (e.g., EFL vs. ESL contexts, tertiary vs. secondary education) but also across linguistic varieties, task types, and learner profiles. Fifth, comprehensive professional development initiatives must be designed to cultivate educators' critical AI literacy: encompassing technical fluency, ethical discernment, pedagogical design competence, and reflective practice in orchestrating human–AI collaboration.

As artificial intelligence capabilities continue to evolve rapidly, the central challenge confronting L2 writing researchers, curriculum designers, and classroom practitioners is no longer framed as a binary choice between automated and human feedback. Instead, the imperative lies in co-designing, empirically validating, and ethically implementing robust hybrid feedback ecosystems—systems wherein AI augments rather than replaces pedagogical judgment, where teacher expertise guides AI deployment, and where learner needs remain the unwavering focal point of all technological and instructional decisions. Such ecosystems must be grounded in principles of transparency, equity, adaptability, and developmental responsiveness, ensuring that feedback practices advance not only linguistic accuracy but also rhetorical sophistication, critical voice, and lifelong writing agency.

References

1. L. Ding, D. Zou, and L. Kohnke, "ChatGPT as an automated writing evaluation tool: how students perceive it and how it affects their writing," *Education and Information Technologies*, pp. 1–23, 2025.
2. N. Chen, W. Xiong, and E. Di Zhang, "Evaluating the effect of teacher written feedback on developing Chinese L2 graduate students' paraphrasing performance," *Journal of English for Academic Purposes*, vol. 77, p. 101553, 2025.
3. J. Han and M. Y. Li, "Exploring ChatGPT-supported teacher feedback in the EFL context," *System*, vol. 126, Art. no. 103502, 2024.
4. D. R. Ilgen, C. D. Fisher, and M. S. Taylor, "Consequences of individual feedback on behavior in organizations," *Journal of Applied Psychology*, vol. 64, no. 4, p. 349, 1979.
5. I. Lee, "Teacher written corrective feedback: Less is more," *Language Teaching*, vol. 52, no. 4, pp. 524–536, 2019.
6. M. Feng Teng, "A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations," *International Journal of TESOL Studies*, vol. 6, no. 3, 2024.
7. C. Van Beuningen, "Corrective feedback in L2 writing: Theoretical perspectives, empirical insights, and future directions," *International Journal of English Studies*, 2010.
8. Y. Yao, X. Zhu, L. Xiao, and Q. Lu, "Secondary school English teachers' application of artificial intelligence-guided chatbot in the provision of feedback on student writing: An activity theory perspective," *Journal of Second Language Writing*, vol. 67, p. 101179, 2025.
9. N. Ismail, S. Maulan, and N. H. Hasan, "The impact of teacher feedback on ESL students' writing performance," *Academic Journal of Social Studies*, vol. 8, no. 1, pp. 45–54, 2008.
10. T. Chhorn, "A study of teacher feedback impact on Cambodian EFL students' L2 writing: perception, performance, and attribution," *Journal of Education Innovation and Language Teaching (JEILT)*, 2025.
11. T. Wongvorachan, K. W. Lai, O. Bulut, Y. S. Tsai, and G. Chen, "Artificial intelligence: Transforming the future of feedback in education," *Journal of Applied Testing Technology*, pp. 95–116, 2022.
12. M. Molina-Soria, J. L. Aparicio-Herguedas, T. Fuentes-Nieto, and V. M. López-Pastor, "The combination of artificial intelligence and formative assessment in teacher education: A systematic review," *Encyclopedia*, vol. 6, no. 3, p. 66, 2026.
13. A. A. Aldino, B. Maheshi, Y. Li, Y. Zhou, Y. S. Tsai, D. Gašević, and G. Chen, "Enhancing learner-centered feedback with AI: teachers' practices and perceptions," *Assessment & Evaluation in Higher Education*, pp. 1–18, 2026.
14. Y. Xu, J. Zhong, and Y. Han, "How teacher feedback literacy changed by using Generative Artificial Intelligence (GenAI) feedback as exemplars: Case studies of English as a foreign language school teachers in China," *Teaching and Teacher Education*, vol. 176, p. 105501, 2026.
15. A. Maleki, "Effects of AI generated and teacher feedback on EFL learners writing performance and emotional experience," *Discover Artificial Intelligence*, vol. 6, no. 1, p. 199, 2026.
16. T. Fuentes-Nieto, J. L. Aparicio-Herguedas, and V. M. López-Pastor, "Artificial Intelligence and Formative and Shared Assessment in Teacher Education," *Applied Sciences*, vol. 15, no. 22, p. 12067, 2025.

17. I. Celik, M. Dindar, H. Muukkonen, and S. Järvelä, "The promises and challenges of artificial intelligence for teachers: A systematic review of research," *TechTrends*, vol. 66, no. 4, pp. 616–630, 2022.