Asia Pacific Science Press

Journal of Educational Theory and Practice
ISSN(O):3005-9313
ISSN(P):3005-9305
Vol. 2 No.4 (2025)

A Review of Research on Artificial Intelligence Writing Feedback: Mechanisms, Learner Behaviours, and Future Directions

Gufeng Wu[1]*, Rosilawati Sueb[1], Qingyi Deng[2]

1.Faculty of Education, Puncak Alam Campus, Universiti Teknologi MARA (UiTM), 40450, Selangor, Malaysia

2.Faculty of Media and Art Design, Guangzhou Huali College, Guangdong, 511325, China

Corresponding author: Gufeng Wu, wgf0829@gmail.com

Abstract: This study systematically reviews the evolution of Artificial Intelligence (AI) writing feedback from Automated Writing Evaluation (AWE) to generative Large Language Models. Existing evidence indicates that AI enhances writing quality. However, research regarding the mechanisms through which feedback translates into concrete revision behaviours remains insufficient. Drawing upon Feedback Intervention Theory and Writing Process Theory, this paper analyses the distinctions between local linguistic feedback and global discourse feedback. It further explores critical variables including learner cognitive assessment, revision intention, and affective responses. A comprehensive conceptual framework is constructed. This framework elucidates the mediating role of feedback satisfaction in connecting technical characteristics with revision depth. Addressing current limitations in process data collection and short-term designs, the article calls for a shift towards longitudinal empirical paradigms combined with behavioural tracking. Ultimately, the study underscores the necessity of enhancing feedback literacy and establishing ethical norms within the context of human-machine synergy to achieve sustainable development in writing education.

Keywords: Artificial Intelligence Writing Feedback; Automated Writing Evaluation; Generative AI; Revision Behaviour; Feedback Satisfaction; Learner Engagement

Published: Dec 27, 2025

DOI: https://doi.org/10.62177/jetp.v2i4.961

1.Introduction

1.1 Research Background and Importance

Over the past decade, the functionality of AI writing feedback tools has undergone a significant transformation, evolving from traditional Automated Writing Evaluation (AWE) systems, such as Criterion and Pigai, to Large Language Models (LLMs) represented by ChatGPT[1]. Initially focused on grammatical error correction and surface-level linguistic monitoring, these tools are now capable of providing multi-layered suggestions regarding content development, discourse structure, logical organisation, and rhetorical strategies. Consequently, the utility of AI has expanded beyond the scope of an auxiliary proofreading tool[2]. It is now increasingly conceptualised as an intelligent writing partner. A wealth of existing literature corroborates the efficacy of these technologies in diverse settings[3]. This includes higher education and classrooms focused on English as a Foreign or Second Language. Both AWE systems and generative AI assistants have been proven to aid students in refining linguistic accuracy. Furthermore, they support the development of textual organisation and elevate the overall quality of writing. Furthermore, their potential to facilitate learner attention to language, alleviate cognitive load, and reinforce process-oriented writing has garnered increasing empirical support[4].

Notwithstanding these advancements, current research exhibits notable limitations that constrain a comprehensive understanding of the mechanisms underlying AI writing feedback. Primarily, studies remain heavily result-oriented, with a scarcity of evidence regarding behavioural processes. Most empirical inquiries utilise pre-test and post-test designs to ascertain improvements in writing quality, yet pay limited attention to how learners interpret feedback, the way they selectively adopt suggestions, and the cognitive and behavioural operations that occur during the actual revision process[5]. Although existing research has revealed that feedback contributes to outcome improvement, explanations regarding how feedback is translated into writing enhancement through behavioural processes remain tenuous[6].

Secondly, there is a prevalent lack of theoretical scaffolding. Reviews have highlighted that AWE research frequently lacks systematic elucidation based on feedback theory, learner engagement theory, or writing process theory. This results in studies that appear more as technical effect tests rather than responses to critical issues in educational and learning theories, particularly in explaining the mechanistic chain linking feedback quality, learner perception, and writing behaviour[7]. Such deficiencies render many studies unable to explain 'why' feedback is effective or ineffective and impede the formation of a more cumulative body of knowledge.

Thirdly, the era of generative AI writing has introduced a new set of challenges. Distinct from traditional AWE systems, which primarily offer local and surface-level feedback, Large Language Models (LLMs) such as ChatGPT can generate deep-level feedback that includes explanations, examples, and even direct rewriting. While this capability significantly enhances the comprehensibility of feedback, it simultaneously triggers novel research concerns: whether high reliance on AI might diminish autonomous revision capabilities; whether generative feedback blurs the boundaries of authorship and originality; and how learner trust or scepticism towards AI feedback influences adoption behaviours. Furthermore, questions arise regarding the specific 'feedback literacy' learners require to utilise these tools effectively in such contexts[8]. These issues extend beyond technical efficacy, touching upon profound challenges in cognitive, ethical, and educational policy domains.

Against this backdrop, a systematic review is essential to delineate the current landscape of AI writing feedback research from a holistic perspective. Although the volume of studies concerning Automated Writing Evaluation (AWE) and generative AI feedback has increased in recent years, a comprehensive framework capable of integrating technical features, learner psychological experiences, behavioural response mechanisms, and writing learning outcomes remains absent. Existing literature tends to focus on isolated dimensions—such as comparisons of feedback types (e.g., immediate vs. delayed; surface vs. deep), the direct impact of feedback on writing quality, or subjective perceptions of system usability—while failing to connect these elements into a cohesive causal chain. Consequently, it is imperative to scrutinise AI writing feedback research through the lens of 'feedback characteristics—learner experience—behavioural response—writing outcome' to drive theoretical deepening and practical optimisation.

1.2 Research Objectives

To address these needs, the present review seeks to achieve four objectives. First, it aims to systematically summarise core findings regarding current AI writing feedback (encompassing both traditional AWE systems and LLM-based generative AI), specifically focusing on major developments in feedback types, functional positioning, and pedagogical effects, thereby mapping the evolutionary path of feedback mechanisms under different technological paradigms. Second, adopting a dual perspective of learner behaviour and psychology, this review intends to explicate engagement patterns, cognitive processing, and characteristics of revision behaviour following the receipt of AI feedback. This includes an analysis of revision depth, adoption rates, emotional experiences, and trust in the tool, to reveal how AI feedback is utilised in authentic learning contexts. Third, building upon existing theories, a behavioural mechanism framework centred on feedback satisfaction is proposed. This framework elucidates how AI writing feedback influences learner value judgements, willingness to revise, and strategy selection, ultimately translating into observable revision behaviours that foster the development of writing competence. Finally, this paper identifies critical gaps in current research regarding methodology, application contexts, learner individual differences, and feedback design, offering theoretical support and reference designs for future empirical inquiries to promote the systematisation and profundity of AI writing feedback research.

2.Theoretical Foundations

2.1 Feedback Intervention Theory and Feedback Functions

Feedback Intervention Theory (FIT)[9] provides a critical perspective for comprehending the effects of feedback. The theory posits that the efficacy of feedback is contingent upon the level at which learner attention is directed, encompassing the task, process, self-regulation, and self-levels. Performance enhancement is effectively promoted when feedback concentrates on the task itself (e.g., error identification, problem explanation) and associated processes (e.g., ideation strategies, logical organisation). Conversely, feedback emphasising individual self-evaluation or capability judgement may instigate negative self-attention, thereby diminishing learning motivation and performance quality. In the context of AI writing feedback, contemporary mainstream systems (such as AWE or generative AI) typically deliver substantial information at the task and process levels. This includes linguistic error annotation, expression paraphrasing suggestions, textual organisation prompts, and logical coherence analysis, thus theoretically possessing high pedagogical potential[10]. However, given the capacity of AI to present large-scale, dense, and multi-dimensional feedback, the information density often exceeds that of human teacher feedback. This disparity may compel learners to process a vast quantity of suggestions within a short timeframe, creating risks associated with increased cognitive load, attention dispersion, or ‘feedback fatigue’. Consequently, achieving an equilibrium between providing sufficient information and maintaining manageability has emerged as a pivotal issue in technological design and pedagogical application.

Within the realm of educational assessment and instructional feedback, the classic feedback function framework proposed by Hattie and Timperley[11] further deepens the understanding of effective feedback. They articulated three core questions from a learning-oriented perspective: ‘Where am I going?’, ‘How am I going?’, and ‘Where to next?’. This framework underscores that feedback should not merely identify issues but must also provide direction and strategies. Applying this framework to AI writing feedback facilitates an effective distinction between the feedback levels and functions of various systems. Certain systems are restricted to ‘problem identification’ (such as highlighting grammatical or spelling errors), which constitutes typical task-level feedback. In contrast, advanced systems (e.g., generative AI based on Large Language Models) can offer ‘revision strategies’ and ‘explanations of writing principles’, thereby addressing deeper questions regarding ‘how to modify’ (process level) and ‘why to modify in this manner’ (self-regulation level). Existing literature also indicates that feedback possessing explanatory, strategic, or model-demonstration characteristics is more likely to stimulate learner willingness to revise and facilitate deep learning[12]. Therefore, basing analysis on the functional classification of Hattie and Timperley allows for a more systematic evaluation of the ‘information quality’ and ‘actionability’ of AI feedback, providing a theoretical foundation for subsequent behavioural mechanism models.

2.2 Writing Process Theory and Revision Behaviour

Process Writing Theory conceptualises writing as a dynamic and recursive cognitive sequence, encompassing distinct stages such as planning, translating, reviewing, and revising[13]. Within this framework, revision behaviour is widely regarded as the central mechanism fostering the development of writing competence. This is because revision entails not only the correction of surface-level linguistic elements but also the profound reprocessing of content logic, discourse structure, and argumentation strategies[14]. During the revision phase, writers are required to continuously compare the ‘goal text’ against the ‘current text’ and adjust based on identified discrepancies. Such a process intrinsically embodies self-monitoring and metacognitive regulation within the writing activity.

In the context of AI writing feedback applications, technological intervention is predominantly concentrated within the ‘reviewing revising’ stage. On the one hand, AI assists learners in identifying deficiencies at the linguistic, content, and structural levels through mechanisms such as grammatical annotation, error detection, and logical analysis, thereby enhancing their capacity for problem detection[15]. On the other hand, generative AI can provide alternative expressions, sentence restructuring, paragraph reorganisation, and even exemplary texts. These features enable learners to acquire more explicit directions for revision alongside richer linguistic input. Furthermore, the immediate feedback provided by AI influences the learners' level of self-assessment regarding text quality, which subsequently regulates their cognitive judgements concerning the necessity of revision.

Drawing upon Process Writing Theory, it is evidently insufficient to evaluate the efficacy of AI feedback solely based on the final scores of written products. Recent scholarship suggests that the impact mechanisms of AI writing feedback should be examined through procedural indicators[16]. These include whether learners generate a revision intention, the depth of revision undertaken (surface versus deep), the specific revision strategies employed (such as substitution, deletion, expansion, or rewriting), and whether the revision behaviour reflects cognitive investment and reflective processing[17]. These mediating process variables serve to unveil the pathways through which AI feedback functions in the development of learner writing, holding significant value for comprehending the genuine merit of technological intervention. In other words, the effectiveness of AI writing feedback depends not critically on 'what is said' by the system, but rather on 'what is done' by the learner and 'how it is executed'[18].

2.3 Learner Engagement and Perspectives on Feedback Literacy

In recent years, the paradigm of feedback research has witnessed a marked transition: moving from a traditional emphasis on 'how teachers provide feedback' towards a focus on 'how students comprehend, utilise, and transform feedback'. This shift underscores the proactive agency of learners within the feedback process. Systemic reviews indicate that learner engagement with written corrective feedback can be delineated into behavioural, cognitive, and affective dimensions, which collectively determine whether feedback translates into tangible writing improvement[19]. Behavioural engagement is manifested in revision frequency and depth; cognitive engagement involves the interpretation, judgement, and integration of feedback content; while affective engagement encompasses emotional factors such as trust, anxiety, and dependency, all of which directly influence the willingness to adopt feedback[20].

Within the specific context of AI writing feedback, learner engagement exhibits more diversified patterns. A subset of learners demonstrates high dependency, tending to accept AI suggestions 'wholesale' without deep judgement regarding their validity. Conversely, other learners display greater agency, forming a mode of 'active regulatory engagement' by comparing multiple feedback sources, adopting suggestions selectively, and adjusting them in alignment with their own writing objectives[21]. Concurrently, certain learners, driven by scepticism regarding AI credibility, reduce interaction frequency, utilising the system merely as an auxiliary tool for checking grammar or spelling[7]. These disparities illustrate that AI writing feedback does not automatically yield learning effects; its potential to genuinely foster writing development is heavily contingent upon the quality of learner engagement[22].

To explicate the disparities, the feedback literacy framework offers substantial theoretical support. This framework posits that learners require not only the ability to comprehend the feedback content itself but also the capacity to evaluate its quality, judge its applicability, and ultimately convert it into concrete revision actions[23]. These three competencies have become particularly critical in the era of generative AI. Given that AI systems may produce inconsistencies, ambiguities, or even erroneous information, the judgement and regulatory capabilities of the learner directly dictate their ability to benefit from the technology. Consequently, enhancing learner feedback literacy—specifically the capacity for critical evaluation of AI feedback and self-regulation—has emerged as a core condition for promoting the pedagogical effectiveness of AI writing feedback.

3.Types and Characteristics of AI Writing Feedback

3.1 Local Linguistic Feedback

Local linguistic feedback primarily centres on surface-level linguistic forms, encompassing the normative aspects of grammar, spelling, punctuation, lexical collocation, and sentence structure. Historically, this category of feedback has been extensively utilised by Automated Writing Evaluation (AWE) systems, such as Criterion, Write Improve, and Pigai, characteristically relying on linguistic rule detection, statistical models, or machine learning algorithms to identify and annotate linguistic errors[24]. A substantial body of research indicates that local linguistic feedback plays a significant role in enhancing learner linguistic accuracy. Particularly within the demographic of elementary and intermediate EFL/ESL learners, the provision of large-scale, immediate error correction effectively reduces syntactic errors and improves lexical appropriateness and linguistic standardisation[25]. Nevertheless, the scope of influence exerted by local linguistic feedback is typically confined to the surface level, offering limited promotion of higher-order writing competencies such as content development, discourse coherence, and logical organisation.

From the perspective of pedagogical application, local linguistic feedback presents multiple advantages. Firstly, given the capacity of systems to instantly generate a multitude of specific and actionable revision suggestions, learners can obtain clear error localisation and rewriting examples within a brief timeframe, which is conducive to forming rapid error awareness and correction cycles[26]. Secondly, local feedback is generally presented through annotations, substitutions, or localised suggestions. These formats are capable of being reviewed repeatedly and are straightforward to manipulate, thereby requiring lower comprehension costs and alleviating cognitive load during the feedback processing stage[27]. Furthermore, for learners with a weaker linguistic foundation, local feedback offers a controllable and explicit learning pathway, which aids in the establishment of writing confidence.

Notwithstanding these benefits, local linguistic feedback also demonstrates distinct limitations. Primarily, an excessive reliance on surface-level error correction tends to lead learners to concentrate on linguistic forms while neglecting content development and the quality of argumentation, thereby reinforcing ‘surface revision’ rather than deep-level meaning construction[28]. Certain studies suggest that when confronted with a high volume of error annotations, learners may experience ‘feedback fatigue’, inclining them towards the mechanical acceptance of modifications without a comprehension of the root causes of errors or linguistic functions. Secondly, the frequent provision of overly granular correction suggestions by systems may attenuate learner self-monitoring capabilities, causing them to become reliant on technology during the writing process and reducing their willingness to actively detect errors. Additionally, as local feedback from AI systems may exhibit inconsistencies or misjudgements, learners lacking feedback literacy find it difficult to assess feedback quality, potentially leading to negative transfer when erroneous feedback is adopted[29].

In summary, while local linguistic feedback remains a vital foundation for improving linguistic accuracy within AI writing feedback, its genuine pedagogical value is contingent upon whether learners can integrate this surface-level information into broader writing objectives and the process of meaning construction. Consequently, in actual instruction and system design, it is necessary to combine local feedback with other types of feedback. This integration serves to guide learners away from overly formalised and fragmented revision behaviours, promoting holistic development at both the linguistic and content levels.

3.2 Global Feedback

Global feedback is primarily orientated towards the macro-organisational and meaning-construction dimensions of writing. Its focal points encompass the rationality of discourse structure, the logical progression between paragraphs, the coherence and completeness of argumentation, and the clarity and sufficiency of thesis statements. Distinct from local linguistic feedback, global feedback is not restricted to isolated sentences or vocabulary; rather, it evaluates the text holistically to determine whether the intended communicative purpose has been achieved. For instance, systems may flag issues such as ‘unclear overall logic’, ‘lack of effective transitions’, ‘indistinct paragraph topic sentences’, or ‘insufficient evidentiary support’, thereby guiding learners to ameliorate their writing from the perspectives of overall structure and content development[30].

Propelled by technological advancements, certain AWE systems and the new generation of generative AI tools (such as the GPT series and Write Wise) have acquired the capability to provide relatively complex evaluations and suggestions at the discourse level. These include paragraph reorganisation, analysis of argumentative coherence, and judgements on content coverage. Research indicates that this category of global feedback holds significant value for fostering higher-order writing competencies in learners. Particularly within academic and argumentative writing, it assists students in comprehending argumentative structures and strengthening logical consistency both within and between paragraphs[31]. Furthermore, global feedback directs learner attention towards writing task requirements and the target audience, thereby enhancing overall writing quality and rhetorical effect[32].

The effectiveness of global feedback varies considerably among learners because it is abstract and relies on metacognitive processes such as clarifying writing purposes, reasoning logically and planning discourse. Learners therefore need strong language proficiency and monitoring abilities to interpret and apply it. When told that "paragraph organisation is problematic" or "argumentation lacks hierarchy," they often have to reassess and reorganise the whole text, which imposes far greater cognitive load than handling linguistic errors and may lead those with weaker foundations to struggle or even avoid using such feedback. At the same time, AI tools still face limitations in providing global feedback; constrained discourse analysis can produce overly general, ambiguous or non-actionable suggestions, reducing learner trust and uptake. As a result, the actual impact of global feedback depends on the interaction among technological quality, pedagogical support and learners' feedback literacy[33].

Overall, global feedback possesses irreplaceable pedagogical value in enhancing higher-order writing skills. However, its effective utilisation must be predicated on the learner possessing sufficient cognitive resources and strategic capabilities. It simultaneously necessitates that systems provide suggestions that are logically clear and highly actionable, to facilitate deep revision by learners at the levels of writing planning and meaning construction.

4.Learner Engagement and Behavioural Response to AI Writing Feedback

4.1 Cognitive Interpretation and Trust Mechanisms

Learners' cognitive interpretation of AI feedback and their degree of trust constitute the psychological mechanism that transforms feedback into concrete revision behaviours. Existing studies show that learners are not passive recipients but make final decisions through judgement, comparison and selective processing, meaning that the effectiveness of AI feedback depends not only on system quality but also on cognitive and affective trust. High trust promotes rapid adoption based on the principle of least effort, especially for clear surface-level errors, leading to a high-adoption surface revision pattern[34]. When trust is ambiguous, learners alternate between acceptance and scepticism or use AI merely to verify their own ideas, increasing cognitive load but also preserving writing autonomy. Low trust relegates AI to a secondary checking tool used mainly for polishing or spelling, while deeper evaluation of logic or structure is rejected and replaced by teacher authority or learners' own linguistic knowledge[34]. External contexts further modulate this relationship: in high-stakes academic settings, learners act cautiously due to fear of misalignment with assessment standards or teacher expectations, with institutional norms playing a decisive role[35]; in low-risk, autonomy-oriented environments, they are more inclined to treat AI as a private tutor for explanations, examples and structural guidance[36].

Overall, accurate comprehension and moderate trust form the psychological foundation for effective AI-assisted writing, and future research should attend to systemic factors influencing trust such as algorithmic transparency, explainable feedback design and the development of student feedback literacy.

4.2 Revision Intention and Extent

Revision intention acts as a critical precursor to behavioural investment. It denotes the learner's plan to modify text following feedback. In generative AI contexts, this intention relies on the perceived value of feedback and its alignment with task objectives. Learners are more likely to revise when suggestions effectively enhance text quality or adhere to specific genre norms. Furthermore, clarity and actionability represent vital mechanisms. Schiller et al.[37] indicate that transparent feedback with executable pathways effectively alleviates cognitive load. Conversely, ambiguous suggestions increase comprehension costs and weaken motivation. External factors such as time pressure and self-efficacy also moderate this process. Empirical studies by Fu and Liu[38] and Crosthwaite and Sun[39] corroborate that specific and high-quality feedback significantly strengthens revision intention.

Revision extent constitutes a multidimensional construct encompassing both quantitative breadth and qualitative depth. Breadth refers to physical variations like word count. Depth pertains to cognitive levels and structural hierarchy. The feedback modality exerts a significant influence here. Traditional AWE often prompts local linguistic corrections. In contrast, GenAI facilitates substantial text modification through rewriting suggestions. However, large-scale revision does not invariably equate to deep cognitive investment. Some students merely adopt AI content without critical reflection. Others engage in deep re-creation. Process data analysis reveals distinct behavioural pathways. Schiller et al.[37] note that high-engagement learners exhibit recursive editing modes. Low-engagement learners frequently demonstrate linear and rapid acceptance. Consequently, mere revision quantity is insufficient to measure effectiveness. It must be evaluated alongside the cognitive investment involved.

4.3 Affective and Motivational Responses

AI writing feedback serves as more than a mere cognitive resource for information transfer. It functions as a psychological trigger capable of significantly stimulating learner affective responses and motivational shifts. This influence is characterised by distinct complexity and duality.

On the one hand, immediate and effective AI feedback can ameliorate emotional experiences by alleviating cognitive load. Language barriers frequently precipitate high levels of anxiety within L2 writing contexts. Research by Rahman et al.[40] indicates that immediate feedback and visualised text improvements provided by AI effectively bolster student self-confidence in writing. These features also mitigate writing anxiety. Consequently, a positive reinforcement loop for learning is established.

Conversely, technological limitations may incite negative affective blockages. Learners are prone to feelings of frustration and confusion when AI systems frequently flag errors without sufficient explainability. Similar reactions occur when there is significant dissonance between AI feedback and teacher evaluation criteria. If such cognitive conflict remains unresolved, it may lead students to develop technological mistrust. This subsequently inhibits their motivation to utilise these tools for learning support[40].

Furthermore, the dimension of affective response has expanded from simple mood fluctuations to deeper crises regarding morality and identity with the intervention of Generative AI (GenAI). The robust generative capabilities of GenAI have reshaped the nature of human-computer interaction. This makes student emotional experiences replete with contradictions. A study by Teng[17] found that some students perceive tools like ChatGPT as collaborative partners. They emphasise emotional reliance on these tools for cognitive expansion and writing support alongside their positive value. In contrast, other students exhibit significant apprehension. They worry that excessive reliance on AI might undermine individual writing agency or trigger ethical concerns regarding academic integrity. This psychological struggle between tool empowerment and competence deprivation constitutes a unique affective landscape in current AI writing education.

5.A Comprehensive Conceptual Framework for AI Writing Feedback

Drawing on the systematic literature review, this study proposes a conceptual framework that explains how AI writing feedback is transformed into revision behaviours and writing outcomes through a chain of cognitive and affective processes. The core pathway consists of six linked components: feedback characteristics, cognitive assessment, feedback satisfaction, revision intention, revision extent, and writing outcome. The framework highlights that AI feedback does not directly improve writing performance. Instead, it works through a series of mediating mechanisms shaped by learner perceptions[41].

Feedback characteristics form the initial input affecting cognitive assessment. Key dimensions include feedback type (e.g., local linguistic vs. global discourse feedback), quality (e.g., accuracy, specificity, consistency), and interactive load (e.g., information volume and interface presentation). Based on these features, learners develop cognitive judgements concerning credibility, task relevance, and comprehensibility[42].

Cognitive assessment then shapes overall feedback satisfaction, which combines evaluative judgement with affective responses. As a central mediator, satisfaction strongly predicts revision intention and influences whether learners engage in surface-level or deep-level revision. Evidence from AWE and generative AI research consistently shows that high satisfaction promotes active uptake, whereas deep revision occurs only when learners invest substantial cognitive effort[43].

The final stage concerns writing outcomes, including both product-oriented indicators (linguistic accuracy, discourse structure, argument quality) and capability-oriented indicators (self-efficacy, metacognitive strategies, long-term writing development), forming a complete loop from technical input to learning effectiveness[44].

Overall, this framework contributes to theory building by clarifying the dynamic relationships between feedback characteristics, learner cognition, and revision behaviour, identifying feedback satisfaction as a key psychological mechanism, and underscoring the value of process data such as revision logs and behavioural trajectories for future empirical research on AI-supported writing.

6.Methodological Challenges

Contemporary scholarship regarding AI writing feedback exhibits a notable deficiency in the depth and multidimensionality of data collection. This deficit is primarily characterised by a scarcity of process data alongside a potential disjunction between self-reported information and actual behaviour. Certain vanguard studies have commenced attempts to utilise keystroke logging, screen recording, and learning analytics to capture micro-revision behaviours within AI feedback contexts[45]. However, the prevailing paradigm remains excessively reliant on summative evaluations derived from pre-test and post-test scores. Consequently, there is a lack of quantitative description and deep mining regarding the intermediate revision process. Current revision classification systems are predominantly predicated on time-consuming manual coding[46]. These systems have not yet fully leveraged large-scale log data and natural language processing tools for automated analysis. As a result, research struggles to unveil the dynamic cognitive trajectories inherent in learner-AI interactions[47]. More critically, exclusive reliance on self-reported data such as questionnaires and interviews may lead to severe measurement bias. Research findings indicate that high levels of self-reported satisfaction by students do not necessarily translate into high levels of actual revision behaviour or substantive writing improvement. This inconsistency between subjective and objective data necessitates that future inquiries adopt multi-source data triangulation strategies[48]. It is essential to combine subjective perceptual data with objective behavioural data, such as revision logs and version comparisons, to enhance the internal validity of research conclusions.

Regarding the temporal dimension and research design, the prevalence of short-term and single-task orientations constitutes another major methodological bottleneck within the field. A vast number of studies tend to employ one-off writing tasks or short-cycle experimental designs[49]. Consequently, researchers often struggle to capture the sustained effects and transfer value of AI writing feedback on the long-term development of learner writing competence. Such cross-sectional designs are highly susceptible to interference from the novelty effect. This implies that positive reactions exhibited by learners may stem more from the freshness of new technology rather than genuine cognitive investment or capability enhancement. This phenomenon often leads to an overestimation of the actual pedagogical efficacy of AI tools[50].

Furthermore, rapid technological iteration and the limitations of research contexts pose severe challenges to the external validity and reproducibility of findings. The update velocity of generative AI technology is extremely fast. It is a common occurrence that the system version used in a study has undergone fundamental changes by the time of publication. This creates a common difficulty for cross-study comparison and replication[51]. Concurrently, sample sources in existing literature exhibit significant geographical imbalance. They are concentrated primarily in higher education contexts within Europe, America, China, Japan, and South Korea. Conversely, empirical evidence from basic education stages in Southeast Asia, the Middle East, and non-English speaking countries remains relatively scarce. This cultural singularity of samples makes it difficult to isolate the contextual influences of culture and educational systems. Consequently, it is challenging to determine whether existing findings possess universality or are merely products of specific cultural backgrounds[40].

7.Conclusion and Future Directions

AI writing feedback technology has undergone a significant paradigm shift. It has evolved from early local error correction tools represented by Automated Writing Evaluation (AWE) to generative AI ecosystems capable of providing multi-layered and conversational support[42]. Overall, existing evidence indicates that AI writing feedback possesses immense potential to enhance learner writing quality and self-efficacy under appropriate instructional design and usage conditions. However, the efficacy of this technology is not solely a result of the tool itself. Rather, it is highly dependent on learner cognitive understanding of the feedback, affective satisfaction, and subsequent concrete revision behaviours. Consequently, this study constructs a comprehensive conceptual framework. This framework encompasses AI feedback, cognitive assessment, satisfaction, revision intention, revision extent, and writing outcomes. It reveals from a theoretical perspective how feedback characteristics are transformed into actual writing behaviours through the subjective psychological experiences of the learner. Furthermore, it confirms the critical mediating role played by satisfaction in connecting technical characteristics with behavioural performance.

Future academic exploration requires a fundamental transformation at the methodological level from a result-orientated approach to a process-orientated one to further validate and deepen this theoretical framework. The research emphasis must shift from mere writing score evaluation to micro-process analysis combining keystroke logging, version control, and learning analytics technologies. It is essential to construct revision mechanism models based on behavioural data. This will allow for the precise quantification of the specific contributions of surface and deep revisions to writing quality. Concurrently, future empirical designs should transcend the laboratory paradigm to address the limitations of short-term effects prevalent in current research. Longitudinal and ecological research pathways are required. Researchers can circumvent the interference of the novelty effect by deeply integrating AI tools into long-cycle curriculum systems. This approach also allows for a profound comparison of the differential efficacy of diverse pedagogical modes. These modes include teacher-feedback dominance supplemented by AI, AI-feedback dominance regulated by teachers, and hybrid models combining AI and peer feedback. Such comparisons will provide empirical grounds for different educational contexts.

Finally, the widespread application of AI writing feedback must be scrutinised under a framework of human-machine synergy sustainability and ethics. At the level of pedagogical practice, educators should design specialised units for feedback literacy and metacognitive training. The focus should be placed on cultivating learner abilities to screen AI suggestions critically and process them deeply. Additionally, the role of the teacher must transform from a basic error corrector to a guide of higher-order thinking. At the institutional level, it is imperative to establish clear ethical norms and policy guidelines. These measures are necessary to balance the tension between technological empowerment and academic integrity within curriculum and institutional spheres. In summary, the future development of AI writing feedback lies not only in the enhancement of algorithmic precision. It relies more significantly on constructing a responsible and learner-centred hybrid feedback ecosystem within multi-cultural contexts. This is the path to truly realising the deep empowerment of writing education by technology.

Funding

This research was supported by funding from Universiti Teknologi MARA (UITM).

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

[1] Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching and Learning Practice, 20(2). https://doi.org/10.53761/1.20.02.07

[2] Mizumoto, A., & Eguchi, M. (2023). Exploring the Potential of Using an Ai Language Model for Automated Essay Scoring (SSRN Scholarly Paper No. 4373111). Social Science Research Network. https://doi.org/10.2139/ssrn.4373111

[3] Wang, L., & Ren, B. (2024). Enhancing Academic Writing in a Linguistics Course with Generative AI: An Empirical Study in a Higher Education Institution in Hong Kong. Education Sciences, 14(12), 1329. https://doi.org/10.3390/educsci14121329

[4] Fan, N., & Ma, Y. (2022). The Effects of Automated Writing Evaluation (AWE) Feedback on Students' English Writing Quality: A Systematic Literature Review. Language Teaching Research Quarterly, 28, 53–73. https://doi.org/10.32038/ltrq.2022.28.03

[5] Kurt, G., & Kurt, Y. (2024). Enhancing L2 Writing Skills: ChatGPT as an Automated Feedback Tool. Journal of Information Technology Education: Research, 23, 024. https://doi.org/10.28945/5370

[6] Saeli, H., Rahmati, P., & Koltovskaia, S. (2023). Learner Engagement with Written Corrective Feedback: The Case of Automated Writing Evaluation. Journal of Response to Writing, 9(2). https://scholarsarchive.byu.edu/journalrw/vol9/iss2/1

[7] Cotos, E. (2023). Automated Feedback on Writing. In O. Kruse, C. Rapp, C. M. Anson, K. Benetos, E. Cotos, A. Devitt, & A. Shibani (Eds), Digital Writing Technologies in Higher Education: Theory, Research, and Practice (pp. 347–364). Springer International Publishing. https://doi.org/10.1007/978-3-031-36033-6_22

[8] Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. Cogent Education, 11(1), 2410101. https://doi.org/10.1080/2331186X.2024.2410101

[9] Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin, 119(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

[10] Zaheer, S., Ma, C., Zhu, Y., & Vasinda, S. (2025). GenAI in Academic Writing- Empowering Learners or Redefining Traditional Pedagogical Practices? A Systematic Review From 2019-2023. International Journal of Artificial Intelligence (AI) in Teaching and Learning (IJAITL), 1(1), 1–34. https://doi.org/10.4018/IJAITL.373582

[11] Hattie, J., & Timperley, H. (2007). The Power of Feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487

[12] Yang, H., Gao, C., & Shen, H. (2024). Learner interaction with, and response to, AI-programmed automated writing evaluation feedback in EFL writing: An exploratory study. Education and Information Technologies, 29(4), 3837–3858. https://doi.org/10.1007/s10639-023-11991-3

[13] Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. College Composition & Communication, 32(4), 365–387. https://doi.org/10.58680/ccc198115885

[14] Li, W., & Zhang, F. (2021). Tracing the Path Toward Self-Regulated Revision: An Interplay of Instructor Feedback, Peer Feedback, and Revision Goals. Frontiers in Psychology, 11. https://doi.org/10.3389/fpsyg.2020.612088

[15] Al-Othman, A. A.-A. M. (2024). EFL Students Develop Cognitive and Metacognitive Self-Regulated Writing Strategies Using Automated Feedback: A Case Study. Theory and Practice in Language Studies, 14(5), 1525–1536. https://doi.org/10.17507/tpls.1405.26

[16] Suraworachet, W., Zhou, Q., & Cukurova, M. (2023). Impact of combining human and analytics feedback on students' engagement with, and performance in, reflective writing tasks. International Journal of Educational Technology in Higher Education, 20(1), 1. https://doi.org/10.1186/s41239-022-00368-0

[17] Teng, M. F., & Huang, J. (2025). Incorporating ChatGPT for EFL Writing and Its Effects on Writing Engagement. International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT), 15(1), 1–21. https://doi.org/10.4018/IJCALLT.367874

[18] Tsao, J.-J. (2024). Unveiling the Links Between EFL Learners' Perceived L2 Writing Classroom Goal Structures, Engagement with Teacher Written Corrective Feedback, and Self-Reported Writing Proficiency Through Structural Equation Modeling. Sage Open, 14(4), 21582440241299161. https://doi.org/10.1177/21582440241299161

[19] Shen, R., & Chong, S. W. (2023). Learner engagement with written corrective feedback in ESL and EFL contexts: A qualitative research synthesis using a perception-based framework. Assessment & Evaluation in Higher Education, 48(3), 276–290. https://doi.org/10.1080/02602938.2022.2072468

[20] Pearson, W. S. (2024). Affective, behavioural, and cognitive engagement with written feedback on second language writing: A systematic methodological review. Frontiers in Education, 9. https://doi.org/10.3389/feduc.2024.1285954

[21] Shi, Y. (2021). Exploring Learner Engagement with Multiple Sources of Feedback on L2 Writing Across Genres. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.758867

[22] Yu, S., & Liu, C. (2021). Improving student feedback literacy in academic writing: An evidence-based framework. Assessing Writing, 48, 100525. https://doi.org/10.1016/j.asw.2021.100525

[23] Rad, H. S., & Mirzaei, A. (2024). Developing feedback literacy, scaffolded writing, and resilience through intervention on feedback processes for L2 writing students. Studies in Educational Evaluation, 81, 101354. https://doi.org/10.1016/j.stueduc.2024.101354

[24] Ding, L., Zou, D., & Kohnke, L. (2025). ChatGPT as an automated writing evaluation tool: How students perceive it and how it affects their writing. Education and Information Technologies. https://doi.org/10.1007/s10639-025-13775-3

[25] Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized controlled trial. Frontiers in Psychology, 14. https://doi.org/10.3389/

fpsyg.2023.1249991

[26] Mao, Z., Lee, I., & Li, S. (2024). Written corrective feedback in second language writing: A synthesis of naturalistic classroom studies. Language Teaching, 57(4), 449–477. https://doi.org/10.1017/S0261444823000393

[27] Thi, N. K., & Nikolov, M. (2023). Effects of teacher, automated, and combined feedback on syntactic complexity in EFL students' writing. Asian-Pacific Journal of Second and Foreign Language Education, 8(1), 6. https://doi.org/10.1186/s40862-022-00182-1

[28] Wang, I. X., Wu, X., Coates, E., Zeng, M., Kuang, J., Liu, S., Qiu, M., & Park, J. (2024). Neural Automated Writing Evaluation with Corrective Feedback (No. arXiv:2402.17613). arXiv. https://doi.org/10.48550/arXiv.2402.17613

[29] Chen, Q. (2025). Students' Perceptions of AI-Powered Feedback in English Writing: Benefits and Challenges in Higher Education. International Journal of Changes in Education. https://doi.org/10.47852/bonviewIJCE52025580

[30] Ozan, O., & Azap, S. (2025). Students' Perceptions of ChatGPT as a Feedback Tool in EFL Writing. Open Access Journal of Education & Language Studies, 3(1). https://doi.org/10.19080/OAJELS.2025.03.555604

[31] Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. Journal of Second Language Writing, 27, 1–18. https://doi.org/10.1016/j.jslw.2014.10.004

[32] Sari, E., & Han, T. (2024). The impact of automated writing evaluation on English as a foreign language learners' writing self-efficacy, self-regulation, anxiety, and performance. Journal of Computer Assisted Learning, 40(5), 2065–2080. https://doi.org/10.1111/jcal.13004

[33] Cheng, X., & Zhang, L. J. (2024). Examining Second Language (L2) Learners' Engagement with AWE-Teacher Integrated Feedback in a Technology-Empowered Context. The Asia-Pacific Education Researcher, 33(4), 1023–1035. https://doi.org/10.1007/s40299-024-00877-8

[34] 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(1), 1086. https://doi.org/10.1057/s41599-024-03543-y

[35] Tang, X. (2025). L2 Writing with AI: Perceptions and Engagement of EFL Learners in China. English Language Teaching, 18(2), p68. https://doi.org/10.5539/elt.v18n2p68

[36] Ekizoğlu, M., & Demir, A. N. (2025). The role of AI assisted writing feedback in developing secondary students writing skills. Discover Education, 4(1), 454. https://doi.org/10.1007/s44217-025-00919-3

[37] Schiller, R., Fleckenstein, J., Höft, L., Horbach, A., & Meyer, J. (2025). On the role of engagement in automated feedback effectiveness: Insights from keystroke logging. Computers & Education, 238, 105386. https://doi.org/10.1016/j.compedu.2025.105386

[38] Fu, H., & Liu, X. (2022). EFL Learner Engagement in Automatic Written Evaluation. Frontiers in Psychology, 13, 871707. https://doi.org/10.3389/fpsyg.2022.871707

[39] Crosthwaite, P., & Sun, S. (2025). Generative AI and L2 Written Feedback Studies: A Scoping Review. RELC Journal, 00336882251386530. https://doi.org/10.1177/00336882251386530

[40] Rahman, N. A. A., Zulkornain, L. H., & Hamzah, N. H. (2022). Exploring Artificial Intelligence using Automated Writing Evaluation for Writing Skills. Environment-Behaviour Proceedings Journal, 7(SI9), 547–553. https://doi.org/10.21834/ebpj.v7iSI9.4304

[41] Xiaosa, L., & Ping, K. (2025). How L2 student writers engage with automated feedback: A longitudinal perspective. Assessing Writing, 64, 100919. https://doi.org/10.1016/j.asw.2025.100919

[42] Zhang, Z. (Victor). (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. Assessing Writing, 43, 100439. https://doi.org/10.1016/j.asw.2019.100439

[43] Cengiz, B. C., Bilki, Z., Ataş, A. H., & Celik, B. (2025). Exploring second language writers' engagement with ChatGPT feedback: Revision behaviors and perceptions. System, 134, 103837. https://doi.org/10.1016/j.system.2025.103837

[44] Ajabshir, Z. F., & Ebadi, S. (2023). The effects of automatic writing evaluation and teacher-focused feedback on CALF measures and overall quality of L2 writing across different genres. Asian-Pacific Journal of Second and Foreign

Language Education, 8(1), 26. https://doi.org/10.1186/s40862-023-00201-9

[45] Youn, C. H., & Salam, A. R. (2025). A Systematic Review on the Use of Artificial Intelligence in Providing Feedback in Teaching Writing. LSP International Journal, 12(2), 133–142. https://doi.org/10.11113/lspi.v12.26003

[46] Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. Computers and Education: Artificial Intelligence, 7, 100270. https://doi.org/10.1016/j.caeai.2024.100270

[47] Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. Frontiers in Artificial Intelligence, 6, 1162454. https://doi.org/10.3389/frai.2023.1162454

[48] Liu, Z., Litman, D., Wang, E., Li, T., Gobat, M., Matsumura, L. C., & Correnti, R. (2025). eRevise+RF: A Writing Evaluation System for Assessing Student Essay Revisions and Providing Formative Feedback (No. arXiv:2501.00715). arXiv. https://doi.org/10.48550/arXiv.2501.00715

[49] Wilson, J., Huang, Y., Palermo, C., Beard, G., & MacArthur, C. A. (2021). Automated Feedback and Automated Scoring in the Elementary Grades: Usage, Attitudes, and Associations with Writing Outcomes in a Districtwide Implementation of MI Write. International Journal of Artificial Intelligence in Education, 31(2), 234–276. https://doi.org/10.1007/s40593-020-00236-w

[50] Aladini, A., Ismail, S. M., Ahmad Saleem Khasawneh, M., & Shakibaei, G. (2025). Self-directed writing development across computer/AI-based tasks: Unraveling the traces on L2 writing outcomes, growth mindfulness, and grammatical knowledge. Computers in Human Behavior Reports, 17, 100566. https://doi.org/10.1016/j.chbr.2024.100566

[51] Almusharraf, A., Bailey, D., Almusharraf, N., & Alotaibi, T. (2025). Students' perceptions of generative AI in EFL writing: Strategies, self-efficacy, satisfaction and behavioural intention. Australasian Journal of Educational Technology. https://doi.org/10.14742/ajet.10045