LLT Journal, e-ISSN 2579-9533, p-ISSN 1410-7201, Vol. 29, No. 2, October 2026, pp. 668-687

LLT Journal: A Journal on Language and Language Teaching
https://ejournal.usd.ac.id/index.php/LLT
Sanata Dharma University, Yogyakarta, Indonesia

GENERATIVE AI AND THE RESHAPING OF SECOND LANGUAGE WRITING: A COGNITIVE PROCESS PERSPECTIVE

Liubov Darzhinova[1*], Lianjiang George Jiang[2], and Mingyue Michelle Gu[3]
[1,2,3] The Education University of Hong Kong, China
liubovdarzh@eduhk.hk[1], jljiang@hku.hk[2], and mygu@eduhk.hk[3]
*correspondence: liubovdarzh@eduhk.hk
https://doi.org/10.24071/llt.v29i2.1359
received 9 June 2026; accepted 12 October 2026

Abstract
Addressing how Generative AI (GenAI) reshapes the cognitive act of second language (L2) writing, this conceptual article draws on recent empirical literature (2023–2026). It grounds its analysis in the Cognitive Process Theory of Writing (CPTW), chosen because its decomposition of planning, translating, and reviewing under working-memory limits illuminates how cognitive labor is redistributed between learner and machine. Through CPTW, three key insights emerge. First, planning undergoes an externalization of productive struggle: offloading idea generation causes learners to bypass essential rhetorical problem framing. Second, translating experiences a voice constraint: model optimization toward conventionality flattens non-standard L2 syntax and homogenizes authorial voice. Third, reviewing suffers a reviewing bifurcation, splitting evaluation into goal-level appraisal and opaque decision-level monitoring. Theoretically, GenAI transforms writing into a co-generative sociotechnical practice requiring GenAI feedback literacy. Pedagogically, educators should adopt process-oriented assessments, using pre-prompting framing logs, annotated drafts, and prompt histories to evaluate cognitive engagement over polished final products.

Keywords: chatbot, ChatGPT, generative AI, L2 writing

Introduction
Writing is a cognitive activity that enables the creation and communication of ideas, meaning, and knowledge through symbolic expression, and it is essential for academic and professional success. Technological advancements have continually reshaped writing, but the advent of Generative AI (GenAI) marks a transformation of a qualitatively different order (Facchinetti, 2021). GenAI does not merely mediate writing but executes drafting, revisions, and feedback provision (Ferrara, 2024). Documented benefits include access to an extensive well of knowledge, assistance with idea generation, timely feedback, and text-level operations such as translation, editing, and paraphrasing (Warschauer et al., 2023). Yet GenAI also produces highly polished, native-like text that exposes L2 learners to heightened scrutiny and the risk of suspicion of academic dishonesty, and its effectiveness in classrooms depends heavily on social norms, teachers' technologist identities, and institutional supports (Prihandoko et al., 2026).

This work is licensed under CC BY-SA.
Creative Commons Attribution-ShareAlike 4.0 International License

Scholarly attention has grown rapidly but unevenly. Previous reviews have examined GenAI in English language teaching and learning (Crompton et al., 2024), English as a foreign language (EFL) writing (Teng, 2025), and L2 learning in general (Yang & Li, 2024). More recent syntheses situate GenAI within the historical trajectory of AI in writing instruction (Feng et al., 2025) and map its affordances and learner attitudes (Yang et al., 2025). Despite these contributions, a critical gap remains: the conversation is predominantly technology-centered and outcome-oriented, privileging what GenAI tools do and produce over how they reconfigure the cognitive processes through which L2 writing is accomplished. Consequently, the fundamental question of how GenAI reshapes the act of L2 writing itself has not yet received adequate theoretical treatment.

Addressing this gap requires a framework capable of illuminating the cognitive processes that constitute writing. This article therefore adopts the cognitive process theory of writing (CPTW), drawing on Flower and Hayes (1981) and Hayes (2000), which, by decomposing writing into the recursive, interlocking processes of planning, translating, and reviewing under working-memory constraints, offers a finer-grained account of the cognitive work writers perform and thus the strongest analytical handle on how GenAI alters it. Alternative cognitive frameworks, while valuable, offer less traction at this level of detail. The knowledge-transforming model (Bereiter & Scardamalia, 1987) draws an important developmental distinction between novice knowledge-telling and expert knowledge-transforming, but it analyzes an overarching composing strategy rather than the real-time, sequential operations of text production, leaving it less able to differentiate among ceded, offloaded, and emergent co-creative processes; CPTW can represent all three, making visible the redistribution of cognitive labor across processes. Sociocultural perspectives, notably Vygotsky's (2012), explain the social, collaborative, and tool-mediated conditions under which writing occurs, but are less equipped to illuminate how such external conditions reconfigure the internal architecture of cognitive processes (Cong-Lem & Daneshfar, 2024).

The article first lays out the foundational tenets of CPTW, then systematically re-examines each stage of the writing process to account for GenAI's pervasive presence and influence, explores the consequences of these transformations, and concludes with limitations and promising avenues for research.

Theory

Core tenets of the cognitive process theory of writing

According to the cognitive process theory of writing (CPTW; Flower & Hayes, 1981), writing is a non-linear, goal-oriented cognitive activity in which multiple cognitive processes operate not as discrete, sequential steps but as a recursive network. The model identifies three principal macro-processes, specifically planning, translating, and reviewing, each placing distinct cognitive demands on the writer and each capable of being activated, interrupted, or re-entered at any point during composition. These processes are embedded in a task environment comprising the writing assignment and the text produced so far, and constrained by the writer's long-term memory, which stores topic knowledge, discourse schemata, and knowledge of the intended audience. The original model's monitor is a metacognitive executive function that governs movement among processes by tracking progress toward rhetorical goals and determining when to

shift from one to another (Flower & Hayes, 1981). Hayes's (1996) revision replaced the undifferentiated resource of long-term memory with a more differentiated cognitive system and, most significantly, introduced working memory, encompassing the phonological loop, the visuospatial sketchpad, and the central executive, as the central, capacity-limited workspace for planning, translating, and reviewing operations. Hayes (2000) further repositioned motivation and affect (e.g., writing apprehension, self-efficacy, and goal setting) as modulators of writing performance and situated cognitive processes within the physical and social dimensions of the writing context.

Central to CPTW is the construct of recursion. Rather than proceeding through a discrete sequence of plan-draft-revise, proficient writers move fluidly among all three macro-processes: a sentence drafted during translating may trigger renewed planning if a deeper conceptual problem is revealed, and a reviewing episode may lead to wholesale restructuring of rhetorical goals. This recursive interplay is not a sign of inefficiency but the generative engine of writing development and the primary mechanism through which writers achieve progressively deeper alignment between their rhetorical intentions and the emerging text (Hayes, 2012). The monitor is integral to this dynamic: it detects the misalignment and prompts a shift from translating back to planning, or signals when a reviewing pass has sufficiently resolved a local problem and composition may resume (Flower & Hayes, 1981).

Hayes's (1996) working-memory constraints qualify this picture of fluent recursion. Because all active operations compete for finite cognitive resources, inexperienced or overloaded L2 writers may find recursive movement disruptive. Cognitive Load Theory (CLT; Sweller et al., 1998, 2019) explains why this disruption can be especially acute for L2 writers. Writing carries a high intrinsic load: content generation, organization, vocabulary, and grammar must be managed simultaneously. Intrinsic load reflects element interactivity relative to prior knowledge, so it is effectively higher for L2 writers whose linguistic systems are not yet automatized (Sweller et al., 2019). Rather than retrieving language in consolidated chunks, L2 learners must process lexical retrieval and grammatical encoding as separate, interacting elements. Révész et al. (2017), for instance, found that even advanced L2 writers produced more frequent pausing and revision alongside less sophisticated lexical choices when content support was withdrawn. Revision is precisely the site at which such schemas are constructed and refined: diagnosing a mismatch and reintegrating the text draws on the working-memory resources devoted to schema construction (germane processing). Because intrinsic and germane processing draw on the same pool, capacity consumed by linguistic demands is capacity unavailable for schema construction. When linguistic demands saturate working memory, little residual capacity remains, and recursion becomes a cognitive tax rather than a developmental catalyst.

Planning: From idea generation to rhetorical problem representation

In the original CPTW, the planning stage entails three interrelated sub-processes: (1) generating ideas, (2) setting goals, and (3) organizing information. The writer is the originating locus of composing. Idea generation draws heavily on long-term memory, requiring the writer to retrieve and elaborate topic-relevant knowledge, associations, and conceptual schemata. Goal setting involves

constructing a hierarchical representation of communicative intentions: what to say, for whom, in what register, and to what rhetorical effect. Organizing entails selecting among possible arrangements of content and constructing a provisional discourse plan. These sub-processes jointly require writers to define the rhetorical problem before attempting to solve it, a step especially consequential for L2 writers, who may lack the automaticity in accessing L2 linguistic resources that would otherwise allow planning to proceed in tandem with other processes (Cumming, 2002).

Translating: Linguistic realization and the limits of automatization

Translating is the process by which the conceptual and rhetorical representations generated during planning are converted into linear, linguistically realized text. Flower and Hayes (1981) characterized translating as placing particularly heavy demands on working memory because writers must manage multiple simultaneous constraints: lexical selection (choosing words that are both accurate and appropriate), syntactic formulation (constructing grammatically well-formed and varied sentences), and rhetorical and semantic monitoring (ensuring that emerging text fulfils the communicative goals established during planning). For L2 writers, these demands are compounded by the relative lack of automatization of L2 linguistic knowledge, which elevates intrinsic load: lower-level processes such as word retrieval and grammatical encoding consume a disproportionate share of available working memory capacity, leaving less available for higher-level rhetorical monitoring (Révész et al., 2017).

Translating is also the site at which the writer's stylistic and rhetorical identity is most directly expressed in language: the particular word chosen over its synonym or the syntactic construction selected over its paraphrase is the material form of the writer's voice and authorial stance. When translating proceeds smoothly, writers can attend to both lower-level linguistic accuracy and higher-level rhetorical coherence. When it is effortful, lower-level demands may crowd out higher-level concerns, a trade-off CPTW treats as central to writing development (Flower & Hayes, 1981) and that CLT explains as competition between intrinsic and germane load for the same finite resources.

Reviewing: Metacognitive regulation and feedback processing

Reviewing encompasses both evaluating and revising. Evaluating involves reading the text produced so far and appraising it against the writer's goals, which requires metacognitive monitoring and the capacity to adopt a reader's perspective on one's own output. Revising, when triggered by evaluation, may operate at the surface level (correcting lexical or grammatical errors) or at the discourse level (restructuring arguments, reorganizing paragraphs, or introducing new content). Crucially, discourse-level revision functions as a recursive gateway back into planning, while surface-level revision may require re-entering translation for specific passages. This recursive quality means that reviewing serves as a regulatory mechanism within the broader writing system, mediating between the writer's intentions and the text's actual adequacy (Hayes, 2012).

For L2 writers, reviewing is further complicated by limited language proficiency, which constrains the ability to detect or diagnose specific types of textual inadequacy, particularly those related to pragmatic appropriateness,

discourse-level coherence, and misalignment of lexical choice with register. From a CPTW perspective, the metacognitive dimension of reviewing is arguably the most theoretically significant. It is the process by which writers actively adopt an external perspective, read their work through the eyes of an objective audience, spot mismatches between what they meant to say and what they actually wrote, and initiate targeted revisions. This capacity for self-regulatory metacognition, or the ability to evaluate, diagnose, and respond to one's own text, is what transforms reviewing from a corrective operation into a developmental one (Xu & Zhu, 2024).

Theory Application

This section applies CPTW (Flower & Hayes, 1981; Hayes, 2000) to formulate new conceptual arguments about GenAI's role in L2 writing, drawing on empirical evidence from extant research. A prior observation is that most GenAI tools do not map neatly onto individual writing stages: they can simultaneously generate ideas (planning), produce draft text (translating), and provide corrective feedback (reviewing) within a single exchange. Rather than a shortcoming of the framework, this constitutes an important theoretical observation. In the original model, recursive transitions between stages are initiated and governed by the writer (e.g., planning leads to translating, reviewing prompts re-planning), and the writer retains executive control over those transitions. When GenAI intervenes across stages, the writer's role as the primary orchestrator of recursion is partially displaced, which clarifies why a theoretical extension is necessary.

Corpus and selection criteria

Because the aim is theoretical re-interpretation rather than exhaustive synthesis, the empirical evidence discussed below is a purposively selected corpus, not the product of a systematic review. Candidate studies were located through Scopus using search parameters combining keywords (e.g., "generative AI," "ChatGPT," "LLM") with language learning contexts (e.g., "L2 writing," "EFL"). Studies were retained if they were (a) peer-reviewed journal articles in English published between 2023 and 2026, (b) focused on GenAI tools in L2/EFL writing, and (c) reported findings on writing process, outcomes, feedback, or learner perceptions that bore on planning, translating, or reviewing. Earlier reviews, research on adjacent technologies such as machine translation, foundational theoretical works, and technical literature on large language models (LLMs) training and bias are cited to situate the argument but are not part of the corpus. The corpus is therefore illustrative rather than representative.

Planning reconsidered: The externalization of productive struggle

Empirical evidence within the corpus commonly reports that GenAI tools now automate core components of the planning stage, including idea generation, outline drafting, and goal clarification, helping L2 learners overcome writer's block, maximize efficiency, and improve their drafts (Allen & Mizumoto, 2024; Mahapatra, 2024; Nugroho et al., 2025; Yan, 2023). From the perspective of efficient text production, such intervention appears unambiguously beneficial, since what was once a slow, effortful, and cognitively demanding internal process has become a rapid, externalized, and algorithmically supplied product. As a result, the locus of ideation has migrated, at least partially, from the cognitive interior of the

writer to GenAI output, unsettling the subject-centered epistemology upon which the CPTW rests. This shift is not merely procedural, but ontological.

When GenAI supplies ideas too readily, learners may bypass productive struggle, treating planning as a procedural task (select from among AI-generated options) rather than a constructive one (define and solve a novel rhetorical problem): a contrast that parallels Bereiter and Scardamalia's (1987) knowledge telling versus knowledge transforming. Yet, as per the original model (Flower & Hayes, 1981), planning requires productive struggle, including problem framing, rhetorical awareness, and the critical evaluation of competing ideas, and should therefore be regarded as a substantive, iterative component rather than a perfunctory preparatory step. Accordingly, planning should not be viewed as an obstacle to eliminate but as an essential developmental condition, since engaging with it supports the gradual emergence of metacognitive control, rhetorical sophistication, and other higher-order writing skills.

Beyond its cognitive dimensions, GenAI-mediated planning reshapes learners' affective orientation toward writing in potentially productive ways: performance and effort expectancy adoption intentions (Dong, 2024), and self-efficacy beliefs, collaborative writing tendencies, and clarity of writing goals are reported to strengthen through GenAI engagement (Dong, 2024; Fathi & Rahimi, 2026; Guo & Li, 2024; Teng, 2024). The affective architecture of planning is thus highly susceptible to GenAI influence. However, the evidence base (e.g., Ghafouri et al., 2024; Song & Song, 2023) is skewed toward short-term motivational outcomes, including reduced anxiety and increased self-efficacy, with limited attention to whether these gains translate into sustained improvements in independent planning competence. Central is whether learners internalize AI-generated ideas or merely consume them, and whether repeated reliance gradually erodes their capacity to represent rhetorical problems independently. Short-term motivational uplift may therefore mask a longer-term developmental cost that CPTW would predict but that outcome-oriented research is not equipped to detect.

In light of these considerations, a reimagined CPTW must incorporate what might be termed distributed planning: the allocation between writer and GenAI of two separable components of planning. Idea retrieval (generating and organizing content) can be shared at relatively low developmental cost. Problem framing (representing audience, purpose, and stance) is what CPTW treats as developmentally essential. What is distributed, and at what cost, plausibly varies with two characteristics that are easily conflated: L2 proficiency and rhetorical expertise. Proficiency governs how much working memory encoding consumes (Schoonen et al., 2003), so externalized retrieval relieves the load most for lower-proficiency writers. Rhetorical expertise, by contrast, determines whether the writer holds the criteria to frame a problem and judge GenAI framings. For low-proficiency, low-expertise writers, externalized retrieval can scaffold a text they could not otherwise produce, but they are also the most likely to adopt the first plausible GenAI framing by default. High-proficiency, high-expertise writers can interrogate GenAI output while keeping framing, but risk anchoring on early GenAI framings and narrowing the idea space unnoticed, because the output looks competent (cf. Doshi & Hauser, 2024). The mixed profiles are revealing: a fluent writer new to a genre gains little relief yet shares the novice's exposure to framing

delegation; a rhetorically expert writer with limited proficiency gains relief while retaining the criteria to frame, making distribution most plausibly benign.

Across profiles, distribution is developmentally benign only to the extent that problem framing stays with the writer. Scaffolds should therefore follow rhetorical expertise: writers low in expertise may need routines requiring them to specify audience, purpose, and stance before consulting GenAI, whereas those high in it may need prompts that elicit divergent alternatives or critique initial framings. This yields two testable propositions. First, the effect of GenAI-supported planning on later independent planning depends on whether the writer frames the problem before seeing GenAI framings, most strongly for low-expertise writers regardless of proficiency. Second, proficiency moderates short-term gains, with lower-proficiency writers benefiting most, which may mask framing-related costs.

Finally, the affective gains documented above may come at the cost of epistemic authorship, and the risk differs by profile: low-expertise writers who adopt GenAI framings by default may struggle to claim ideas they did not frame, whereas high-expertise writers may feel ownership intact while early GenAI framings have in fact shaped their ideas. The reimagined CPTW must therefore treat planning as a negotiated, iterative, and sometimes contested space in which learners engage GenAI outputs rather than passively consuming them.

Translating reconsidered: The paradox of surface-level enhancement

Concerning the translating stage, studies report improvements across multiple dimensions of L2 writing quality attributable to GenAI use, including syntactic complexity, lexical richness, accuracy, and fluency (Boudouaia et al., 2024; Tsai et al., 2024; Yu et al., 2024), as well as enhanced clarity and cohesion at the discourse level (Allen & Mizumoto, 2024). Extending these findings into multimodal territory, Liu et al. (2024) showed that GenAI fosters creativity by enabling learners to integrate text with images and other media. These findings jointly suggest that during the translating stage, GenAI tools function as a powerful compensatory mechanism that improves the surface quality of L2 learners' written output by supplementing or, in more intensive use cases, supplanting the linguistic encoding process. Yet this characterization sits in tension with what the translating stage demands theoretically: in CPTW, translating is not reducible to producing correct and fluent sentences, but a recursive engagement with meaning-making, rhetorical choice, and the gradual development of a writer's voice (Flower & Hayes, 1981).

The trade-off between surface improvement and cognitive depth is real, and it presents a paradox that a reimagined CPTW must address head-on. The efficiency with which GenAI augments translating performance raises the specter of what Zou and Huang (2024) identified as diminished authorial voice and potential learning loss. If linguistic encoding is outsourced to an algorithmic system, translating as a cognitive practice, through which L2 learners develop procedural and declarative knowledge of their target language, is rendered superfluous, or at least truncated. The paradox deepens when structural factors are considered, since limited access and lack of digital literacy complicate GenAI adoption and mean that its augmentative benefits are not uniformly distributed across learner populations, thereby compounding existing educational inequities rather than dissolving them (Warschauer et al., 2023).

Another closely related dimension of this paradox is homogenization. Because LLMs are pretrained on corpora that privilege dominant discourses and standard language varieties, GenAI-assisted texts tend to converge toward formulaic, standardized patterns that suppress linguistic diversity and cultural voice (Fleisig et al., 2024). The mechanism is likely more specific than corpus bias alone: contemporary LLMs are post-trained with reinforcement learning from human feedback (RLHF) and related preference-optimization methods that reward outputs judged as fluent, helpful, and conventional. Such training has been associated with reduced output diversity (Kirk et al., 2024), and LLMs have been shown to treat non-standard dialects less favorably than standard ones (Moon et al., 2025). We therefore hypothesize that when a model rewrites or extends learner text, non-standard syntax, including L2-influenced constructions, dialectal forms, and idiosyncratic rhetorical patterns, is disfavored relative to conventional alternatives. For L2 writers, this exerts a dual and conflicting pressure: it corrects learner errors and raises formal quality while potentially flattening the non-standard choices that create personal voice. Because the model is optimized for conventionality, it does not by default distinguish a genuine error from a stylistic choice. Writers can instruct it to preserve their voice, but doing so requires the metacognitive and prompting literacy that less experienced writers may lack.

Whether this model-level tendency translates into measurable convergence in actual L2 writing remains an open empirical question. Boudouaia et al. (2024) observed syntactic gains among EFL participants, but their design did not measure stylistic convergence and cannot distinguish genuine development from alignment with a standardized register. Direct investigation, for example, comparing the stylistic diversity of learner texts before and after GenAI revision, is needed.

These compounding dimensions of the paradox, that is, cognitive displacement, structural inequity, and homogenization, jointly expose the limits of incremental pedagogical fixes and demand a deeper theoretical response. When reimagining the translating stage, we need to distinguish between the lower and upper thresholds of writing quality. The corpus commonly reports that GenAI elevates the lower threshold by reducing surface errors and improving cohesion. There is a parallel countereffect: GenAI constrains the upper threshold of learner agency, critical thinking, and expressive range. One practically valuable response is the complementary-use approach advocated by Zou and Huang (2024) and Yu et al. (2024), in which learners produce independent drafts before using GenAI for refinement. This sequencing preserves at least some of the developmental affordances of unmediated translating and mitigates the risk of cognitive outsourcing.

Nevertheless, this approach addresses the paradox at the level of classroom practice without resolving the theoretical problem it poses for CPTW. The full weight of the paradox (cognitive, structural, and expressive) compels a reimagined CPTW to expand the translating stage beyond its original cognitive-individualist scope and accommodate a sociotechnical text production model. Such a model treats the written artifact not as the product of a single cognitive agent's encoding process but as the outcome of a collaborative and structurally conditioned interaction between human writers and GenAI systems.

This reconceptualization entails three theoretical obligations. First, the framework must theorize how GenAI transforms the developmental trajectory of

L2 learners' language acquisition, particularly whether algorithmically mediated practice can produce the same procedural and declarative gains as unmediated writing. Second, it must account for how GenAI reshapes learners' affective relationship to authorship, including their sense of ownership, investment, and identity as writers, given that voice and agency are progressively delegated and that model-level pressure toward standardized expression may work against the expression of a distinctive voice. Third, and most fundamentally, the framework must establish the conditions under which algorithmically augmented writing constitutes a genuine advance in communicative competence rather than a sophisticated simulation of it, since without this distinction the evaluative basis of L2 writing assessment is called into question.

Reviewing reconsidered: Bifurcation of the evaluative process

GenAI most profoundly reshapes the reviewing stage in L2 writing, and it is here that the explanatory limits of CPTW become most visible. Although fine-tuned GPT models demonstrate notable consistency in rubric-based grading (Jiang et al., 2023; Shin & Lee, 2024), several studies reported moderate reliability and inconsistencies with human scoring, cautioning against uncritical adoption in high-stakes assessment contexts (Bucol & Sangkawong, 2025; Shabara et al., 2024). These studies position GenAI as an assessor of finished texts; our concern here is different – the writer's own reviewing, in which GenAI becomes a participant in the writing process.

What the corpus overlooks, however, is a dimension of CPTW's own account of reviewing that makes the evaluative act itself theoretically legible – its metacognitive quality. In CPTW, reviewing is not a passive appraisal but a self-regulatory process in which writers read their emerging text against their developing intentions, drawing on memory traces of the compositional decisions that shaped it: why a particular argument was put forth, why one lexical choice prevailed over another. This epistemic access is constitutive of genuine reviewing. GenAI bifurcates the act into two levels: goal-level appraisal (whether the text reflects what the writer intended) and decision-level monitoring (why the text was realized this way and whether it could have been otherwise). GenAI leaves goal-level appraisal largely intact, but changes whose decisions are available for monitoring; how this occurs depends on the configuration, and the literature tends to collapse two distinct ones: reviewing AI-generated text, and evaluating GenAI feedback on text the writer has authored.

In the first configuration, the writer reviews text that GenAI produced. Writers who prompt GenAI retain memory traces of their communicative goals, that is, what they intended the output to accomplish and the arguments they planned to make, but lack access to the linguistic and rhetorical decisions through which those goals were realized: why GenAI structured an argument as it did, which discursive moves it selected, or how it resolved competing organizational possibilities. Reviewing here supports goal-level appraisal but cannot sustain the decision-level monitoring that CPTW presupposes, because there are no authorial decisions of the writer's own to monitor. The act resembles the reading of others' work more than self-regulatory revision, and its characteristic risk is accepting text on the strength of its fluency and its apparent fit with intentions, consistent with research on processing fluency as a cue to judgement (Alter & Oppenheimer, 2009)

and on complacency toward automated aids (Parasuraman & Manzey, 2010). Writers can therefore judge whether the text broadly reflects their goals yet remain structurally unable to scrutinize the compositional logic behind its production. Whether L2 writers reviewing AI-generated text show this pattern is an empirical question that the corpus does not address.

In the second configuration, the writer evaluates GenAI feedback on writing they authored. The feedback studies in the corpus bear on this configuration: GenAI delivers comprehensive, evenly distributed comments across content, organization, and language (Guo et al., 2024; Guo & Wang, 2024; Wiboolyasarin et al., 2024), though these studies document feedback provision rather than how writers evaluate it. The writer retains decision-level access to their own text, and knows why each choice was made, but lacks access to the reasoning behind GenAI's proposed revision: why a particular change was recommended, why one organizational move was preferred over another, or how competing rhetorical possibilities were weighed. GenAI often supplies explanations with its suggestions, but these are generated alongside the revision and need not reflect how it was actually produced, as unfaithful model explanations show (Turpin et al., 2023). The writer can check them for plausibility but not verify them against the system's decision process. Evaluation is therefore one-sided, requiring the writer to adjudicate between a transparent set of their own decisions and an opaque set of GenAI's. The empirical record bears this out: L2 writers apply GenAI feedback with limited scrutiny of its rhetorical appropriateness or fit with their communicative goals (Allen & Mizumoto, 2024). This need not reflect disengagement: when the rationale for a revision can be assessed only for plausibility, uncritical acceptance may be a rational response to altered epistemic conditions. The developmental consequences, however, differ across the two configurations: the first may undermine the writer's capacity to make compositional decisions, the second may threaten the internalization of the evaluative standards writers use to assess their own writing.

Accordingly, reviewing in GenAI-assisted writing should be reconceptualized as a bifurcated process. Reviewing GenAI output is not an extension of CPTW's reviewing construct but structurally external to it: it introduces an authorial agent whose decisions carry rhetorical consequence yet remain opaque, and for whose presence CPTW provides no analytical architecture. Evaluating GenAI feedback approximates CPTW's construct more closely, but the evaluative standard is partially externalized and unverifiable. Where writers move between both configurations, as in iterative drafting with GenAI, the resulting text is only partially authored by the writer, undermining the epistemic completeness that genuine reviewing presupposes. Both configurations demand substantial metacognitive resources, particularly feedback literacy and self-regulatory capacity, which L2 writers may not yet possess (Gozali et al., 2024; Tam, 2024).

Implications

Cognitive reconfigurations: Evaluative load, task environment, and recursive displacement

The transformations described for planning, translating, and reviewing converge to form three structural configurations of CPTW, as summarized in Figure 1.

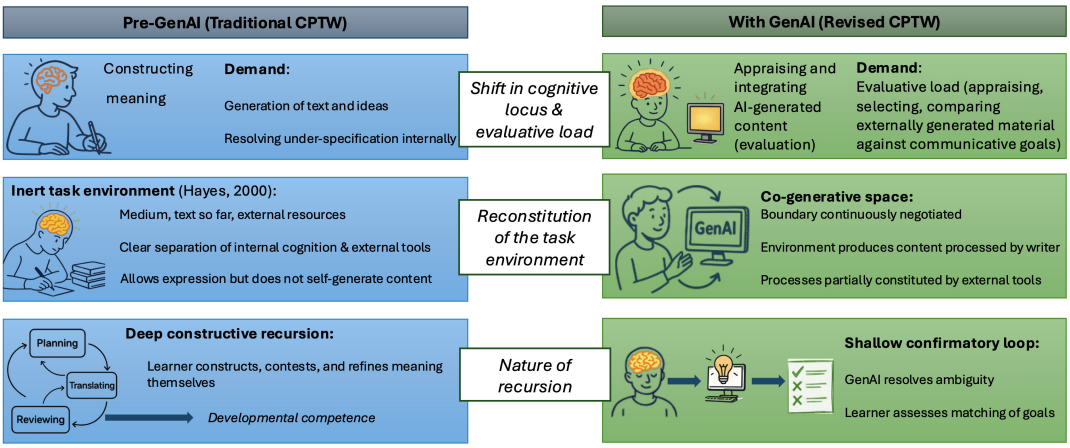


Figure 1. Applying CPTW to L2 writing: Pre-GenAI vs. with GenAI

GenAI shifts the locus of cognitive challenge from constructing ideas and text anew to evaluating externally produced material against one's own communicative goals. This shift stems from what can be conceptualized as distributed planning: while idea retrieval is offloaded to the machine, problem framing (representing audience, purpose, and stance) must remain anchored in the writer. Because a learner's capacity to evaluate GenAI framings depends on rhetorical expertise rather than L2 proficiency alone, delegating framing risks default acceptance of algorithmically supplied content. Fan et al. (2024) found that learners using GenAI spent less time generating ideas and more time evaluating AI-suggested content, indicating that attentional resources are redistributed rather than reduced. This demand constitutes an evaluative load – the cognitive effort needed to appraise, compare, and selectively integrate algorithmically generated material with one's goals. The developmental cost is that constructing ideas is replaced by evaluating them, shifting but not lightening cognitive effort.

GenAI also reconfigures the task environment. CPTW treats the task environment as largely inert, providing affordances and constraints without self-generating content (Hayes, 2000). Unlike a pen, dictionary, or word processor, a GenAI tool produces linguistic and ideational content, reconceptualizing the task environment as a co-generative, sociotechnical space in which the boundary between internal cognitive processes and external textual production is continuously negotiated. Within translation, this co-generation introduces a subtle risk: because modern LLMs are post-trained with RLHF, which reduces output diversity (Kirk et al., 2024), AI-mediated text formulation may flatten non-standard L2 syntax and homogenize stylistic voice.

This co-generation also alters recursion. When GenAI supplies outlines and drafts, the planning-translating loop can collapse into a shallow confirmatory cycle where learners check whether the AI-generated text matches their goals rather than constructing, contesting, and refining meaning themselves. We term this recursive displacement. Evaluating a ready-made draft requires less cognitive effort than constructing one from underspecified goals, reducing opportunities for developmental learning.

Pedagogical implications: Restoring evaluative and recursive engagement

The risks of confirmatory processing and stylistic homogenization necessitate assignment designs that preserve recursive engagement and authorial agency rather than displace them. Protocols should require learners to interrogate, contest, and, where appropriate, reject AI-generated content. For planning, learners (especially those low in rhetorical expertise) need upfront problem-framing tasks that mandate defining audience, purpose, and stance prior to consulting GenAI, recorded in an initial framing log; they can then engage in divergent alternative prompting, requesting several rhetorically distinct framings or drafts and justifying which, if any, they adopt.

Because the reviewing bifurcation is structural rather than dispositional, uncritical acceptance reflects the absence of a stable decision-level referent; exhortations to engage more critically with GenAI suggestions are insufficient. What is required are designs that reconstruct access to decision-level logic: instructor-led modeling that makes the rhetorical reasoning behind revisions explicit; comparative draft analysis in which learners examine how competing organizational and lexical choices serve a communicative purpose and audit whether GenAI revisions flatten authentic authorial voice; and structured interrogation protocols requiring learners to articulate why a particular discursive move was preferable to the alternatives GenAI did not select.

These designs reconstitute the epistemic conditions for genuine decision-level reviewing and are supported by sociocultural theories emphasizing interaction and meaning negotiation (Poehner & Lantolf, 2022). Collaborative drafting and structured dialogue around AI-generated text, consonant with Socratic seminar principles, require learners to articulate and defend their positions, refine their arguments in real time, and heighten metacognitive awareness of their own rhetorical choices, cultivating the feedback literacy suited to the GenAI context.

Assessment validity: Distinguishing competence from output

These pedagogical concerns carry direct consequences for assessment validity. Learners who produce a polished text following GenAI feedback may have engaged only in goal-level appraisal, confirming that the text reflects their intentions without interrogating the compositional logic behind it. From the perspective of the reviewing bifurcation, they have not fully engaged in reviewing as a self-regulatory process, yet their product may be indistinguishable from one emerging from genuine decision-level engagement. Product-only evaluation cannot detect whether the developmental substance of reviewing is there at all.

This gap reflects a conflation L2 writing scholarship must keep analytically distinct: an instrumentalist view, which treats GenAI as a performance enhancer and evaluates outcomes accordingly, and a developmental view, which requires evidence of internalized skill growth rather than textual improvement alone. If the construct that L2 writing assessment measures includes the self-regulatory competencies reviewing is meant to develop, product-only evaluation systematically misrepresents learning (Wiboolyasarin et al., 2026).

Addressing this requires assessment designs that make the process visible (Table 1): revision process logs and prompt histories document whether learners' goals preceded and shaped their prompts; annotated drafts and metalinguistic commentary tasks externalize the decision-level engagement GenAI-assisted

reviewing leaves structurally implicit; and stylistic divergence tracking, that is, comparing successive drafts against the GenAI output and recording what learners retained, changed, or rejected, makes visible whether authorial voice survived co-generation. Recent scholarship (Lim, 2026) similarly contends that teachers need to prioritize multimodal, process-oriented tasks over judging final products.

Table 1. Mapping re-conceptualized CPTW constructs to instruction and assessment

Re-conceptualized Construct	**Core risk**	**Actionable Classroom Tasks**	**Assessment Evidence (what makes it visible)**
Evaluative load (including distributed planning)	Default acceptance of GenAI framing and content	Upfront problem-framing protocols (audience, purpose, stance) before consulting GenAI; divergent alternative prompting	Initial framing logs; comparative prompt histories (whether goals preceded and shaped prompts)
Co-generative task environment	Stylistic homogenization (RLHF flattening)	Comparative draft analysis auditing voice; Socratic seminars; collaborative drafting	Metalinguistic commentary tasks; stylistic divergence tracking (what the writer retained, changed, or rejected)
Recursive displacement	Shallow confirmatory cycles; no decision-level referent	Instructor-led modeling of rhetorical reasoning; structured interrogation and contestation protocols	Revision process logs; annotated drafts with decision-level rationale

Feedback literacy as a prerequisite competency

Both the pedagogical and assessment implications presuppose a form of feedback literacy that existing frameworks have not fully theorized. Feedback literacy, as traditionally conceived, involves the capacity to understand, evaluate, and act on feedback to support learning rather than merely improve a single text (Carless & Boud, 2018). That conception was developed for human-generated feedback, where rhetorical reasoning could be made accessible through dialogue or scaffolding. GenAI feedback is structurally different: the decision logic through which it generates suggestions is constitutively opaque, rather than merely withheld. GenAI-oriented feedback literacy is therefore not a more demanding version of the conventional competency but a categorically distinct one. It requires learners to reason about the plausible decision space from which choices were drawn, asking not only whether a revision improves the text, but what rhetorical values it encodes, whose discursive norms it reflects, and what communicative possibilities it forecloses.

This demand is compounded by feedback volume. Because GenAI generates abundant formative commentary, the limiting factor shifts from access to quantity management. Learners confronting high volumes of suggestions risk treating them as instructions to be followed rather than invitations to reason, forfeiting the

metacognitive engagement that makes revision educationally productive (Vinall et al., 2023). Whether repeated engagement cultivates this capacity or habituates learners to externalize evaluative judgment to the machine remains an open question, one best addressed by the process-oriented methodologies recommended below.

Conclusion

This conceptual paper explored how GenAI reshapes the act of L2 writing by re-reading the empirical record through the CPTW framework and extending it with theoretical resources from adjacent domains. It makes three theoretical contributions. The first is conceptual: by centering the recursive processes of planning, translating, and reviewing as the unit of analysis, the paper offers a lens through which GenAI's effects can be examined not merely in terms of what writers produce but in terms of how the act of writing as a cognitive process is constituted and developed. The second is diagnostic: the analysis surfaces unresolved tensions in current research, including the conflation of process and product and the tendency to treat pedagogical risks as dispositional rather than structural. The third entails normative guidance: by positioning GenAI not as another entry in the lineage of educational technologies but as a force that reconfigures the cognitive conditions under which writing competence is built, the article argues that the field's primary obligation has shifted. The challenge of documenting what GenAI can do has largely run its course; the task is now to theorize what GenAI should do – how it should be designed and deployed to cultivate linguistically proficient and epistemically responsible L2 writers.

The reframing also bears methodological implications. The most fundamental gap in the literature is its predominant reliance on product-focused, cross-sectional designs: learner texts and self-report data remain the dominant data sources, and the cognitive processes through which GenAI-assisted writing unfolds are largely undocumented. Tackling this gap demands temporally sensitive, process-centric methodologies: keystroke logging, think-aloud protocols, screen capture, and naturalistic observation (Séror & Gentil, 2023). Accessible tools make this feasible. Inputlog (Leijten & Van Waes, 2013), freely available for research, records typing bursts, pauses, revisions, and window switching that expose planning, translating, and reviewing episodes, while open-source software such as OBS Studio (Lin et al., 2026) can capture how learners prompt, evaluate, and revise GenAI output and can anchor think-aloud or stimulated-recall sessions. More ethnographic and longitudinal studies are also necessary to examine whether engagement with GenAI cultivates transferable competencies that persist beyond the immediate presence of the tool. The available evidence from adjacent domains offers limited cause for optimism: research on neural machine translation demonstrates that while learners gain short-term benefits from automated tools, these advantages disappear on delayed post-tests after tool access is removed (O'Neill, 2019). Whether the same pattern of transient benefits applies to GenAI-assisted L2 writing remains an open empirical question, and one that process-oriented longitudinal research is uniquely suited to address.

Another methodological priority is more systematic reporting of L2 proficiency, as the cognitive demands of interacting with GenAI may vary across ability levels. Many of the studies reviewed do not explicitly document participants'

proficiency levels, undermining cross-study comparability and diminishing the translational value of the research for educators. Future studies should report proficiency against an externally referenced scale (e.g., CEFR levels), anchored by a named and validated measure, and, where sample size permits, analyze process and outcome data by proficiency band.

In closing, several limitations warrant acknowledgement. This article draws on scholarship published in English, so research conducted in other linguistic traditions falls outside its purview. Because GenAI tools and the research documenting their effects advance rapidly, any synthesis of this kind represents a snapshot. Living systematic reviews, adapted from health research, offer one model for maintaining currency as the field evolves.

Acknowledgments

This research is supported by JC_AI research fund (Project number: 02186), funded by the Education University of Hong Kong and the Hong Kong University of Science and Technology, Hong Kong SAR, China.

References

Allen, T. J., & Mizumoto, A. (2024). ChatGPT over my friends: Japanese English-as-a-foreign-language learners' preferences for editing and proofreading strategies. *RELC Journal, 57*(1), 89-106. https://doi.org/10.1177/00336882241262533

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219-235. https://doi.org/10.1177/1088868309341564

Bereiter, C., & Scardamalia, M. (Eds.). (1987). *The psychology of written composition* (1st ed.). Routledge. https://doi.org/10.4324/9780203812310

Boudouaia, A., Mouas, S., & Kouider, B. (2024). A study on ChatGPT-4 as an innovative approach to enhancing English as a foreign language writing learning. *Journal of Educational Computing Research, 62*(6), 1289-1317. https://doi.org/10.1177/07356331241247465

Bucol, J. L., & Sangkawong, N. (2025). Exploring ChatGPT as a writing assessment tool. *Innovations in Education and Teaching International, 62*(3), 867-882. https://doi.org/10.1080/14703297.2024.2363901

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Cong-Lem, N., & Daneshfar, S. (2024). Generative AI and second/foreign language education from Vygotsky's cultural-historical perspective. In H.P. Bui & E. Namaziandost (Eds.), *Innovations in technologies for language teaching and learning* (Vol. 1159, pp. 175- 188). Springer Nature. https://doi.org/10.1007/978-3-031-63447-5_10

Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology, 55*(6), 2503-2529. https://doi.org/10.1111/bjet.13460

Cumming, A. (2002). Assessing L2 writing: Alternative constructs and ethical dilemmas. *Assessing Writing*, *8*(2), 73–83. https://doi.org/10.1016/S1075-2935(02)00047-8

Dong, L. (2024). "Brave new world" or not? A mixed-methods study of the relationship between second language writing learners' perceptions of ChatGPT, behaviors of using ChatGPT, and writing proficiency. *Current Psychology, 43,* 19581-19495. https://doi.org/10.1007/s12144-024-05728-9

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, *10*(28), Article eadn5290. https://doi.org/10.1126/sciadv.adn5290

Facchinetti, R. (2021). News discourse and the dissemination of knowledge and perspective: From print and monomodal to digital and multisemiotic. *Journal of Pragmatics, 175*, 195–206. https://doi.org/10.1016/j.pragma.2021.01.015

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2024). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489-530. https://doi.org/10.1111/bjet.13544

Fathi, J., & Rahimi, M. (2026). Utilising artificial intelligence-enhanced writing mediation to develop academic writing skills in EFL learners: A qualitative study. *Computer Assisted Language Learning, 39*(1-2), 263-302. https://doi.org/10.1080/09588221.2024.2374772

Feng, H., Li, K., & Zhang, L. J. (2025). What does AI bring to second language writing? A systematic review (2014-2024). *Language Learning & Technology, 29*(1), 1-27. https://doi.org/10.64152/10125/73629

Ferrara, E. (2024). GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science, 7,* 549-569. https://doi.org/10.1007/s42001-024-00250-1

Fleisig, E., Smith, G., Bossi, M., Rustagi, I., Yin, X., & Klein, D. (2024). Linguistic bias in ChatGPT: Language models reinforce dialect discrimination. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 13541–13564. https://doi.org/10.18653/v1/2024.emnlp-main.750

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. https://doi.org/10.2307/356600

Ghafouri, M., Hassaskhah, J., & Mahdavi-Zafarghandi, A. (2024). From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Language Teaching Research*. https://doi.org/10.1177/13621688241239764

Gozali, I., Wijaya, A. R. T., Lie, A., Cahyono, B. Y., & Suryati, N. (2024). Leveraging the potential of ChatGPT as an automated writing evaluation (AWE) tool: Students' feedback literacy development and AWE tools integration framework. *The JALT CALL Journal, 20*(1), 1–22. https://doi.org/10.29140/jaltcall.v20n1.1200

Guo, K., & Li, D. (2024). Understanding EFL students' use of self-made AI chatbots as personalized writing assistance tools: A mixed methods study. *System, 124,* Article 103362. https://doi.org/10.1016/j.system.2024.103362

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29*(7), 8435-8463. https://doi.org/10.1007/s10639-023-12146-0

Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. *The Internet and Higher Education, 63*, Article 100962. https://doi.org/10.1016/j.iheduc.2024.100962

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 1–27). Lawrence Erlbaum Associates.

Hayes, J. R. (2000). A new framework for understanding cognition and affect in writing. In R. Indrisano & J. R. Squire (Eds.), *Perspectives on writing: Research, theory, and practice* (pp. 6–44). International Reading Association.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, *29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Jiang, Z., Xu, Z., Pan, Z., He, J., & Xie, K. (2023). Exploring the role of artificial intelligence in facilitating assessment of writing performance in second language learning. *Languages, 8*(4), Article 247. https://doi.org/10.3390/languages8040247

Kirk, R., Mediratta, I., Nalmpantis, C., Luketina, J., Hambro, E., Grefenstette, E., & Raileanu, R. (2024). Understanding the effects of RLHF on LLM generalisation and diversity. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, & Y. Sun (Eds.), *International Conference on Learning Representations* (pp. 1-34). https://proceedings.iclr.cc/paper_files/paper/2024/file/5a68d05006d5b05dd9463dd9c0219db0-Paper-Conference.pdf

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using inputlog to analyze and visualize writing processes. *Written Communication*, *30*(3), 358–392. https://doi.org/10.1177/0741088313491692

Lim, F. V. (2026). Designing literacy learning with multimodality in the age of Generative AI. *LLT Journal: A Journal on Language and Language Teaching*, *29*(1), 1-17. https://doi.org/10.24071/llt.v29i1.694

Lin, H., Wei, W., Zou, B., & Zheng, Y. (2026). Interactive reading patterns and metacognitive strategies from surface to deep levels of L2 learners in GenAI-assisted critical reading. *Discover Computing, 29*, Article 187. https://doi.org/10.1007/s10791-026-10069-1

Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Computers & Education, 211*, 1–21. https://doi.org/10.1016/j.compedu.2023.104977

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments, 11*(9), 1–18. https://doi.org/10.1186/s40561-024-00295-9

Moon, K., Green, A. E., & Kushlev, K. (2025). Homogenizing effect of large language models (LLMs) on creative diversity: An empirical comparison of human and ChatGPT writing. *Computers in Human Behavior: Artificial Humans*, *6*, 1-10. https://doi.org/10.1016/j.chbah.2025.100207

Nugroho, A., Andriyanti, E., Widodo, P., & Mutiaraningrum, I. (2025). Students' appraisals post-ChatGPT use: Students' narrative after using ChatGPT for writing. *Innovations in Education and Teaching International, 62*(2), 499–511. https://doi.org/10.1080/14703297.2024.2319184

O'Neill, E. M. (2019). Training students to use online translators and dictionaries: The impact on second language writing scores. *International Journal of Research Studies in Language Learning, 8*(2), 47–65. https://doi.org/10.5861/ijrsll.2019.4002

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Poehner, M. E., & Lantolf, J. P. (2022). Sociocultural theory and second language development. In C. Coombe & H. Mohebbi (Eds.), *Research questions in language education and applied linguistics* (pp. 871–877). Springer Nature. https://doi.org/10.1007/978-3-030-79143-8_150

Prihandoko, L.A., Amalia, S.N., Ardi, P., & Waluyo, B. (2026). Applying AI for English Language instruction and material development in schools: A PLS-SEM approach. *International Journal of TESOL Studies*, *8*(2), 49–72. https://doi.org/10.58304/ijts.251020

Révész, A., Michel, M., & Lee, M.J. (2017). *Investigating IELTS academic writing task 2: Relationships between cognitive writing processes, text quality, and working memory*. (IELTS Research Report Online Series). British Council, Cambridge English Language Assessment and IDP.

Schoonen, R., Gelderen, A. V., Glopper, K. D., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning*, *53*(1), 165–202. https://doi.org/10.1111/1467-9922.00213

Séror, J., & Gentil, G. (2023). Direct observation of writing activity: Screen capture technologies. In R. M. Manchón & J. Roca de Larios (Eds.), *Research methods in the study of L2 writing processes* (Vol. 5, pp. 141–160). John Benjamins Publishing Company. https://doi.org/10.1075/rmal.5.07ser

Shabara, R., ElEbyary, K., & Boraie, D. (2024). Teachers or ChatGPT: The issue of accuracy and consistency in L2 assessment. *Teaching English with Technology, 24*(2), 71–92. https://doi.org/10.56297/vaca6841/LRDX3699/XSEZ5215

Shin, D., & Lee, J. H. (2024). Exploratory study on the potential of ChatGPT as a rater of second language writing. *Education and Information Technologies, 29,* 24735-24757. https://doi.org/10.1007/s10639-024-12817-6

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14,* 1–14. https://doi.org/10.3389/fpsyg.2023.1260843

Sweller, J., Van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*, 261–292. https://doi.org/10.1007/s10648-019-09465-5

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296. https://doi.org/10.1023/A:1022193728205

Tam, A. C. F. (2024). Interacting with ChatGPT for internal feedback and factors affecting feedback quality. *Assessment & Evaluation in Higher Education, 50*(2)*,* 219-235. https://doi.org/10.1080/02602938.2024.2374485

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,* 1–10. https://doi.org/10.1016/j.caeai.2024.100270

Teng, M. F. (2025). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies, 6*(3), 36–57. https://doi.org/10.58304/ijts.20240304

Tsai, C.-Y., Lin, Y.-T., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies*, *29*, 22427–22445. https://doi.org/10.1007/s10639-024-12722-y

Turpin, M., Michael, J., Perez, E., & Bowman, S. (2023). Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. *Advances in Neural Information Processing Systems, 36*, 74952–74965. https://doi.org/10.52202/075280-3275

Vinall, K., Wen, W., & Hellmich, E. A. (2023). Investigating L2 writers' uses of machine translation and other online tools. *Foreign Language Annals, 57*(2), 499–526. https://doi.org/10.1111/flan.12733

Vygotsky, L. S. (2012). *Thought and language, revised and expanded edition.* MIT Press.

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing, 62*, 1–7. https://doi.org/10.1016/j.jslw.2023.101071

Wiboolyasarin, W., Chomchuen, F., Kamonsawad, R., Boonprawes, S., Wiboolyasarin, K., & Jinowat, N. (2026). Timing is everything: Investigating the impact of immediate, delayed, and combined AI-generated feedback on L2 writing proficiency. *Social Sciences & Humanities Open*, *13*, 1-13. https://doi.org/10.1016/j.ssaho.2026.102790

Wiboolyasarin, W., Wiboolyasarin, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. *Computers and Education: Artificial Intelligence, 6*, 1–10. https://doi.org/10.1016/j.caeai.2024.100228

Xu, W., & Zhu, X. (2024). Examining metacognitive strategy use in L1 and L2 task-situated writing: Effects, transferability, and cross-language facilitation. *Metacognition and Learning, 19*, 773–792. https://doi.org/10.1007/s11409-024-09387-w

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, *28*, 13943–13967. https://doi.org/10.1007/s10639-023-11742-4

Yang, C., Li, R., & Yang, L. (2025). Revisiting trends in GenAI-assisted second language writing: Retrospect and prospect. *Journal of Educational Computing Research, 63*(7-8), 1819-1863. https://doi.org/10.1177/07356331251367309

Yang, L., & Li, R. (2024). ChatGPT for L2 learning: Current status and implications. *System, 124,* Article 103351. https://doi.org/10.1016/j.system.2024.103351

Yu, L., Kong, Q., & Hao, H. (2024). Artificial intelligence-assisted Chinese L2 writing: An empirical study on educational sustainability in Africa. *Rupkatha Journal, 16*(2), 1–15. https://doi.org/10.21659/rupkatha.v16n2.01

Zou, M., & Huang, L. (2024). The impact of ChatGPT on L2 writing and expected responses: Voice from doctoral students. *Education and Information Technologies, 29,* 13201-13219. https://doi.org/10.1007/s10639-023-12397-x