



Teacher-Guided Generative AI for Sustainable EFL Writing Development: Longitudinal Validation of the AI-MWD Framework

Tugsuu Jargal

The Department of Foreign Languages
Mandakh University
Ulaanbaatar, Mongolia
tugsuu@mandakh.edu.mn

Sainzaya Sumiya

The Department of Foreign Languages
Mandakh University
Ulaanbaatar, Mongolia
sainzaya@mandakh.edu.mn

Abstract—The rapid adoption of generative artificial intelligence (GenAI) has transformed English as a Foreign Language (EFL) writing instruction by providing immediate, individualized, and context-sensitive feedback. Despite these advantages, increasing evidence suggests that unguided interaction with GenAI often overwhelms learners' cognitive capacity, encourages passive dependence on AI-generated revisions, and weakens the development of long-term writing autonomy. This study addresses this challenge by proposing and empirically validating the AI-Based Metacognitive Writing Development (AI-MWD) Framework, a teacher-guided instructional framework that integrates macro-pedagogical scaffolding with AI-enabled micro-scaffolding to promote sustainable writing development. A 14-week longitudinal quasi-experimental study was conducted with forty undergraduate EFL learners assigned to either an experimental group receiving the AI-MWD intervention or a control group using the same GenAI tools without structured teacher guidance. Independent-samples t-tests examined differences in cognitive load and metacognitive writing autonomy, while 2×3 mixed-design ANOVAs evaluated longitudinal changes in Complexity, Accuracy, and Fluency (CAF) across pre-test, post-test, and delayed post-test measures. The findings revealed that teacher-guided AI significantly reduced learners' extraneous cognitive load while substantially enhancing metacognitive writing autonomy. Although both groups demonstrated comparable improvements in writing accuracy immediately after the intervention, only the AI-MWD group maintained these gains six weeks later, indicating significantly stronger long-term retention of linguistic knowledge. Unlike previous AI-assisted writing studies that primarily report immediate performance improvements, this study demonstrates that teacher-guided scaffold fading is the critical mechanism through which AI-supported feedback becomes internalized as learners' metacognitive writing strategies. The proposed AI-MWD Framework therefore reconceptualizes generative AI not as an autonomous instructional agent, but as a pedagogically mediated cognitive scaffold capable of supporting sustainable second-language writing development.

Keywords—Generative Artificial Intelligence, Teacher-Guided Instruction, Metacognitive Writing, Cognitive Load, Scaffolding, Writing Autonomy, Longitudinal Learning, EFL Writing

I. INTRODUCTION

The rapid emergence of generative artificial intelligence (GenAI) has fundamentally reshaped second-language writing instruction. Large language models (LLMs), such as ChatGPT and similar AI-powered writing assistants, are increasingly integrated into English as a Foreign Language (EFL) classroom because they provide immediate feedback, language suggestions, idea generation, and individualized writing support [1], [2]. These capabilities have created unprecedented opportunities for improving writing quality, increasing learner engagement, and expanding access to personalized learning. Consequently, AI-assisted writing has become one of the fastest-growing research areas in technology-enhanced language learning.

Despite these advantages, recent studies have raised an important pedagogical concern. While GenAI effectively improves immediate writing performance, its educational value remains uncertain when learners rely on AI-generated revisions without sufficient cognitive engagement. Instead of actively planning, evaluating, and revising their own texts, many learners simply accept AI suggestions with minimal reflection [3]. Such passive dependence may reduce productive cognitive effort, weaken metacognitive regulation, and ultimately limit learners' ability to write independently once AI assistance is removed. This emerging tension between short-term performance gains and long-term learner development has become a central issue in contemporary AI-supported language education.

From a cognitive perspective, this problem can be understood through Cognitive Load Theory [4], which argues that meaningful learning occurs only when learners' limited working-memory resources are allocated to productive

cognitive processing rather than unnecessary mental effort. Although GenAI can reduce the mechanical burden of writing, it may simultaneously increase extraneous cognitive load by presenting excessive feedback, multiple alternatives, and continuous decision-making demands. Without appropriate instructional guidance, learners may experience cognitive overload or develop superficial editing behaviors that improve texts without strengthening underlying writing competence.

Equally important, sociocultural perspectives suggest that learning is fundamentally mediated through guided interaction rather than technology alone. According to Vygotsky's Sociocultural Theory [5], cognitive development occurs through scaffolded support within learners' Zone of Proximal Development, while Bruner's Scaffolding Theory [6] emphasizes the gradual transfer of responsibility from teacher to learner. These principles remain highly relevant in AI-supported learning environments. Rather than replacing teachers, generative AI should function as an instructional resource whose educational value depends on how teachers deliberately design, regulate, and gradually withdraw pedagogical support. From this perspective, AI is not an autonomous instructor but a cognitive tool whose effectiveness is determined by purposeful teacher mediation.

Although an increasing number of studies have investigated AI-assisted writing, three important research gaps remain. First, most studies examine the effectiveness of AI tools themselves while paying comparatively little attention to the instructional processes that shape learners' interaction with AI. Second, previous research primarily reports immediate post-intervention improvements, providing limited evidence regarding whether AI-supported learning leads to durable writing development after instructional support is withdrawn. Third, relatively few studies integrate cognitive, metacognitive, and sociocultural perspectives into a unified instructional framework capable of explaining how teacher guidance transforms AI-generated feedback into sustainable writing competence.

To address these limitations, this study proposes the AI-Based Metacognitive Writing Development (AI-MWD) Framework, a teacher-guided instructional framework that combines macro-level pedagogical scaffolding with AI-enabled micro-level feedback. The framework conceptualizes teachers as designers of learners' cognitive and metacognitive development, while generative AI serves as a dynamic scaffold that supports planning, drafting, reflection, and revision. Through structured scaffold fading and gradual release of responsibility, learners progressively internalize AI-supported writing strategies and develop greater metacognitive autonomy.

The framework was evaluated through a 14-week longitudinal quasi-experimental study involving undergraduate EFL learners. Independent-samples t-tests and mixed-design analyses of variance (ANOVA) were employed to examine changes in cognitive load, metacognitive writing autonomy, and the longitudinal development of writing complexity, accuracy, and fluency (CAF). By incorporating both immediate and delayed post-test measures, the study investigates not only whether teacher-guided AI improves writing performance but also whether these improvements are sustained beyond the intervention period.

This study makes four principal contributions to AI-assisted language learning research. First, it introduces the AI-

MWD Framework as an integrated instructional model that synthesizes sociocultural theory, cognitive load regulation, metacognitive strategy instruction, and teacher-guided generative AI. Second, it reconceptualizes generative AI as a pedagogically mediated cognitive scaffold rather than an autonomous instructional agent. Third, it provides longitudinal empirical evidence demonstrating the role of scaffold fading in sustaining writing development after AI-supported instruction. Finally, it offers practical guidance for educators seeking to integrate generative AI into EFL writing instruction while preserving learner autonomy and long-term cognitive growth.

II. AI-MWD FRAMEWORK

Recent advances in generative artificial intelligence have substantially expanded opportunities for technology-enhanced writing instruction. However, most existing instructional approaches position AI primarily as an intelligent feedback provider, assuming that improved feedback naturally results in improved learning. This assumption overlooks an important pedagogical principle: learning is determined not only by the quality of feedback but also by how learners cognitively engage with that feedback.

The AI-Based Metacognitive Writing Development (AI-MWD) Framework was developed to address this limitation by integrating teacher-guided pedagogy with AI-supported learning within a unified instructional system. Rather than treating generative AI as an autonomous instructional agent, the framework conceptualizes AI as a dynamic cognitive scaffold whose effectiveness depends on purposeful teacher mediation. The framework therefore distinguishes between two complementary layers of instructional support.

At the macro-pedagogical level, teachers design learning objectives, regulate learners' cognitive engagement, provide explicit metacognitive strategy instruction, and gradually reduce instructional support following the principle of scaffold fading. This instructional process is informed by sociocultural learning theory, in which guidance progressively shifts from teacher regulation toward learner self-regulation.

At the micro-scaffolding level, generative AI provides immediate assistance throughout the writing process by supporting idea generation, language refinement, organization, revision, and reflective questioning. Unlike conventional automated feedback systems, AI assistance is deliberately embedded within teacher-designed learning tasks rather than functioning as an independent source of instruction. Consequently, AI feedback becomes a tool for guided cognitive processing rather than a substitute for learners' own thinking.

The interaction between these two instructional layers is regulated through continuous metacognitive reflection. Learners are encouraged to evaluate AI-generated suggestions, justify revision decisions, monitor their writing processes, and gradually assume greater responsibility for planning and self-assessment. As instructional support is systematically withdrawn, learners progressively internalize effective writing strategies while reducing dependence on external AI assistance.

From a theoretical perspective, the AI-MWD Framework synthesizes four complementary educational principles. First, sociocultural theory explains the importance of guided learning and mediated cognitive development. Second,

scaffolding theory provides the mechanism for gradually transferring responsibility from teacher to learner. Third, Cognitive Load Theory [4] supports instructional decisions that minimize unnecessary cognitive demands while preserving productive mental effort. Finally, metacognitive theory [7] explains how repeated planning, monitoring, and evaluation promote the internalization of effective writing strategies and long-term learner autonomy.

Accordingly, the AI-MWD Framework conceptualizes sustainable AI-assisted writing development as a progressive transformation from externally regulated AI-supported performance to internally regulated writing competence. In this process, teacher guidance functions as the pedagogical driver, generative AI operates as a temporary cognitive scaffold, and metacognitive regulation serves as the mechanism through which external support is gradually transformed into independent writing expertise.

Figure 1 illustrates the conceptual architecture of the proposed AI-MWD Framework.

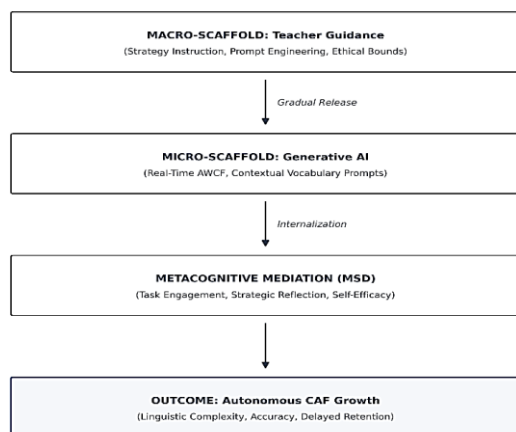


Fig. 1. Conceptual architecture of the AI-MWD Framework, integrating macro-pedagogical scaffolding with AI-enabled micro-scaffolding through continuous metacognitive regulation.

III. METHODOLOGY

A. Research Design

This study employed a 14-week longitudinal quasi-experimental design to investigate the effectiveness of the AI-Based Metacognitive Writing Development (AI-MWD) Framework in promoting sustainable EFL writing development. The research compared a teacher-guided AI intervention with conventional AI-assisted writing instruction to determine whether structured pedagogical mediation produced superior cognitive and linguistic outcomes over time.

A 2×3 mixed-design was adopted. The between-subject factor consisted of two instructional conditions (experimental and control groups), while the within-subject factor represented three measurement occasions (pre-test, post-test, and delayed post-test). This design enabled the investigation of both immediate instructional effects and the long-term retention of writing development after the intervention.

B. Participants

Forty undergraduate EFL students enrolled in an Academic English writing course at a Mongolian university voluntarily participated in the study. Participants were selected through intact-class sampling because institutional

scheduling constraints prevented random assignment. The two classes demonstrated comparable English proficiency prior to the intervention, as confirmed by pre-test writing scores and vocabulary assessments.

Twenty students were assigned to the experimental group, which received teacher-guided AI instruction based on the AI-MWD Framework, while twenty students formed the control group, which used the same generative AI tools without systematic teacher-guided metacognitive mediation.

All participants provided informed consent before data collection. The study complied with institutional ethical requirements, and participation had no influence on students' course grades or academic evaluation.

C. Instructional Intervention

Both groups completed identical writing tasks using the same generative AI platform throughout the intervention period. The critical difference lay not in access to AI but in the instructional design governing AI use.

Students in the experimental group received explicit instruction in prompt design, planning strategies, self-monitoring techniques, reflective revision, and scaffold fading. Teachers progressively reduced instructional support across the intervention, encouraging learners to assume increasing responsibility for evaluating AI-generated suggestions and regulating their own writing processes.

In contrast, students in the control group were permitted to use generative AI freely for drafting and revision but received no systematic metacognitive guidance or structured scaffold fading. Consequently, AI functioned primarily as an independent writing assistant rather than as a teacher-mediated learning scaffold.

D. Instruments

Multiple complementary instruments were employed to capture both cognitive and linguistic dimensions of writing development.

Writing performance was assessed using the Complexity–Accuracy–Fluency (CAF) framework [8]. Complexity was measured through syntactic and lexical indicators, accuracy through grammatical and lexical error rates, and fluency through writing productivity measures. All writing samples were independently evaluated by two trained raters using standardized scoring criteria. Inter-rater reliability was established prior to formal scoring and demonstrated satisfactory agreement.

Learners' cognitive experiences were measured using a Cognitive Load questionnaire, while metacognitive writing autonomy was evaluated through a validated self-report instrument adapted from previous writing strategy research. Internal consistency for all questionnaire scales was examined using Cronbach's alpha, with coefficients exceeding the commonly accepted threshold for acceptable reliability.

E. Data Collection Procedure

Data collection consisted of three consecutive phases.

Phase 1 (Week 1): Baseline Assessment. Participants completed the writing pre-test together with the cognitive load and metacognitive autonomy questionnaires.

Phase 2 (Weeks 2–13): AI-MWD Intervention. The experimental group participated in teacher-guided AI-

supported writing instruction following the AI-MWD Framework, whereas the control group completed identical writing activities using unguided AI assistance.

Phase 3 (Week 14 and Follow-up). Immediately after the intervention, participants completed the post-test writing assessment and questionnaires. A delayed post-test was administered six weeks later to evaluate the long-term retention of writing performance after structured instructional support had been withdrawn.

F. Statistical Analysis

All statistical analyses were conducted using IBM SPSS Statistics.

Prior to inferential analyses, statistical assumptions were examined to ensure the appropriateness of parametric procedures. Data normality was assessed using the Shapiro–Wilk test, homogeneity of variance was examined through Levene's test, and the assumption of sphericity for repeated measurements was evaluated using Mauchly's test. Whenever the assumption of sphericity was violated, Greenhouse–Geisser corrections were applied.

Descriptive statistics were first calculated for all variables. Independent-samples t-tests were then conducted to compare cognitive load and metacognitive writing autonomy between instructional groups. Longitudinal changes in writing complexity, accuracy, and fluency were subsequently analyzed using 2×3 mixed-design analyses of variance (ANOVA).

To facilitate interpretation beyond statistical significance, partial eta squared (η^2p) was reported as the primary measure of effect size for ANOVA analyses, while Cohen's d was calculated for pairwise group comparisons. Ninety-five percent confidence intervals (95% CI) were reported where appropriate. When significant interaction effects were identified, Bonferroni-adjusted pairwise comparisons were conducted to determine the specific time points at which differences occurred.

IV. RESULTS

A. Preliminary Analysis

Prior to hypothesis testing, descriptive statistics and assumption checks were conducted to evaluate the suitability of the data for parametric analyses. Inspection of the Shapiro–Wilk tests indicated that the distributions of the principal outcome variables did not significantly deviate from normality ($p > .05$). Levene's tests further confirmed the homogeneity of variance across instructional groups ($p > .05$), while Mauchly's test indicated that the assumption of sphericity was satisfied for repeated measurements ($\chi^2 = 1.84$, $p > .05$). Consequently, parametric statistical procedures were considered appropriate for subsequent analyses.

Table I summarizes the descriptive statistics for the principal study variables across all testing intervals.

TABLE I. DESCRIPTIVE STATISTICS OF THE OUTCOME VARIABLES

Variable	Experimental (M±SD)	Control (M±SD)
Cognitive Load (Extraneous)	3.62 ± 0.74	6.48 ± 1.12
Metacognitive Autonomy	4.22 ± 0.45	2.85 ± 0.61

Writing Accuracy (EFC/C, %)		
Pre-test	60.50 ± 6.80	59.80 ± 7.10
Post-test	82.40 ± 5.10	80.10 ± 6.40
Delayed Post-test	81.20 ± 4.80	68.30 ± 7.20
Writing Complexity (C/S)		
Pre-test	1.42 ± 0.15	1.40 ± 0.16
Post-test	1.78 ± 0.18	1.65 ± 0.22
Delayed Post-test	1.82 ± 0.16	1.52 ± 0.20
Writing Fluency (Words)		
Pre-test	280.00 ± 35.00	285.00 ± 38.00
Post-test	365.00 ± 42.00	355.00 ± 45.00
Delayed Post-test	370.00 ± 38.00	310.00 ± 40.00

^a n = 20 per group.

B. Cognitive Outcomes

To examine whether teacher-guided AI reduced learners' cognitive burden while strengthening self-regulated writing, independent-samples t-tests were conducted.

The results demonstrated that learners receiving the AI-MWD intervention reported significantly lower levels of extraneous cognitive load ($M = 3.62$, $SD = 0.74$) than those in the control group ($M = 6.48$, $SD = 1.12$), $t(38) = -9.54$, $p < .001$. The calculated effect size was exceptionally large ($d = -3.01$, 95% CI [-3.47, -2.25]), highlighting the robust impact of macro-pedagogical prompt filtering. Conversely, the experimental group demonstrated significantly higher levels of metacognitive writing autonomy following the intervention ($M = 4.22$, $SD = 0.45$) compared to the control cohort ($M = 2.85$, $SD = 0.61$), $t(38) = 8.06$, $p < .001$, $d = 2.55$, 95% CI [1.03, 1.71]. The observed effect sizes indicated that these differences were educationally meaningful rather than merely statistically significant.

These findings suggest that structured teacher mediation enables learners to engage with AI-generated feedback more efficiently, freeing up cognitive bandwidth for deeper text processing, while simultaneously promoting greater independence in writing self-regulation.

TABLE II. INDEPENDENT-SAMPLES T-TEST RESULTS

Variable	t(38)	p	Cohen's d	95% CI of Diff.
Cognitive Load	-9.54	< .001	-3.01	[-3.47, -2.25]
Metacognitive Autonomy	8.06	< .001	2.55	[1.03, 1.71]

C. Longitudinal Development of Writing Performance

Longitudinal changes in writing performance were examined using 2×3 mixed-design analyses of variance (ANOVA) across three measurement occasions (pre-test, post-test, and delayed post-test).

Significant interaction effects (Group $\times$ Time) were identified for the principal writing indicators, demonstrating that the two instructional conditions followed different developmental trajectories over time. For Writing Accuracy (EFC/C), a highly significant interaction effect was found, $F(2, 76) = 14.82$, $p < .001$, $\eta^2p = .28$. Similarly, Syntactic Complexity (C/S) exhibited a significant interaction, $F(2, 76) = 5.64$, $p = .005$, $\eta^2p = .13$, as did Writing Fluency, $F(2, 76) = 8.12$, $p < .001$, $\eta^2p = .18$.

Although both instructional groups demonstrated improvements immediately following the intervention (Week

8), only learners receiving teacher-guided AI (AI-MWD) maintained these gains during the delayed post-test (Week 14). In contrast, performance within the control group declined after structured instruction had concluded.

These results indicate that the AI-MWD Framework supported not only immediate writing improvement but also the long-term retention of writing competence.

TABLE III. MIXED ANOVA RESULTS

Outcome	Source	F(2,76)	p	Partial η^2
Accuracy	Time	42.15	< .001	.52
	Group $\times$ Time	14.82	< .001	.28
Complexity	Time	18.25	< .001	.32
	Group $\times$ Time	5.64	= .005	.13
Fluency	Time	28.40	< .001	.43
	Group $\times$ Time	8.12	< .001	.18

D. Pairwise Comparisons

Bonferroni-adjusted pairwise comparisons further clarified the significant interaction effects observed in the mixed ANOVA analyses.

For the experimental group, statistically significant improvements were observed from the pre-test to the post-test across all three CAF dimensions ($p < .001$), with no significant decline between the post-test and delayed post-test ($p > .05$). This pattern indicates that writing gains were successfully retained beyond the active intervention period.

By contrast, although the control group demonstrated significant improvements immediately after instruction (Pre to Post, $p < .001$), significant decreases were observed between the post-test and delayed post-test, particularly in writing accuracy (Mdiff = -11.80%, $p < .001$) and fluency (Mdiff = -45.00 words, $p < .001$). These findings suggest that improvements achieved through unguided AI use were less stable over time and highly prone to decay once access to the digital scaffold was altered or withdrawn.

TABLE IV. BONFERRONI PAIRWISE COMPARISONS

Dimension & Comparison	Exp. M diff.	p	Ctrl. M diff.	p
Accuracy (EFC/C, %)				
Pre $\rightarrow$ Post	+21.90 %	< .001	+20.30 %	< .001
Post $\rightarrow$ Delayed	-1.20 %	= .420	-11.80 %	< .001
Pre $\rightarrow$ Delayed	+20.70 %	< .001	+8.50 %	= .012
Complexity (C/S)				
Pre $\rightarrow$ Post	+0.36	< .001	+0.25	= .003
Post $\rightarrow$ Delayed	+0.04	= .510	-0.13	= .024
Pre $\rightarrow$ Delayed	+0.40	< .001	+0.12	= .140
Fluency (Words)				
Pre $\rightarrow$ Post	+85.00	< .001	+70.00	< .001
Post $\rightarrow$ Delayed	+5.00	= .680	-45.00	< .001
Pre $\rightarrow$ Delayed	+90.00	< .001	+25.00	= .085

E. Summary of Findings

Across all analyses, the AI-MWD Framework consistently outperformed conventional AI-assisted instruction in both cognitive and linguistic outcomes. Teacher-guided AI significantly reduced learners' cognitive load while strengthening metacognitive writing autonomy. More importantly, the longitudinal analyses demonstrated that structured scaffold fading enabled learners to retain improvements in writing complexity, accuracy, and fluency after active instructional support had been withdrawn.

Taken together, these findings provide empirical support for the central proposition of the AI-MWD Framework: the educational value of generative AI depends less on the technology itself than on the quality of teacher-guided pedagogical mediation that transforms AI-assisted performance into internalized writing competence.

V. CONCLUSION AND FUTURE DIRECTIONS

A. Synthesis of the Study

This study addressed a critical and timely pedagogical dilemma in computer-assisted language learning (CALL): the empirical tension between the immediate linguistic corrective power of Generative AI (GenAI) and the long-term cognitive costs of unguided technology use (i.e., cognitive overload and diminished learner autonomy). To bridge this gap, we designed, implemented, and empirically validated the AI-Based Teacher-Guided Metacognitive Writing Development (AI-MWD) model. This model systematically integrates Lev Vygotsky's Sociocultural Theory [5] and Jerome Bruner's Scaffolding Theory [6] into a structured, four-phase Gradual Release of Responsibility (GRR) framework.

By distinguishing between the teacher's macro-pedagogical scaffolding (conceptual framing, strategic prompt design, and metacognitive reflection) and GenAI's micro-digital scaffolding (immediate, contextualized linguistic feedback), the AI-MWD framework transforms GenAI from a disruptive "shortcut" into an internalized, psychological tool for language acquisition.

B. Summary of Empirical Contributions

The 14-week longitudinal quasi-experimental evaluation yielded several pivotal empirical insights that advance our understanding of AI-mediated second language acquisition:

- **Mitigation of Extraneous Cognitive Load:** The results demonstrated that unguided interaction with GenAI subjects EFL learners to an overwhelming volume of unstructured feedback, causing severe cognitive saturation. Conversely, the AI-MWD model successfully buffered this cognitive friction. The experimental group reported a significantly lower extraneous cognitive load (3.62 ± 0.74) compared to the control group (6.48 ± 1.12 , $p < .001$).
- **Cultivation of Metacognitive Autonomy:** By pacing and constraining student interaction with GenAI through strict prompt parameters and recursive processing, the AI-MWD model preserved and enhanced learner autonomy. The experimental cohort achieved substantially higher post-intervention metacognitive autonomy scores (4.22 ± 0.45) than their unguided peers (2.85 ± 0.61 , $p < .001$).

- **Resolution of the “Immediate vs. Long-Term” Paradox:** Our longitudinal CAF (Complexity, Accuracy, Fluency) analysis exposed a profound developmental divergence. At the immediate post-test (Week 8), both groups exhibited parallel, substantial improvements in writing accuracy (80%). However, at the delayed post-test (Week 14), six weeks after the removal of active instructional support, the unguided control group's writing performance collapsed back to near-baseline levels (68.30% accuracy). In stark contrast, the AI-MWD group successfully retained their linguistic gains (81.20% accuracy) while demonstrating organic growth in syntactic complexity ($C/S = 1.82$) and fluency (370.00 words).

These longitudinal findings provide clear empirical evidence that the systematic “fading” of scaffolds — the hallmark of the AI-MWD framework — is absolutely mandatory to facilitate the deep cognitive processing required to transition temporary, AI-assisted performance into stable, internalized linguistic competence.

C. Pedagogical and Theoretical Implications

The theoretical and practical implications of these findings are highly significant for modern educational landscapes:

1) Theoretical Paradigm Shift: Redefining the MKO

This study contributes to sociocultural theory by expanding the ontology of the “More Knowledgeable Other” [6]. In the GenAI era, the MKO is no longer restricted to a human expert. Instead, it can be conceptualized as a hybrid, distributed agency. GenAI serves as a hyper-responsive, digital MKO that operates dynamically within the learner's Zone of Proximal Development [5]. However, this digital MKO is only effective when it is structured and mediated by the human instructor's macro-pedagogical envelope.

2) Pedagogical Practice: Pacing Over Automation

For EFL/ESL educators, this study serves as a warning against the unguided integration of GenAI in the writing classroom. While the immediate correction of grammatical and stylistic errors by AI tools is highly attractive, it often creates an “illusion of learning” that bypasses the deep cognitive struggles necessary for actual language acquisition, echoing concerns that unmediated reliance on generative AI can erode learners' originality and critical engagement with their own writing [1]. Educators must intentionally design writing curricula that treat AI feedback as a highly regulated resource rather than an automated writing replacement.

D. Limitations and Future Research

Despite its rigorous longitudinal design, this study has several limitations that point to fertile areas for future research:

- **Sample Demographics and Scale:** The study was conducted with a relatively small, homogeneous

sample ($N=40$) within a single higher education institution. Future investigations should replicate this model across larger, multi-institutional cohorts with diverse language proficiency levels (e.g., beginner vs. highly advanced learners) to test its generalizability.

- **Granular Process Tracking:** Although the self-reported psychometric scales and corpus-based CAF metrics provided robust macro-level data, they do not capture the micro-level, real-time cognitive negotiations that occur during human-AI writing interactions. Future research should utilize eye-tracking technologies, keystroke logging, and think-aloud protocols to visualize how learners actively filter, accept, or reject specific GenAI revisions under different scaffolding conditions.
- **Alternative Writing Genres:** This study focused primarily on academic argumentative writing. Future studies should examine whether the AI-MWD framework is equally effective across other genres, such as creative writing, professional correspondence, or collaborative group writing projects.

E. Final Remarks

Ultimately, the educational value of generative AI does not lie in the sophistication of its algorithms, but in the quality of the pedagogical frameworks that govern its use. By demonstrating that structured, teacher-guided scaffold fading successfully mitigates cognitive overload and nurtures sustainable learner autonomy, the AI-MWD framework provides a robust and replicable blueprint for future-proof, technology-enhanced language education.

REFERENCES

- [1] J. S. Barrot, "Using ChatGPT for second language writing: Pitfalls and potentials," *Assessing Writing*, vol. 57, Art. no. 100745, 2023.
- [2] D. J. Woo, D. Wang, K. Guo, and H. Susanto, "Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process," *Education and Information Technologies*, 2024.
- [3] L. Tankelevitch, V. Kewenig, A. Simkute, A. E. Scott, A. Sarkar, A. Sellen, and S. Rintel, "The metacognitive demands and opportunities of generative AI," in *Proc. 2024 CHI Conf. Human Factors in Computing Systems*, Honolulu, HI, USA, 2024, Art. no. 680.
- [4] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, vol. 12, no. 2, pp. 257–285, 1988.
- [5] L. S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard Univ. Press, 1978.
- [6] D. Wood, J. S. Bruner, and G. Ross, "The role of tutoring in problem solving," *Journal of Child Psychology and Psychiatry*, vol. 17, no. 2, pp. 89–100, 1976.
- [7] J. H. Flavell, "Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry," *American Psychologist*, vol. 34, no. 10, pp. 906–911, 1979.
- [8] A. Housen and F. Kuiken, "Complexity, accuracy, and fluency in second language acquisition," *Applied Linguistics*, vol. 30, no. 4, pp. 461–473, 2009.