Article

A Practice-Based Study of Generative AI–Empowered "Teacher–Student–Machine" Triadic Collaborative Writing Instruction in College Foreign Language Education at Application-Oriented Universities

Liang Zhou[1] and Yue Huang[1,*]

[1] Chengdu College of Arts and Sciences, Chengdu, China
* Correspondence: Yue Huang, Chengdu College of Arts and Sciences, Chengdu, China

Abstract: Generative artificial intelligence (GenAI) is rapidly reshaping foreign language education, yet organizing teachers, students, and intelligent machines into effective collaborative structures remains significantly underexplored, especially within application-oriented universities. Grounded in a novel "teacher–student–machine" (TSM) triadic framework, this study designed and rigorously evaluated a GenAI-empowered college English writing model via a comprehensive 16-week quasi-experiment at a central Chinese university. Second-year non-English majors (N = 96) were assigned to an experimental group (n = 48) utilizing the TSM model with a domestic large language model, or a control group (n = 48) receiving conventional process-writing instruction. Mixed methods data included pre- and post-writing tests, self-efficacy questionnaires, experience surveys, process records, and semi-structured interviews. Results demonstrated baseline equivalence but significant post-test advantages for the experimental group (78.42 vs. 73.56; $p < .001$, $d = 0.75$), featuring notable gains in content and organization. Furthermore, error density decreased by 51.2%, feedback uptake reached 78%, and self-efficacy improved significantly ($p < .001$). Qualitative findings revealed a complementary division of tasks: machines provided instant diagnostic feedback, teachers orchestrated instructional tasks, and students retained authorship by critically evaluating AI outputs. Ultimately, this study provides an empirically validated, operationalized account of triadic collaboration, offering highly transferable pedagogical insights for resource-constrained educational contexts globally.

Keywords: generative ai; collaborative learning; writing instruction; human ai collaboration; higher education

Received: 03 October 2026
Revised: 11 November 2026
Accepted: 23 November 2026
Published: 26 November 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Research Background

The maturation of large language models has integrated generative artificial intelligence (GenAI) into core classroom practice. Since late 2022, GenAI tools capable of producing fluent, context-sensitive text within seconds have attracted hundreds of millions of users, with education emerging as one of the most significantly transformed sectors [1]. Research has documented both the opportunities and challenges that large language models present for education: they enable personalized tutoring, instantaneous feedback, and scalable support for self-regulated learning, while also raising important questions regarding academic integrity, cognitive dependency, algorithmic bias, and the reliability of machine-generated content. In second language writing specifically, GenAI is redefining the process of composing in an additional language, as tasks such as text generation, translation, error identification, and stylistic refinement—previously requiring substantial human effort—can now be partially or fully supported by intelligent systems.

Concurrently, educational discourse in China has been guided by the concept of new quality productive forces and its educational counterpart, new quality education. Since its introduction in September 2023, this framework has prompted calls for a qualitative transformation in educational goals, content, and pedagogical approaches to cultivate talent characterized by innovative capacity, complex problem-solving ability, and effective human–machine collaboration. New quality education is understood as a paradigm shift from uniform, standardized instruction toward lifelong, personalized, and cooperative learning environments [2]. Digital-intelligent technologies are recognized as key enablers of this transformation, though attention is directed toward avoiding overreliance on technology, ensuring equitable digital literacy development, and strengthening institutional governance capacity. The implementation of AI-empowered education is structured around four interrelated pillars: a digitally enhanced teaching environment, professionally developed educators, pedagogically sound instructional strategies, and the holistic development of learners. This evolution reflects a distinctive approach that integrates domestic educational priorities with global perspectives, while consistently emphasizing the foundational importance of students' cognitive development and problem-solving competencies. Within higher education, the integration of artificial intelligence is prompting a reexamination of university classroom dynamics, calling for a transition from technological supplementation toward comprehensive ecological adaptation.

Foreign language writing instruction lies at the intersection of these two developments. College English writing remains a compulsory component of undergraduate education in China, and writing proficiency continues to represent one of the most challenging aspects of students' overall language competence. In this domain, GenAI has been met with both strong interest and cautious consideration: interest stems from the potential of automated feedback to support instructors in managing large cohorts, while cautious consideration reflects concerns that unguided use of such tools may undermine the very competencies that writing instruction aims to develop [3].

1.2. Research Status and Existing Problems in the Subfield

Research on generative artificial intelligence–supported second language writing has expanded rapidly [3]. Empirical studies indicate that such systems can serve as sources of ideas, language models, and feedback during the writing process; however, their educational effectiveness depends critically on task design and the nature of student engagement with machine-generated output. Surveys reveal differing perspectives between students and educators: students often prioritize efficiency and accuracy improvement, while educators express concerns about overdependence, diminished development of drafting abilities, and challenges in assessing authorship. Systematic reviews emphasize that the pedagogical advantages of large language models are neither inherent nor consistent; they require deliberate instructional design, ethical considerations, and sustained human involvement.

Within China, scholarly discourse has increasingly centered on the concept of triadic collaboration among teacher, student, and machine. A classroom paradigm based on the principle of peer-assisted learning integrates a digital-intelligent assistant as a dynamic scaffold that bridges intuitive understanding with explicit instruction, shifting from fixed role allocation to adaptive interdependence. Drawing on intersubjectivity theory, research highlights the importance of avoiding reductive subject–object framing in human–machine interaction and identifies three practical models of collaborative instruction: teacher and machine co-delivering instruction while students learn; teacher-led instruction supplemented by student-guided machine training; and teacher-led instruction combined with joint student–machine learning activities. Other work reexamines interdisciplinary thematic teaching under this triadic framework, illustrating how intelligent teaching agents support scenario construction, knowledge synthesis, and individualized diagnostic analysis—while affirming the irreplaceable educative function of the teacher. At the institutional level, a large-scale survey of undergraduate students

across diverse higher education institutions indicates that college English instructors are generally regarded as competent; however, capacities for adaptive instructional adjustment, individualized feedback delivery, integration of value-oriented education, and effective use of AI tools remain areas requiring targeted enhancement [4]. Frameworks have been proposed to define competencies for educators in this evolving context, including multi-dimensional models for high-quality teaching professionals and for digital-intelligent curriculum designers aligned with triadic collaboration.

Despite rapid theoretical and empirical progress, several persistent challenges remain in classroom implementation. First, AI tools are frequently integrated superficially—as automated answer generators or grammar correctors—resulting in shallow, sequential human–machine interaction rather than authentic, synergistic collaboration. Second, feedback in large-enrollment courses tends to be delayed and generic; instructors lack sufficient time to respond meaningfully to multiple drafts, leading students to engage in only surface-level revision. Third, students often adopt machine output uncritically, yielding improved surface features without corresponding growth in core composing competence—a pattern consistent with broader concerns about cognitive offloading [5]. Fourth, application-oriented universities confront these issues more acutely: their students typically begin with lower baseline writing proficiency and less developed self-management strategies, class sizes are larger, and instructor capacity in AI-related pedagogy varies significantly.

1.3. Research Gaps

Three gaps in the existing literature follow from this review. First, there is a gap between theory and operationalization: the triadic collaborative instruction construct has been elaborated conceptually and teacher competency frameworks have been proposed, but few studies translate these constructs into a concrete, replicable instructional model for a specific course with specified role assignments, task flows, and data loops. Second, there is a gap between conceptual advocacy and empirical verification: much of the domestic literature relies on theoretical deduction or case illustration, and quasi-experimental evidence on whether triadic collaborative instruction actually improves measurable writing outcomes remains scarce, particularly in college foreign language writing [6]. Third, there is a gap in context: existing empirical work on generative artificial intelligence in second language writing is concentrated in research-intensive universities or in international contexts, whereas application-oriented universities—which enroll a large share of Chinese undergraduates and serve an explicitly employment-facing mission—are rarely examined as a distinct research site. How triadic collaboration should be configured under conditions of limited student proficiency and constrained instructional resources is therefore an open question of both theoretical and practical significance.

1.4. Research Purpose, Content, and Approach

The present study addresses these gaps by designing, implementing, and evaluating a generative artificial intelligence–enhanced triadic collaboration model for teacher–student–machine (TSM) interaction in college English writing instruction at an application-oriented university [7]. Three objectives are pursued: (1) to operationalize the TSM framework into a writing instruction model with clearly defined roles for the teacher, the student, and the machine across the entire composing cycle; (2) to assess, through a 16-week quasi-experimental design, whether the model produces improved outcomes in writing performance, writing sub-competencies, revision behavior, and writing self-efficacy compared with conventional process-writing instruction; and (3) to clarify, using process records and interviews, the collaborative mechanisms underlying the model's effectiveness and the contextual conditions influencing its success or limitation.

The research proceeds in four phases. First, scholarly literature from domestic and international sources is synthesized to inform foundational design principles. Second, the model is developed as a closed-loop system encompassing task design, machine-generated feedback, critical revision, and teacher orchestration across pre-writing, while-

writing, and post-writing stages. Third, a mixed-methods quasi-experiment is carried out with two intact classes, integrating pre- and post-writing assessments, questionnaires, process documentation, and semi-structured interviews [8]. Fourth, quantitative and qualitative data are triangulated to address the research questions and iteratively refine the model.

1.5. Theoretical and Practical Significance

Theoretically, the study makes three contributions. It converts the TSM framework from a descriptive construct into an operationalized instructional design amenable to cumulative empirical testing; it extends intersubjectivity-based accounts of human–machine collaboration into foreign language writing by specifying how agency is distributed among the three actors across composing stages; and it supplies context-sensitive evidence on how the new quality education agenda can be enacted in a concrete disciplinary setting. Practically, the study offers application-oriented universities a transferable instructional model together with prompt scaffolds, evaluation rubrics, and evidence-based guidance on the allocation of responsibilities between human and machine. The findings further inform teacher development programs by identifying the specific competencies the model demands of writing teachers—prompt design, data-informed diagnosis, and critical AI literacy instruction [6].

1.6. Structure of the Paper

The remainder of this paper is organized as follows. Section 2 reviews relevant research conducted internationally and domestically, followed by a critical appraisal. Section 3 outlines the research design, covering model construction, participant selection, instrumentation, and procedural implementation. Section 4 presents and analyzes both quantitative and qualitative findings. Section 5 discusses the principal results in relation to existing scholarship, addresses potential sources of bias, and clarifies theoretical and practical implications. Section 6 concludes by summarizing key findings, acknowledging limitations, and suggesting directions for future research [3].

2. Literature Review

2.1. Research Abroad

International research relevant to this study clusters into three strands. The first concerns the opportunities and challenges of large language models for education in general. Large language models have been found to support personalized learning, formative assessment, and teacher workload reduction, while simultaneously introducing risks such as misinformation, bias amplification, and overreliance [9]. Realizing the benefits requires deliberate instructional integration rather than unrestricted access. Systematic analyses of practical and ethical challenges highlight dimensions including reliability, fairness, accountability, privacy, and human agency, indicating that governance frameworks and pedagogical redesign must evolve in tandem with technological advancement.

The second strand focuses on generative artificial intelligence in second language and foreign language writing [10]. Generative AI has been observed to supply ideas, linguistic models, and feedback previously inaccessible to many learners, yet it may also encourage plagiarism, homogenize voice, and reduce the cognitive effort essential for developing writing competence. Studies indicate that both students and educators acknowledge efficiency gains, but perspectives differ regarding appropriate levels of machine involvement: educators emphasize process integrity and authorship, whereas students prioritize product quality and time savings. Collectively, findings suggest that generative AI supports writing development most effectively when learners are required to evaluate, select, and transform machine output rather than accept it passively. Classroom-based observations reinforce this view, indicating that unrestricted access tends to widen performance gaps: learners with stronger metacognitive strategies engage critically with

machine feedback, whereas those with weaker strategies tend to delegate compositional tasks rather than actively compose.

The third strand addresses human–AI collaboration as an instructional arrangement. A recurring finding is that effective integration necessitates redefining the roles of teachers and learners rather than merely introducing a new tool: teachers assume responsibilities as designers of learning experiences and critical interpreters of machine output, while learners are expected to develop new literacies for purposeful interaction with AI [3]. This conceptual shift aligns with frameworks developed in domestic scholarship, although international literature typically frames the relationship through concepts of distributed cognition and augmented teaching rather than the ternary 'teacher–student–machine' model.

Across the three strands, methodological patterns are also noteworthy. Much of the early evidence relies on self-report surveys, think-aloud protocols, and short-term writing tasks, whereas longitudinal, course-length empirical studies remain rare; outcome measures vary widely—from holistic scores to error counts and learner perceptions—complicating cross-study comparison. The field therefore requires longer-duration, process-sensitive investigations that track not only whether learners write more effectively with generative AI, but also how their composing behaviors evolve over time [6].

2.2. Research at Home

Domestic research has developed along four interrelated lines. The first establishes the macro framework of new quality education, which breaks through the fence of schooling to develop lifelong education, personalized learning, and human–machine cooperative learning, shifting the essence of education from cultivating people "with knowledge" to cultivating people who "can cooperate and innovate." Digital-intelligent technologies are examined for their transformative impact on knowledge, teaching, learning, management, and evaluation, supporting a practical pathway of "response–transformation–implementation–enhancement–innovation–safeguard." The internal mechanism of AI-empowered new quality education is specified in terms of environment, teacher, pedagogy, and talent. Conceptual foundations and implementation strategies for a Chinese approach to new quality education have been articulated, while analyses have also addressed core continuities and necessary adaptations in alignment with evolving new quality productive forces [11]. From the perspective of talent cultivation paradigms, integrative innovation in science-and-humanities education is proposed to cultivate new quality talent suited to the digital-intelligent society.

The second line concerns the reconstruction of classroom ecology. An "AI + classroom ecology" paradigm for universities highlights TSM triadic collaboration, three-dimensional element chains, knowledge-graph networking, and agile iteration of ternary structures, with implementation prototypes including human–machine collaborative elastic generation. Although focused on adult education, similar work positions digital-intelligent empowerment as central to systemic transformation of educational form. Evaluation approaches in new liberal arts education are examined for responsiveness to the logic of new quality productive forces, and higher vocational education is analyzed for functional realignment and adaptive response pathways—highlighting distinct challenges faced by application-oriented institutions [12].

The third line develops the TSM triadic collaboration construct itself. A "curriculum–behavior–data" analytical framework and a three-stage "knowing–doing–creating" process have been introduced to concretize interactions among teachers, digital-intelligent teaching assistants, and artificial intelligence teachers. A theoretical treatment decomposes the human–machine collaborative instruction model from the perspective of intersubjectivity into interaction rules grounded in educational communication, a shared "perception–cognition–action" pathway, multiple dialogic relationships, and a balance between preset and generative goals. Triadic collaboration has also been applied to interdisciplinary thematic instruction, specifying how teachers use thinking-chain

prompts to guide teaching agents in integrating multidisciplinary knowledge while retaining responsibility for critical verification and educative judgment.

The fourth line addresses the teacher workforce. Empirical evidence indicates that college English teachers show relative weakness in areas most critical to AI-era instruction—dynamic adjustment, personalized feedback, and AI tool application—with student expectations varying systematically across institutional types, suggesting that instructional models developed elsewhere may not be directly transferable to application-oriented universities. A three-dimensional, twelve-competency framework for new quality teachers has been proposed, spanning value orientation, practice proficiency, and innovation capacity. The emerging role of the digital-intelligent curriculum designer has been identified, supported by a five-dimensional competency framework encompassing design, operational coordination, decision-making, creativity, and leadership.

2.3. Critical Review

Taken together, the literature establishes three points of consensus: GenAI is transforming language education irreversibly; its benefits are conditional on instructional orchestration; and the teacher–student–machine relationship must be re-specified rather than left to improvisation. Four shortcomings nevertheless remain. First, operationalization is insufficient: domestic triadic collaboration research is strong on conceptual architecture but rarely specifies instructional procedures at the granularity a writing teacher can enact. Second, empirical verification is unbalanced: large-scale surveys exist for teacher competence and case illustrations for triadic classrooms, but controlled comparisons measuring student learning outcomes under triadic collaborative instruction are lacking [13]. Third, disciplinary specificity is weak: the triadic collaboration literature spans basic education, vocational education, and general higher education, but foreign language writing—a domain distinguished by recursive drafting, language-form feedback, and authorship sensitivity—has not been systematically addressed. Fourth, the application-oriented university context is under-represented, even though its combination of large classes, modest student proficiency, and employment-focused curricula makes it arguably the site where a scalable human–machine division of responsibilities is most needed and most challenging. The present study is designed in direct response to these four shortcomings.

3. Research Design

3.1. Research Questions

The study addresses four research questions:

RQ1. Does GenAI-empowered triadic collaborative instruction produce greater improvement in overall writing performance among students in an application-oriented university, compared with conventional process-writing instruction?

RQ2. In which sub-competencies of writing (content, organization, language use) and in which writing-process behaviors (revision depth, feedback uptake, error density) are the differences, if any, most pronounced?

RQ3. Does the instructional model differentially affect students' writing self-efficacy?

RQ4. How do students and teachers perceive and implement the roles of teacher, student, and artificial intelligence system, and what mechanisms explain the observed outcomes?

3.2. Participants

The study involved 96 second-year non-English major undergraduates from an application-oriented university in central China, enrolled in two intact classes of the compulsory College English (Writing Module) course during the autumn semester of 2025. Classes were randomly assigned to the experimental group (EG, n = 48; 22 male, 26 female; mean age = 19.8 years) and the control group (CG, n = 48; 24 male, 24 female; mean age = 19.7 years). All participants had completed approximately nine years of English instruction, demonstrated intermediate proficiency, and had no prior systematic

experience with generative artificial intelligence–assisted writing. A single instructor delivered both courses to minimize instructor-related variability. Classes met twice weekly, with writing instruction comprising roughly one-third of total class time; the remainder addressed reading and listening components as specified in the syllabus [14]. Informed consent was obtained from all participants, the study adhered to institutional ethical guidelines for educational research, and all reported data are fully anonymized.

3.3. The TSM Triadic Collaboration Model for Writing Instruction

The instructional model implements the TSM triadic collaboration construct for the full composing cycle, assigning complementary and non-substitutable functions to each of the three actors across three stages (see Figure 1).

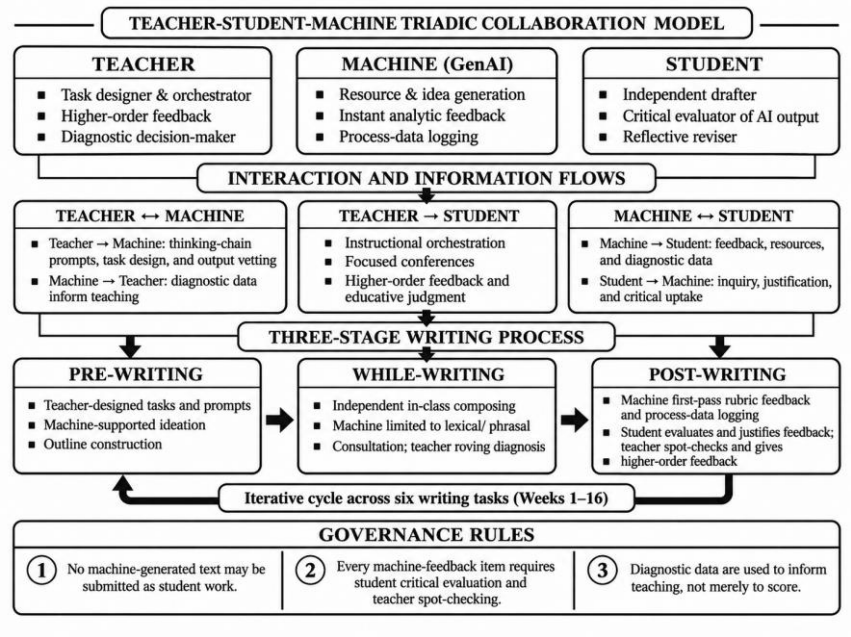

Figure 1. The TSM Triadic Collaboration Model for College English Writing Instruction.

In the pre-writing stage, the teacher designs genre- and task-appropriate assignments and converts instructional intentions into structured thinking-chain prompts; the machine, guided by these prompts, generates topic-relevant background materials, brainstormed idea clusters, and exemplar language features at calibrated difficulty; students interact with the machine dialogically to clarify the task, explore positions, and build an outline, while the teacher monitors the quality of machine output and corrects factual or ideological deviations. In the while-writing stage, students compose independently in class without machine text generation—a deliberate design feature intended to safeguard authentic composing processes and to mitigate cognitive offloading; the machine functions only as a consultable resource for lexical and phrasal retrieval, simulating a dictionary-plus-corpus role, while the teacher conducts roving diagnosis. In the post-writing stage, the machine provides immediate analytic feedback on drafts against the instructional rubric and logs diagnostic data on error types and revision traces; students must evaluate machine feedback critically—accepting, modifying, or rejecting each suggestion with a brief written justification—before producing a revised draft; and the teacher, relieved of routine error correction by the machine's first-pass feedback, concentrates on higher-order issues of argumentation, coherence, and voice, spot-checks the machine's feedback quality, conducts focused conferences informed by the diagnostic data, and adjusts subsequent task design. This closed loop of task design, machine feedback, critical revision, and teacher orchestration constitutes the model's core mechanism: machine speed and data-processing capacity are exchanged for teacher time and attention, which are reinvested in educative judgment that machines cannot supply.

Three governance rules, derived from risk analyses in the literature, bound the model: (1) the machine may not generate student-facing text that is submitted as student work; (2) every piece of machine feedback must pass through student critical evaluation and teacher spot-checking; and (3) diagnostic data are used by the teacher to teach, not merely to score.

3.4. Instruments

Four instruments were used. (1) Writing tests. Parallel argumentative writing tasks (e.g., "Should college students be required to learn AI-related skills?") were administered as pre- and post-assessments, each completed in 30 minutes in class without technological assistance. Essays were scored on a 100-point analytic rubric adapted from the scoring framework of China's national college English testing, comprising content (30), organization (30), and language use (40). Two trained raters scored all essays independently; inter-rater reliability was satisfactory (Pearson r =.86 for the pre-assessment and.88 for the post-assessment), and discrepancies exceeding five points were resolved by a third rater. (2) Writing self-efficacy questionnaire. An adapted 12-item, five-point Likert scale covering ideation, organization, language control, and revision self-efficacy was administered at pre- and post-assessment; internal consistency was good (Cronbach's α =.87). (3) Instructional experience survey. At the end of the instructional period, EG students completed a 10-item, five-point survey on perceived usefulness, feedback quality, teacher guidance, and concerns about over-reliance [15]. (4) Semi-structured interviews. Eight EG students (four high-gainers and four low-gainers selected by gain-score quartiles) and both raters/teachers were interviewed for 25–40 minutes each; the interviews explored the implemented task distribution, perceived changes, and challenges. In addition, writing-process records were collected for all EG drafts, including machine feedback logs, student justification notes, revision counts, and feedback-uptake coding. All instruments were piloted with a parallel class not included in the formal study, and minor wording revisions were made on the basis of the pilot results.

3.5. Instructional Procedure

The instructional program spanned 16 weeks [16]. Both groups completed 32 writing instruction sessions, addressing six writing tasks (narrative, descriptive, expository, two argumentative essays, and an integrated practical genre), with identical instructional time, materials, and assignment topics. The control group followed a conventional process-writing approach: teacher-led pre-writing discussion, in-class drafting, collection of drafts by the teacher, delayed whole-class feedback, and optional revision. The experimental group implemented the triadic collaborative routine outlined in Section 3.3, utilizing a domestically accessible large language model platform via classroom terminals and students' personal devices, in full compliance with the three governance rules. During weeks 1–2, experimental group students received training in critical AI interaction—including prompt formulation, verification strategies, and documentation of feedback evaluation—and task complexity increased progressively across the semester. Data collection adhered to a fixed schedule: pre-test and self-efficacy questionnaire administered in week 1; post-test and follow-up questionnaires in week 16; continuous process records; and interviews conducted in weeks 15–16.

3.6. Data Analysis

Quantitative data were analyzed using SPSS 26.0. Independent-samples t-tests compared group differences at baseline and post-test; paired-samples t-tests evaluated within-group changes. Effect sizes were reported as Cohen's d. Error density—defined as errors per 100 words—was calculated from rater-marked drafts. Revision behavior was measured by the mean number of substantive revisions per draft and the feedback-uptake rate, which reflected the proportion of machine or teacher feedback points followed by corresponding revisions. Interview recordings were transcribed verbatim and subjected to thematic analysis through two iterative coding rounds, with coder discrepancies

resolved via consensus discussion [11]. Quantitative and qualitative findings were then triangulated to address RQ4.

4. Results and Data Analysis

4.1. Overall Writing Performance

At baseline, the two groups showed no significant difference in overall writing performance (EG: M = 68.95, SD = 6.42; CG: M = 69.38, SD = 6.21; t(94) = −0.33, p =.740), confirming initial equivalence [7]. After the 16-week instructional period, the EG significantly outperformed the CG (EG: M = 78.42, SD = 6.13; CG: M = 73.56, SD = 6.87; t(94) = 3.66, p <.001, Cohen's d = 0.75). Within-group analyses revealed significant improvement in both groups, though the magnitude differed substantially: the EG gained 9.47 points (t(47) = 12.63, p <.001), while the CG gained 4.18 points (t(47) = 5.62, p <.001); the between-group difference in gain scores was also statistically significant (t(94) = 3.86, p <.001). As summarized in Table 1 and Figure 2, the triadic collaborative approach yielded approximately 2.3 times the average gain observed under conventional instruction over the same timeframe (As shown in Table 2).

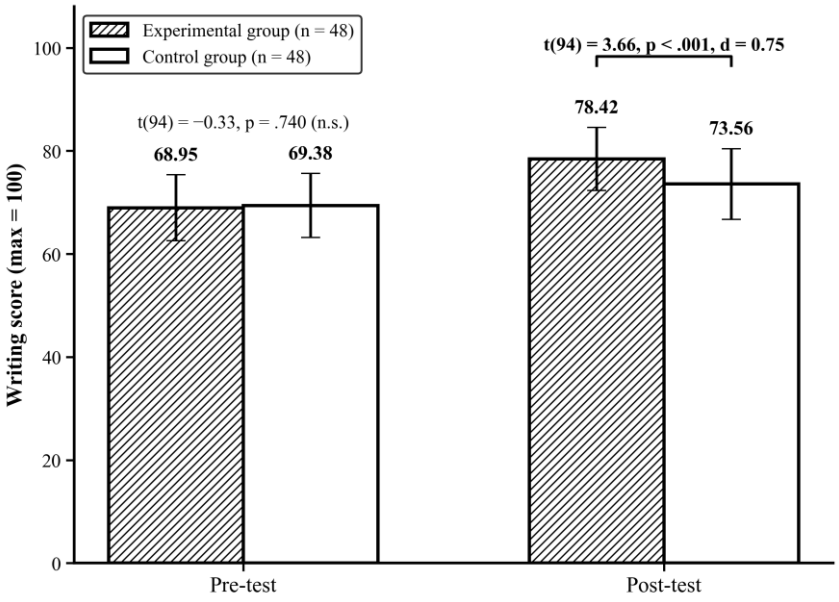

Figure 2. Pre-test and Post-Test Writing Scores by Group (Error Bars: ±1 SD).

Table 1. Pre-Test and Post-Test Writing Performance by Group

Group	Pre-test M(SD)	Post-test M(SD)	Gain	Paired t	Between-group (post)	Cohen's d
EG (n = 48)	68.95 (6.42)	78.42 (6.13)	9.47	12.63***	3.66***	0.75
CG (n = 48)	69.38 (6.21)	73.56 (6.87)	4.18	5.62***	—	—

Table 2. Post-Test Sub-Scores by Group

Sub-competency (max)	EG M(SD)	CG M(SD)	t(94)	p	Cohen's d
Content (30)	24.10 (2.85)	22.30 (2.91)	2.98	0.004	0.61
Organization (30)	23.85 (2.92)	21.96 (2.99)	3.12	0.002	0.64
Language use (40)	30.47 (3.78)	29.30 (3.25)	2.05	0.043	0.42

Note. ***p < .001. Between-group pre-test difference: t(94) = −0.33, p = .740.

4.2. Writing Sub-Competencies and Process Behaviors

Post-test sub-scores revealed a differentiated pattern (see Table 2). The experimental group's advantage was largest in organization (M = 23.85 vs. 21.96; t(94) = 3.12, p =.002, d = 0.64) and content (M = 24.10 vs. 22.30; t(94) = 2.98, p =.004, d = 0.61), and smaller though

still significant in language use (M = 30.47 vs. 29.30; t(94) = 2.05, p =.043, d = 0.42). This ordering is pedagogically informative: the model's pre-writing machine-supported idea exploration and outline construction directly target content and organization, whereas language-form improvement depends more on long-term internalization. Within the experimental group, sub-score gains were uneven across individuals: students in the lower baseline quartile improved most in language use, whereas those in the upper quartile improved most in content, a pattern consistent with the ceiling structure of the rubric.

Process-behavior data corroborated the test-score pattern. Error density declined from 8.6 to 4.2 errors per 100 words in the experimental group (−51.2%) and from 8.4 to 6.7 in the control group (−20.2%). Experimental group students produced a mean of 5.8 substantive revisions per draft cycle, compared with 2.3 in the control group, and the experimental group's feedback-uptake rate reached 78%, against 54% in the control group [12]. Notably, 19.4% of machine feedback points were explicitly rejected by experimental group students with written justification, indicating that the critical-evaluation requirement was genuinely enacted rather than formally present.

4.3. Writing Self-Efficacy

Writing self-efficacy increased significantly in both groups, with a larger gain observed in the experimental group. The experimental group rose from M = 3.12 (SD = 0.54) to M = 4.05 (SD = 0.48) (t(47) = 10.24, p <.001), whereas the control group increased from M = 3.15 (SD = 0.56) to M = 3.48 (SD = 0.52) (t(47) = 3.98, p <.001); the between-group difference at post-test was statistically significant (t(94) = 5.59, p <.001, d = 1.13). Sub-scale analysis revealed the greatest improvements in the experimental group for ideation and revision self-efficacy, aligning with the instructional emphasis on machine-supported brainstorming and iterative revision.

4.4. Instructional Experience and Interview Findings

The end-of-course survey (EG, five-point scale) showed strongly positive perceptions: timeliness of feedback (M = 4.32), reduction of writing-related stress (M = 4.18), and—strikingly—perceived indispensability of teacher guidance (M = 4.41). Students also reported moderate concern about over-reliance (M = 3.62) and rated their own critical evaluation of AI output at M = 3.85, suggesting awareness rather than uncritical acceptance [10].

Thematic analysis of the interviews yielded three organizing themes that illuminate the mechanism behind the quantitative gains. First, redistribution of attention: both teachers described the machine's first-pass feedback as "freeing two-thirds of correction time," which was reallocated to conferences on argumentation and coherence, and high-gaining students similarly reported that feedback arrived "while the draft was still warm," enabling revisions that "actually stuck." Second, supported self-direction: students characterized the prompt scaffolds and the justification requirement as initially burdensome but eventually internalized, such that by the final tasks they "argued with the machine by habit." Third, boundary negotiation: teachers reported recurring episodes in which the machine generated fluent but hollow or factually inaccurate content, and both groups converged on the view that the teacher's professional judgment and the student's critical evaluation constituted irreplaceable safeguards. Low-gaining students, by contrast, tended to accept machine feedback verbatim in early tasks and required teacher support to restore critical engagement—evidence that the governance rules were a necessary, not decorative, component of the model.

A further pattern concerned individual differences in machine interaction. High-gaining students tended to use the machine recursively, returning to it after each revision with increasingly specific questions, whereas low-gaining students typically issued a single broad request and stopped—a behavioral contrast that mirrors the distinction between dialogic and one-shot use described in collaborative learning research [17]. Teachers also reported that the feedback-justification sheets, although designed as an accountability device, doubled as an assessment resource that revealed students'

rhetorical decision-making more transparently than final drafts alone. These observations suggest that the model's benefits are partly mediated by the quality of students' machine interaction, a mediating variable that future iterations of the model could target directly through explicit interaction-strategy training.

5. Discussion

5.1. Principal Findings

Three principal findings emerge from the converging evidence. First, the TSM triadic collaborative model yielded a statistically significant and educationally meaningful advantage over conventional process-writing instruction, with an overall effect size of d = 0.75—falling within the medium-to-large range—and a gain-score ratio of approximately 2.3:1. Second, this advantage varied across sub-competencies: it was strongest in organization and content, moderate in language use, and corresponded to observable shifts in process behaviors—including deeper revision, higher feedback uptake, and lower error density—indicating that the model primarily restructures the composing process rather than enhancing surface-level linguistic features. Third, the model substantially increased writing self-efficacy (d = 1.13), and qualitative evidence suggests this improvement was facilitated by timely feedback and by students' active negotiation of machine-generated output rather than passive acceptance [14]. Collectively, these results indicate that the model's effectiveness arises not from any isolated component but from the coordinated sequencing of human and machine contributions throughout the composing cycle.

5.2. Comparison with Prior Research

The findings are consistent with, and in several respects extend, prior work. The conditional-benefit conclusion of international GenAI research is confirmed: machine involvement produced gains precisely because it was embedded in a structured instructional arrangement with explicit governance rules, echoing findings that product gains and process integrity can coexist only when roles are clearly assigned. The redistribution-of-attention mechanism instantiates accounts of AI empowering education through the reconstruction of teacher roles, and it empirically substantiates claims that triadic collaboration can move classrooms from static task allocation toward dynamic symbiosis—though the present data suggest that, at least in a writing course, the collaboration is better described as deliberately asymmetrical, with the machine carrying speed and data while humans retain judgment. The differentiated sub-competency gains are consistent with observations that teaching agents are strongest in knowledge integration and situation generation and weakest where educative judgment is required. The finding that students rated teacher guidance as indispensable directly addresses workforce concerns raised in recent surveys: rather than displacing teachers, the model exposed and amplified the value of precisely those competencies—dynamic adjustment and personalized feedback—that such surveys identified as most challenging to sustain. The study also adds nuance to intersubjectivity-based theory: the "perception–cognition–action" pathway was enacted, but the balance between preset and generative goals tilted toward preset goals in this proficiency-limited context, suggesting that the generative pole of human–machine collaboration may need to be scaffolded more gradually for learners with developing proficiency.

Where the results diverge from optimistic accounts, the divergence is instructive. The model did not produce large gains in language use relative to other dimensions, and 19.4% of machine feedback was rejected by students—findings that temper any narrative of the machine as a superior language tutor and align with warnings about the hollowness and error-proneness of machine-generated text. The persistence of moderate over-reliance concern (M = 3.62) despite governance rules indicates that the risk identified internationally persists in managed form rather than disappearing.

5.3. Sources of Bias and Alternative Explanations

Several design features constrain interpretation. First, the same instructor taught both groups, which controls for instructor-related variability but introduces the possibility of differential enthusiasm; although materials and time were equated, a Hawthorne-type effect in the experimental group cannot be excluded. Second, the novelty of generative AI tools may have temporarily elevated engagement in the experimental group; the 16-week duration mitigates but does not eliminate this concern. Third, the analytic rubric, while reliable, assigns fixed weightings to content, organization, and language, so conclusions about sub-competencies are partly dependent on the scoring scale. Fourth, process-behavior measures such as revision counts and feedback uptake rate were collected under differing instructional conditions across groups and therefore reflect those contextual factors rather than inherent learner tendencies. Fifth, although gains observed under test conditions without machine assistance indicate genuine competence transfer rather than tool-assisted performance inflation, increases in self-efficacy may stem partly from familiarity and successful use of the tool rather than from writing development alone.

5.4. Theoretical Implications and Practical Value

Theoretically, the study specifies the TSM triadic collaboration construct as an operable mechanism: a closed loop in which machine speed and diagnostic data are converted, through teacher orchestration and student critical evaluation, into redistributed pedagogical attention. This specification refines ecological accounts of AI-infused classrooms by identifying attention redistribution as a concrete, measurable pathway from technological empowerment toward ecological reconstruction, and it grounds the new quality education agenda in disciplinary evidence: the model cultivates exactly the human–machine collaborative competence and critical disposition that new quality talent is said to require. Practically, the model offers application-oriented universities—whose operational conditions are well documented—a low-cost, scalable routine that requires no specialized platform beyond an accessible large language model, together with replicable artifacts including thinking-chain prompt templates, feedback-justification sheets, and diagnostic dashboards. For teacher development, the study implies that training should prioritize prompt engineering for genre pedagogy, diagnostic-data literacy, and the teaching of critical AI evaluation, in alignment with established competency frameworks. For assessment practices, the process data generated by the model—revision traces and justification notes—offer precisely the kind of process evidence that contemporary liberal arts assessment approaches emphasize. In applying the model, institutions should nevertheless adapt rather than adopt it wholesale: class size, device availability, and students' baseline proficiency condition the feasible granularity of teacher conferences and machine-feedback review, and the governance rules, rather than any specific platform, constitute the non-negotiable core of the design [14].

6. Conclusions and Prospects

6.1. Main Conclusions

This study designed a GenAI-empowered "teacher–student–machine" triadic collaboration model for college English writing instruction and evaluated it in a 16-week quasi-experiment at an application-oriented university. The model significantly outperformed conventional process-writing instruction in overall writing performance ($d = 0.75$), with the largest advantages in organization and content; it deepened revision behavior, more than doubled feedback uptake, halved error density, and substantially improved writing self-efficacy. Qualitative evidence identified attention redistribution, scaffolded independence, and boundary negotiation as the operative mechanisms and confirmed that governance rules preventing uncritical machine reliance were functionally necessary. The study thereby provides the first controlled, discipline-specific empirical evidence that TSM triadic collaboration, when operationalized with explicit role assignment and critical-evaluation safeguards, improves measurable writing outcomes.

6.2. Limitations

Four limitations should be acknowledged. First, the sample was limited to two intact classes at a single application-oriented university, which constrains generalizability across proficiency levels, majors, and institutional types. Second, the instructional sequence lasted one semester, so the durability of gains and the long-term trajectory of reliance on or disengagement from machine scaffolding remain unknown. Third, the design compares the full model against conventional instruction and therefore cannot isolate the contribution of individual components such as machine feedback, the justification requirement, or diagnostic-informed conferences; the mechanism claims, while triangulated, rest partly on qualitative inference. Fourth, measures such as error density and uptake rate were derived from drafts produced under different feedback conditions, and the self-report instruments remain susceptible to social desirability bias.

6.3. Future Directions

Future research may proceed in four directions. First, multi-site replications across institutional types and proficiency strata are needed to test the boundary conditions of the model, particularly in the application-oriented sector where student needs differ systematically. Second, longitudinal designs should track whether machine scaffolding can be faded without performance decay and how the balance between preset and generative goals shifts as learner competence grows. Third, component-analysis designs should isolate the active ingredients of the model, and learning-analytics approaches should utilize the process data the model naturally generates to produce finer-grained accounts of revision and uptake. Fourth, as new quality education advances from concept to systemic practice, future work should connect classroom-level triadic collaboration with curriculum-level design roles and evaluation frameworks, and examine how similar models can be adapted to other language skills and to interdisciplinary writing contexts, so that the triadic collaboration paradigm matures from a classroom technique into an institutional capability. Finally, comparative studies across GenAI platforms would clarify which findings are platform-invariant and which depend on specific model capabilities and interface designs.

Funding: A Project Supported by Research Center for Reform and Development of Newly Established Universities, 2026 Research Project: Research on the Construction and Practice of the "Teacher-Student-Machine" Ternary Collaborative Model of College Foreign Language Teaching in Application-Oriented Universities from the Perspective of New-Quality Education (Project No.: XJYX2026B17)

References

1. E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023.
2. L. Yan et al., "Practical and ethical challenges of large language models in education: A systematic scoping review," *British Journal of Educational Technology*, vol. 55, no. 1, pp. 90–112, 2024.
3. M. Warschauer et al., "The affordances and contradictions of AI-generated text for writers of English as a second or foreign language," *Journal of Second Language Writing*, vol. 62, 2023.
4. A. Barrett and A. Pack, "Not quite eye to AI: Student and teacher perspectives on the use of generative artificial intelligence in the writing process," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, p. 59, 2023.
5. C. Hou et al., "The effects of critical thinking intervention on reliance behaviors, problem-solving quality, and creativity during human-generative AI collaborative learning," *Computers & Education*, p. 105576, 2026.
6. X. Zhou and L. Schofield, "Using social learning theories to explore the role of generative artificial intelligence (AI) in collaborative learning," *Journal of Learning Development in Higher Education*, no. 30, 2024.
7. S. C. Tan, W. Chen, and B. L. Chua, "Leveraging generative artificial intelligence based on large language models for collaborative learning," *Learning: Research and Practice*, vol. 9, no. 2, pp. 125–134, 2023.
8. J. Liu, S. Li, and Q. Dong, "Collaboration with generative artificial intelligence: An exploratory study based on learning analytics," *Journal of Educational Computing Research*, vol. 62, no. 5, pp. 1014–1046, 2024.
9. N. Law et al., "The role of generative AI in collaborative problem-solving of authentic challenges," *British Journal of Educational Technology*, 2025.

10. K. Pahi et al., "Enhancing active learning through collaboration between human teachers and generative AI," *Computers and Education Open*, vol. 6, p. 100183, 2024.
11. A. Naik et al., "Generating situated reflection triggers about alternative solution paths: A case study of generative ai for computer-supported collaborative learning," in *International Conference on Artificial Intelligence in Education*, Cham: Springer Nature Switzerland, Jul. 2024, pp. 46–59.
12. J. Tian, W. Hu, R. Li, and Y. Li, "Exploring students' cognitive engagement in generative AI-enhanced online collaborative learning: does prior knowledge matters?," *Distance Education*, vol. 47, no. 2, pp. 237–267, 2026.
13. Q. Xia et al., "Is generative artificial intelligence (GenAI) a game changer for interdisciplinary collaborative learning in higher education?," *Education and Information Technologies*, vol. 31, no. 1, pp. 143–166, 2026.
14. F. R. Baskara, "From AI to we: Harnessing generative AI tools to cultivate collaborative learning ecosystems in universities," in *Proceeding International Conference on Learning Community (ICLC)*, vol. 1, no. 1, 2024.
15. Y. Wei and M. Perkins, "Generative AI and student collaboration: a scoping review of group work processes, outcomes, and risks," *International Journal for Educational Integrity*, vol. 22, no. 1, p. 8, 2026.
16. S. Strauß and N. Rummel, "Fostering collaborative learning and promoting collaboration skills: What generative AI could contribute," *OECD Digital Education Outlook*, p. 91, 2026.
17. Y. Yang et al., "Mapping the horizon: A scoping review of generative AI in collaborative learning," in *2025 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, Dec. 2025, pp. 1–8.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.