World Journal of Educational Studies
Print ISSN: 2959-9989
Online ISSN: 2959-9997
DOI: https://doi.org/10.61784/wjes3222

AI AS A COGNITIVE ENHANCEMENT TOOL: A FOUR-TIER INSTRUCTIONAL MODEL FOR TEM-8 WRITING

FeiWu Li
School of Foreign Studies, South China Agricultural University, Guangzhou 510640, Guangdong, China.

Abstract: The writing task in Test for English Majors Grade Eight (TEM-8) places high demands on learners' ability to synthesize information and express critical perspectives. Drawing on teaching practice in the English Writing and Translating Skills Training course for English majors at a Double First-Class agricultural university, this study develops a four-tier model of AI-enhanced writing instruction, with the four tiers corresponding to four ascending levels of cognitive engagement: evaluative regulation, critical analysis, comparative discernment, and selective revision. Using ChatGPT and DeepSeek as supporting tools, the model treats AI as a tool for cognitive enhancement rather than a substitute writer. By analyzing the strengths and weaknesses of AI-generated essays, comparing outputs from different platforms, and using AI critically to revise their own work, students make a cognitive shift from passively receiving feedback to actively directing their use of the technology. This paper explains the rationale, key components, and implementation of the model, offering a replicable framework for integrating AI into writing instruction for English majors in higher education.
Keywords: Artificial Intelligence; Test for English Majors Grade Eight (TEM-8); Writing instruction; Critical AI literacy; Instructional model

1 INTRODUCTION

Since the 2016 reform of its writing section, the Test for English Majors Grade Eight (TEM-8) has adopted a composite task format that combines information synthesis based on reading materials with independent viewpoint commentary. This change has shifted the demands on senior English majors from mere linguistic accuracy to multi-layered cognitive abilities involving "summarizing — evaluating — constructing", with critical thinking as a central construct [1]. For English majors at the C1-C2 proficiency level, problems at the level of linguistic form are no longer the primary obstacle; the real bottleneck lies in how to accurately extract the core arguments of the source materials within a limited time and, on that basis, formulate an independent commentary with logical force and intellectual depth.

At the same time, the rapid development of generative artificial intelligence is reshaping the ecology of writing instruction. Large language models such as ChatGPT and DeepSeek can instantly generate English texts that are structurally complete and linguistically fluent, which offers unprecedented assistive tools for teaching while also giving rise to the concern of whether students will over-rely on AI and lose their capacity for autonomous writing. Existing studies indicate that AI feedback is markedly effective at the grammatical and lexical levels but shows clear limitations in organizational coherence and depth of content [2-3]; AI suggestions that have not been mediated by teachers tend to be overly general or to deviate from task requirements [4]. In other words, the stronger AI's writing ability becomes, the more students need the metacognitive capacity to judge "whether AI writes well, why it does so, and where it falls short."

Drawing on teaching practice in the "English Writing and Translating Skills Training" course for English majors at a Double First-Class agricultural university, this study explores how ChatGPT and DeepSeek can be transformed into cognitive tools for cultivating students' critical writing abilities in the context of TEM-8 preparation. The core question the study seeks to answer is: in TEM-8 writing instruction for C1-C2 level learners, how should AI-enhanced activities be designed so that technology use shifts from a "ghostwriting tool" to a "medium for thinking training"?

2 TEACHING CONTEXT AND DESIGN RATIONALE

2.1 Learner Characteristics and Course Positioning

The participants in this study are third-year English majors at a Double First-Class agricultural university, whose overall English writing proficiency falls in the C1-C2 range. At this agriculture-focused institution, English majors have a distinctive disciplinary background: on the one hand, the TEM-8 certificate is a key indicator of their professional competence, and the writing section carries considerable weight; on the other hand, students have accumulated some knowledge of agricultural science and technology, ecological policy, and related issues, which provides potential content resources for commenting on materials involving topics such as the environment, food security, and rural revitalization in TEM-8 writing.

The course "English Writing and Translating Skills Training" is positioned as a dual-objective course combining TEM-8 preparation and writing proficiency development. The TEM-8 writing scoring criteria cover three dimensions—content development (10 points), language use (7 points), and textual organization (3 points)—among which content development, including the accuracy of material summarization and the sufficiency of viewpoint commentary, carries the greatest

© By the Author(s) 2026, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

weight. This scoring orientation dictates that instructional design cannot remain at the level of language error correction but must penetrate into argumentative quality and depth of critical thinking.

2.2 Theoretical Positioning of AI's Role

Social cognitive theory and distributed cognition theory provide a framework for the instructional positioning of AI. From a social cognitive perspective, learning is shaped by the interaction among learners, their behavior, and the learning environment, and from a distributed cognition perspective, cognitive activity is mediated and shared across people and tools [5-6]. Building on these perspectives, Zhou's "teacher–student–AI" tri-subject collaborative model proposes that AI should undertake lower-order cognitive tasks in order to release the higher-order cognitive resources of teachers and students , and that the coordinated development of teachers' strategic regulation of AI intervention intensity and students' critical decision-making ability constitutes the core pathway to "technological empowerment—cognitive leap"[7]. This implies that the ideal role of AI in the writing classroom is neither "another teacher" nor "a ghostwriter," but rather a text generator available for scrutiny, comparison, and critique—its value lies precisely in the fact that its output is not perfect, thereby providing students with material for practicing judgment.

On this basis, the study establishes three principles for AI use: first, AI does not replace students in completing the entire process from conception to drafting; second, students must make explicit evaluations of and revisions to AI output; third, teachers screen and re-regulate AI feedback rather than transmitting it directly to students. The third principle is particularly well supported by recent evidence syntheses. Wiboolyasarin and Jinowat report that AI feedback is consistently effective for lower-order writing skills—especially grammatical accuracy and lexical diversity—while the most comprehensive benefits emerge when AI feedback is combined with human feedback [8]. This lends empirical weight to treating teacher regulation not as a replacement for AI but as a complementary layer: AI handles the surface-level dimensions where it is reliable, and teacher screening addresses the organizational and content-level dimensions where AI suggestions are more prone to error.

3 A FOUR-TIER MODEL OF AI-ENHANCED WRITING INSTRUCTION

In teaching practice, the use of ChatGPT and DeepSeek is organized into four tiers, each corresponding to a distinct mode of cognitive engagement, with the learner's role shifting from observer to evaluator, comparer, and finally decision-maker.

3.1 Tier 1: Evaluative Regulation — Teacher-Guided Evaluation of Writing Quality

The first tier of application occurs in the teacher evaluation phase. In traditional TEM-8 writing instruction, the core difficulty teachers face is that class size makes fine-grained marking unsustainable, whereas TEM-8 writing precisely requires in-depth feedback on content logic and argumentative quality. The introduction of AI does not replace teacher judgment but serves as a "pre-screening tool."

In practice, the teacher inputs student essays into DeepSeek, requiring it to provide diagnoses of the three dimensions of content development, language use, and textual organization according to the TEM-8 scoring criteria, and to flag "whether the summary is accurate," "whether the commentary is supported by evidence," and "whether there are logical breaks in paragraph transitions." The teacher then reviews each AI diagnostic report and makes two types of handling: for problems accurately identified by AI (such as missing key points from the materials or grammatical errors), these are directly incorporated into the marking record; for points where AI misjudges or offers only generalities (such as judging a paragraph as "disorganized" without identifying the specific break), the teacher supplements with specific analysis. This process does not reduce the teacher's workload, but it shifts the teacher's attention from "discovering all problems" to "screening and deepening key problems," markedly enhancing the targetedness of feedback.

It is worth emphasizing that the instructional value of this phase lies not in AI itself but in how the teacher models critical engagement with AI. The teacher will display a particular misjudgment in AI diagnosis in class and explain to students: "AI says this argument is 'insufficiently developed,' but it fails to notice that the author actually used a counterexample to qualify the claim—and that counterexample is precisely the highlight. AI cannot read irony." This public critical scrutiny provides a metacognitive demonstration for students' subsequent autonomous use of AI.

3.2 Tier 2: Critical Analysis — Student-Led Analysis of AI-Generated Essays

The second tier of application requires students to examine directly AI-generated essays in the TEM-8 style. The teacher selects writing tasks in actual TEM-8 tests (e.g., the writing task in TEM-8 2022), generates model essays using ChatGPT or DeepSeek, but does not reveal their source to students.

The student task is divided into two steps. The first step is independent scoring: students rate the AI essays according to the TEM-8 scoring criteria and write comments, which must identify at least one strength and one weakness. The second step is comparative verification: the teacher announces that the essay was AI-generated, and students re-examine their own comments, reflecting on "whether I changed my judgment because I knew it was written by AI." This design draws on a core strategy in critical AI literacy instruction: defamiliarizing AI texts so that students first encounter AI as evaluators rather than users.

A typical finding in practice is that students initially tend to give AI essays relatively high scores, with reasons centering on "fluent language" and "clear structure." But after being required to examine argumentative quality paragraph by

paragraph, most students are able to identify the common weaknesses of AI essays: arguments are often presented in a symmetrical "on the one hand… on the other hand" structure, lacking a genuine stance; evidence consists mostly of generalized statements ("research shows," "many people believe"), lacking specific data or cases; and concluding paragraphs often repeat the preceding text with "in conclusion," failing to elevate the level of discussion. One student wrote in a reflective journal: "An AI essay is like a perfect resume—it says everything, but you don't know what this person actually believes." This recognition is precisely the negative reference point for the "clear viewpoint, sufficient evidence" emphasized in TEM-8 writing.

3.3 Tier 3: Comparative Discernment — Cross-Platform Comparison of AI Outputs

The third tier of application extends the object of critique from "AI-generated essays" to "AI platforms themselves." Students obtain generated texts from ChatGPT and DeepSeek on the same topic and conduct a systematic comparison along four dimensions: lexical richness (whether the same expressions are repeatedly used), syntactic diversity (whether simple and compound sentences predominate), argumentative strategy (whether concession or rebuttal is attempted), and "editable space" (which paragraphs can be effectively revised to improve quality, and which need to be scrapped and rewritten).

The design intent of this activity is to dispel the naive assumption that "AI is a unified whole." Through comparison, students gradually realize that differences between models in training data and alignment strategies manifest as differences in textual features: one platform tends to generate longer sentences and more formal vocabulary but is looser in logical cohesion; another produces more concise output but may appear thin in argument development. One group summarized in a class presentation: "ChatGPT writes like a textbook, DeepSeek writes like a student essay—a textbook isn't necessarily suitable for TEM-8, but a student essay is actually easier to revise." Although this judgment is not rigorous, it marks the point at which students begin to treat AI output as an analyzable, attributable, and utilizable textual object rather than an authoritative answer.

3.4 Tier 4: Selective Revision — AI-Assisted Revision of One's Own Essays

The fourth tier of application returns to students' own writing. Unlike conventional "AI grading," this study requires students to follow a "self-assessment first—AI suggestions second—decision last" procedure: after completing a first draft, students first mark three places they are least satisfied with and explain why; they then input the essay into AI, requiring it to provide revision suggestions only for the marked places; finally, students must selectively adopt the suggestions—rejecting at least one AI suggestion and explaining the reason for rejection in writing.

The theoretical basis for this design comes from research on self-regulated writing strategies. The value of large language models in the revision stage lies not in the "correctness" of their suggestions but in the fact that they provide a dialogic feedback object, enabling learners to deepen their understanding of their own writing problems through questioning, challenging, and comparing. What this design addresses, therefore, is not the quality of AI suggestions per se but the learner's capacity to judge them: without explicit training in critical screening, revision is likely to gravitate toward surface-level changes that are easy to accept, while deeper structural problems remain unexamined.

In the context of TEM-8 preparation, the unique value of this stage lies in simulating time pressure. TEM-8 writing requires completing material reading, summarization, and commentary within 45 minutes, leaving students with extremely limited time for autonomous revision. AI's instant feedback enables students to enter the revision cycle quickly after a mock exam, compressing "write—evaluate—revise" within an acceptable time cost. But the teacher repeatedly emphasizes: there is no AI in the examination room; the purpose of using AI now is to internalize certain AI judgment criteria into one's own revision intuition. In end-of-term reflections, several students mentioned that after a semester of training, they would subconsciously ask themselves when writing: "If I let AI revise this, where would it change first?"—and then revise that place themselves first.

4 STRUCTURAL FORMATION OF THE INSTRUCTIONAL MODEL

The four tiers of practice described above are not arranged linearly over the semester but form a spiral, progressive instructional model. Its structure can be summarized as "dual tracks, four stages": the teacher track and the student track run in parallel, with the teacher track centering on evaluative regulation (Tier 1) and the student track progressing through critical analysis, comparative discernment, and selective revision (Tiers 2–4). The four stages as a whole trace a path from evaluating others' AI-generated texts to regulating one's own use of AI, with cognitive demand escalating along the sequence of identification, analysis, judgment, and application.

The function of the teacher track is to model critique and guarantee quality: by publicly scrutinizing the limitations of AI diagnosis, it provides students with a metacognitive template for using AI critically; by screening and supplementing AI feedback, it ensures that the revision suggestions students receive do not deviate from the TEM-8 scoring orientation. The function of the student track is to move from evaluating others to regulating the self: first as a judge of AI essays, then as a comparer of AI platforms, and finally as a decision-maker regarding AI suggestions, with cognitive load escalating progressively along the path of "identifying — analyzing — judging — applying."

This model echoes the tri-subject collaborative model proposed by Zhou but makes adjustments under the specific constraints of TEM-8 preparation [7]: the teacher's regulation of AI is mainly embodied in front-end screening during the evaluation phase rather than accompaniment throughout the whole process, while students' critical practice is

decomposed into operationalizable and assessable specific tasks, so as to accommodate the current metacognitive capacity of C1-C2 level learners.

5 REFLECTIONS AND RECOMMENDATIONS

5.1 Boundaries and Ethics of AI Use

The most prominent problem encountered in practice was not technical malfunction but the translation of differences in students' AI-use proficiency into differences in revision quality. Some students were able to pose precise questions to AI (e.g., "Please identify one key point omitted in the summary of the source material in paragraph 3"), thereby obtaining high-quality feedback; others' prompts remained at the level of "help me revise this," and what they received was naturally generic. This suggests that teachers need to dedicate a specific instructional segment to prompt design early in the course; otherwise, AI enhancement may widen rather than narrow the gap between learners.

Another issue that must be confronted squarely is the boundary of academic integrity. This course explicitly distinguishes between "using AI to revise one's own essay" (permitted and encouraged) and "using AI to generate an essay for submission" (prohibited), but this boundary is not always clear at the operational level. When students adopt sentence structures or even entire sentences provided by AI, does this count as revision or ghostwriting? The teacher's strategy is to require students to attach an "AI use statement" when submitting revised drafts, annotating the source of each AI suggestion and how they handled it. This practice increases administrative burden but renders AI use transparent, which is itself a form of academic integrity education.

5.2 Applicability to English Majors at Agriculture and Forestry Universities

The disciplinary ecology of the institution where this study was conducted has its particularities. English majors at agricultural universities usually have less resource support for TEM-8 preparation than those at foreign language universities; AI tools compensate for this gap to some extent—DeepSeek's free and open access strategy enables students from less advantaged economic backgrounds to obtain high-quality feedback as well. At the same time, the disciplinary context of agricultural universities provides convertible content resources for TEM-8 writing. In the AI comparison activity, the teacher deliberately selected materials related to food security and the protection of agricultural cultural heritage, guiding students to transform disciplinary knowledge into writing evidence. One student, when commenting on vertical farming, introduced specific data from her own undergraduate innovation project on protected agriculture—evidence that never appeared in AI-generated texts because it came from the student's own academic background. This points to a path worth further exploration: AI-enhanced writing instruction should focus not only on "how to use the technology" but also on "what content to use to counter AI's tendency toward generalization."

5.3 Limitations and Future Improvements

The limitations of this study lie first in the absence of a strict experimental design with a control class, which means that judgments of effectiveness rely mainly on qualitative materials and therefore carry limited causal inferential power. Second, AI tools update rapidly, and the capability boundaries of ChatGPT and DeepSeek changed during the semester, challenging the stability of the instructional design. Future research could consider implementing the "critical AI use" model and a "conventional AI-assisted" model in two parallel classes respectively, so as to test the value-added effect of critical use strategies with a more rigorous design.

At the instructional level, the next step for improvement is to embed critical AI analysis more tightly into TEM-8 task-type training. The current four tiers of practice mainly target the "viewpoint commentary" component, while AI assistance for the "material summarization" component has not yet been fully developed. Material summarization requires accurately extracting key points from the source text and restating them in different wording [9]; AI's performance on this task (especially its tendency to make trade-offs during information compression) is itself a sample worthy of student analysis. Combining summarization training with critical AI analysis may be the next entry point for improving the efficiency of TEM-8 writing instruction.

6 CONCLUSION

At a time when generative artificial intelligence is increasingly embedded in language education, the choice facing English writing instruction is not "whether to use AI" but "with what stance to use AI." The practice in this study, conducted in a TEM-8 preparation course for English majors at a Double First-Class agricultural university, shows that positioning AI as a criticizable text generator rather than an authoritative feedback source can make technology serve a more fundamental instructional goal: cultivating students' independent judgment of writing quality. This judgment is the key to obtaining a high score in the TEM-8 essay within the examination room, and beyond it, it is the foundation for English majors to maintain their professional subjectivity in an increasingly intelligent language environment.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This research is supported by the Education and Teaching Reform Project of South China Agricultural University (Grant No. JG2023061).

REFERENCES

[1] Zhang L, Xu H. Reflections on teaching reform of English major: Based on new TEM-8 writing. Zhejiang Gongye Daxue Xuebao (Social Sciences), 2025(2): 224-229.
[2] Cheng X, Zhang L J. Examining second language (L2) learners' engagement with AWE-teacher integrated feedback in a technology-empowered context. The Asia-Pacific Education Researcher, 2024, 33: 1023-1035. DOI: 10.1007/s40299-024-00877-8.
[3] Chen S Y R. Two voices, one draft: a diachronic comparison of human and AI feedback in EFL essay writing. Australian Journal of Applied Linguistics, 2026, 9: 103200. DOI: 10.29140/ajal.2026.103200.
[4] Zhang D, Hoang T, Zhu Y, et al. Boundary Conditions for LLM-Generated Feedback in Primary Writing: An Educator-Aligned Evaluation and Design Considerations. Computers. 2026; 15(6): 393. DOI: 10.3390/computers15060393.
[5] Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Prentice-Hall, 1986.
[6] Hutchins E. Cognition in the wild. MIT Press, 1995.
[7] Zhou H. Reconceptualizing a tri-subject collaborative model for college English writing instruction: Teacher-student-AI collaboration. Yingyu Guangchang, 2025, 34: 109-112.
[8] Wiboolyasarin K, Jinowat N. AI-powered feedback in second language writing: A systematic scoping review of technological innovations in digital language education. International Journal of Language Studies, 2025, 19(4): 85-114. DOI: 10.5281/zenodo.15632738.
[9] Feng M. Reflections and Explorations on Enhancing the Summary Writing Skills of EFL. Exploration of Educational Management. 2025, 3(3):1-5. DOI: 10.12417/3029-2328.25.03.001.