An Iterative Risk-Governed Model for AI-Enhanced Intercultural Writing Instruction

Zhiyan Gao
School of Tourism / School of Artificial Intelligence, Xi'an International Studies University, Xi'an, Shaanxi, China

Abstract: Artificial Intelligence (AI) is rapidly reshaping university language education, yet its integration into second language writing remains fraught with concerns regarding academic integrity and the erosion of critical thinking. This study proposes a structured model for AI-enhanced intercultural writing instruction that balances technological utility with robust risk governance. The model integrates assessable goals, tasks, scaffolds, assessment, and feedback into an iterative closed-loop cycle. By embedding governance strategies across the full instructional cycle, it helps mitigate misinformation, stereotyping and threats to academic integrity. It also ensures that AI serves as a constrained learning tool without adding to instructors' workload. While this model offers a practical pedagogical framework for writing-intensive courses, it requires empirical testing to validate and refine its components.

***Keywords*:** intercultural writing, artificial intelligence, risk governance, intercultural communicative competence

1. Introduction

AI has brought about great changes to language classrooms in recent years. AI chatbots, like ChatGPT and DeepSeek, provide real-time support for tasks difficult for instructors to manage at scale. These features are particularly attractive in required College English courses in China where instructors often face large class sizes, mixed student proficiency levels, and limited time for individual feedback. This challenge is even more pronounced in writing-intensive courses designed to develop intercultural communicative competence (ICC). ICC-oriented writing requires students to explain culturally bound meanings, anticipate audience assumptions, compare perspectives, and avoid stereotypical representations of self and others [1-2]. While AI tools can assist with writing tasks, they also pose risks. They may produce fluent yet inaccurate content or encourage students to outsource interpretive work essential to intercultural learning. For instance, AI-generated responses often flatten cultural nuance or reinforce stereotypes under the guise of fluency. Thus, integrating AI into writing instruction raises urgent questions about academic integrity, misinformation and ethical issues [3-4]. This study therefore explores an effective model for advancing AI-enhanced intercultural writing instruction. This model is designed to clarify learning outcomes, define appropriate scaffolding, and establish classroom risk governance to ensure that AI supports the intercultural interpretive work that students must learn to do themselves.

2. Literature Review

2.1 AI in L2 writing instruction

Researchers in writing studies and applied linguistics are increasingly documenting how AI affects second language writing, with a focus on feedback, learner support, and potential risks [5]. AI tools can enhance writing instruction for both students and teachers [6-7]. However, the educational value of AI depends critically on pedagogical use and becomes clearer when it is assigned specific roles [8]. Using AI as a substitute for composing may boost efficiency at the cost of learning. This risk is especially acute in writing classrooms, where planning, drafting, revising, and reflecting are core to the learning process. In response, scholars have advocated for process-oriented pedagogies, transparent AI use policies and staged drafting requirements to align AI use with instructional goals [9]. Bounded use of AI is more compatible with learning-centered pedagogy because it keeps human judgment and accountability at the center of the writing process.

2.2 Risks of AI use in intercultural writing

The use of AI in intercultural writing introduces risks related to misinformation, stereotyping, and academic integrity. AI can generate inaccurate statements, fabricated references, and misleading generalizations, making errors in cultural content especially difficult for students to detect. At the same time, because AI is trained on large datasets, it often reflects dominant and oversimplified views of cultures, reinforcing stereotypes [10-11] and simplistic binaries that conflict with intercultural teaching's emphasis on nuance and reflexivity. Finally, generative AI blurs the boundary between student-

generated and externally generated text, complicating assessment and making it difficult to determine whether work reflects intended learning outcomes or merely access to better tools [12]. This is particularly problematic in ICC-oriented writing, where rhetorical fluency can mask weak intercultural reasoning or insufficient evidence.

While current scholarship explores AI as a support tool for writing, limited attention has been paid to how its functions can be deliberately constrained to enhance intercultural learning. Risks of AI are well recognized, yet practical governance models for everyday classroom use are largely lacking. To address the gap, this study proposes a closed-loop governance model for AI-enhanced intercultural writing instruction. It centers on two guiding questions: (1) How can AI serve as a constrained tool to translate ICC into assessable goals, tasks, scaffolds, assessments, and feedback? (2) What classroom risk governance strategies are needed to address misinformation, stereotyping, and academic integrity in AI-enhanced intercultural writing instruction?

3. Conceptual Foundations

3.1 AI as a constrained scaffold

The primary conceptual premise of this study is that AI should function as a constrained scaffold for student writing. In education, scaffolding refers to temporary support that helps learners complete tasks while maintaining their own cognitive responsibility. In the context of AI, this support is pedagogically valuable when it aids specific steps, like raising audience questions, clarifying ambiguity, or rephrasing, but becomes problematic when it performs the interpretive and rhetorical work that the assignment aims to develop. This distinction matters because AI can do both, which raises a key pedagogical question: should its use be bounded or not?

3.2 Classroom risk governance

In this study, risk governance refers to the practical strategies through which teachers anticipate, limit, and monitor risks related to AI in classroom conditions. It concerns what can be implemented by instructors during course design, task instructions, assessment, and feedback. Practical classroom instruction is underpinned by four principles: integration, transparency, proportionality, and a learning orientation. Governance should be integrated across the full instructional cycle, supported by clear rules on AI use. Above all, its purpose should be learning-oriented, prioritizing the development of students' habits of critically evaluating and responsibly using AI-generated content.

4. A Closed-Loop Model for AI-enhanced Intercultural Writing

The proposed model (Figure1) conceptualizes AI-enhanced intercultural writing as an iterative cycle comprising five interconnected instructional components: goals, tasks, scaffolds, assessment, and feedback. To ensure responsible AI integration, risk governance strategies are embedded across all stages to mitigate misinformation, stereotyping, and threats to academic integrity. Through the integration of pedagogy and governance, this model positions AI as a constrained tool that supports the critical intercultural reasoning processes essential to ICC development.



Figure 1. A Closed-Loop Model for AI-Enhanced Intercultural Writing Instruction

4.1 Goals

Instructors must begin by clarifying desired learning goals and translating them into specific observable writing behaviors, for example, identifying cultural bias, comparing multiple perspectives, using culturally sensitive language, and supporting claims with verified evidence. These precise goals replace vague outcomes like "improving awareness" and provide a clear basis for task design, assessment and feedback.

4.2 Tasks

Tasks should create a genuine need for intercultural mediation. Overly general prompts, such as "discuss cultural differences", often invite stereotyped or AI-generated responses because they can be answered with shallow content. By contrast, tasks that require students to explain a specific cultural concept to a particular audience, anticipate cross-cultural misunderstandings, or negotiate competing perspectives are more likely to elicit authentic interpretive work and provide richer material for formative feedback.

4.3 Scaffolds

Such scaffolds as guiding questions, source evaluation checklists, and structured peer-review protocols help students navigate the interpretive demands of intercultural writing. These supports should be designed to fade over time, encouraging students to internalize strategies and apply them independently. At this stage, AI is constrained in three ways. Functionally, students may use AI only for specified purposes. In terms of contents, AI outputs cannot serve as evidence for factual or cultural claims and must be verified through approved sources. As for procedures, evaluation includes drafts, verification logs, and reflective accounts. These constraints ensure that AI functions as a collaborative assistant in student writing. To maintain academic integrity, instructors must provide students with model prompts and explicitly prohibit the use of AI to generate full drafts or unsupported cultural claims.

4.4 Assessment

In AI-mediated writing, the final text alone is insufficient to demonstrate learning. Instructors can require a comprehensive portfolio of process evidence: an unassisted baseline draft, a revised final essay, a source verification log, an AI-use disclosure, and a reflective statement on how AI suggestions were managed. This method pushes students to reflect on their writing while making learning visible. It also deters outsourcing and helps instructors gauge whether AI aided or undermined learning.

4.5 Feedback

Feedback closes the instructional loop by translating performance data into guidance for improvement. Evidence gathered through assessment feeds back into course design. If many students rely on essentialist language, the teacher may revise future scaffolds to include counter-stereotyping prompts. If verification logs show confusion about evidence, source instruction may need strengthening. If students misuse AI despite disclosure rules, assignment wording or tool restrictions may need revision. The loop is therefore closed at two levels: students revise their texts, and instructors revise their designs.

4.6 Governance across the loop

To ensure academic integrity and pedagogical control, this model embeds risk governance at every stage of the writing process. At the Goals stage, the instructor ensures all AI-integrated tasks are strictly aligned with specific intercultural competence outcomes. During the Task phase, students must establish an unassisted baseline through an in-class draft, preventing over-reliance on AI from the outset. Throughout the Scaffolding phase, students are required to maintain a verification log, creating a mandatory audit trail that links every AI-generated claim to reliable sources. The Assessment stage utilizes active instructor review, requiring the instructor to evaluate both the final essay and the process documentation to identify potential AI hallucinations or biases. Finally, the Feedback stage requires students to submit a reflection on their AI collaboration, forcing them to critically evaluate how they managed, challenged, or corrected AI suggestions. By positioning the teacher as the designer and guarantor of evidence, this framework guarantees that AI remains a constrained tool for critical thought.

5. Implications

5.1 Pedagogical implications

The model's pedagogical implications can be understood in terms of the responsibilities it creates for different stakeholders. For instructors, the model requires a shift from generic teaching to intentional design. They must translate ICC into specific, observable writing behaviors and design tasks that demand genuine intercultural mediation rather than surface-

level responses. They are also responsible for making the writing process visible through checkpoints and providing timely, process-oriented feedback that guides students' interpretive work. For students, the model asks them to take ownership of their writing process. This means documenting their revisions, reflecting on their use of AI, and critically evaluating AI-generated content. In doing so, students develop not only ICC but also the habits of critical thinking essential for responsible AI use. For course designers, the model implies that AI literacy should be integrated as a core component of intercultural writing instruction.

5.2 Classroom risk governance implications

In the proposed model, risk governance relies on practical classroom procedures that are sustainable for everyday teaching. By distinguishing between different types of AI use, the model allows instructors to apply governance strategies proportionally, targeting the stages where risks are most acute. It addresses three main risks through concrete, low-burden strategies. First, it mitigates misinformation by requiring students to verify all sources. Students must locate and cite original sources for any factual claims, statistics, or cultural descriptions generated by AI. This practice encourages inquiry and reinforces the ICC habit of grounding claims in evidence. Second, it counters stereotyping through assignment design. Rather than relying on generic prompts that invite surface-level or AI-generated responses, tasks are designed to push students beyond clichés, thus reducing stereotypical outputs and deepening students' intercultural reasoning. Third, it protects academic integrity by requiring students to submit both their final work and a record of their writing process. This approach shifts the focus from policing AI use to fostering student accountability, while providing instructors with clear evidence of student thinking. Students should also be made aware of these expectations from the outset, so they understand what constitutes responsible AI use in the course. By integrating risk governance into the routine practices of teaching, the model enables instructors to manage AI-related risks without sacrificing instructional time or pedagogical focus.

6. Conclusion

This paper proposes a structured model for AI-enhanced intercultural writing instruction. Its primary contribution is a risk governance framework that mitigates misinformation, stereotyping, and threats to academic integrity across the instructional cycle. Complementing this is a pedagogical framework that translates ICC into assessable goals, tasks, scaffolds, assessments, and feedback. This practical framework repositions AI from a potential threat to a pedagogical resource supporting student learning. However, as a conceptual framework, this model requires empirical validation. It is intended as a flexible guide for College English and L2 writing contexts and future research is needed to examine and refine its components.

Acknowledgments

This work was supported by the Teaching Reform Project of Xi'an International Studies University (25BYSZ01).

References

[1] Byram, M. (2021). Teaching and Assessing Intercultural Communicative Competence: Revisited (2nd ed.). Multilingual Matters & Channel View Publications. https://doi.org/10.2307/jj.22730614.
[2] Jackson, J. (2020). Introducing language and intercultural communication (2nd ed.). Routledge.
[3] Sachwani, D. (2025). The role of ChatGPT in education: Opportunities, challenges, and future directions. International Journal of Applied Research, 11(9), 203-206.
[4] Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. Education Sciences, 13(4), 410. https://doi.org/10.3390 educsci13040410.
[5] Almusharraf, A. (2026). AI chatbots in L2 writing: anxiety, usefulness, and writing outcome. Humanities and Social Sciences Communications, 13, 393. https://doi.org/10.1057/s41599-026-06816-w.
[6] Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing, 57, 100745. https://doi.org/10.1016/j.asw.2023.100745.
[7] Luo, Y., Xia, Y., & Lu, X. (2025). A Systematic Review of the Role of Artificial Intelligence in Second Language Writing Education. Digital Studies in Language and Literature, 2, 302 - 325.
[8] Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching & Learning Practice, 20(2). https://doi.org/ 10.53761/1.20.02.07.
[9] Hui Wang, Anh Dang. (2024). Enhancing L2 Writing with Generative AI: A Systematic Review of Pedagogical Integration and Outcomes. System, 10: 1-30.

[10] Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334), 183–186.
[11] Zhu, H., Sun, Y. & Yang, J. (2025). Towards responsible artificial intelligence in education: a systematic review on identifying and mitigating ethical risks. Humanities and Social Science Communications 12, 1111. https://doi.org/10.1057/s41599-025-05252-6.
[12] Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching & Learning Practice, 20(2). https://doi.org/ 10.53761/1.20.02.07.

Author Bio

Zhiyan Gao (born December 1985), female; Han Chinese; hometown: Yan'an, Shaanxi; educational level: MA, English Language and literature, Xi'an International Studies University; title: lecturer; research interests: English teaching, intercultural communication, American environmental history.