DOI: 10.66069/ojspub.5170261013

Teaching With AI, Assessing with Uncertainty: Chinese English Teachers' Practices and Dilemmas

Bacui Chen

School of Foreign Studies, Lingnan Normal University, Zhanjiang, Guangdong, China

Abstract: ***AI tools offer English teachers resources for lesson preparation, classroom interaction, and feedback, while also complicating the interpretation of students' work. This qualitative study examines everyday classroom practices and assessment tensions reported by 11 English teachers from 3 universities in southern mainland China. Chinese-language interviews, follow-up exchanges, and teaching documents and artifacts formed the wider project's evidence base. Drawing on its boundary-crossing-informed and inductive thematic analysis, this article focuses on classroom coordination and reflection. Teachers described combining AI platforms for different tasks, incorporating generated materials into guided activities, and using AI-supported feedback alongside classroom evaluation. These arrangements were accompanied by concerns about unreliable outputs, students' off-task device use, and the difficulty of identifying language-learning needs in AI-assisted writing. The findings show how opportunities for efficient support and creative activity coexist with uncertainty about what student products reveal. Teachers' responses involved continuing to organize interaction, interpret outputs, and examine student work. The study contributes a situated account of these human–AI arrangements and suggests that classroom guidance should connect permitted AI assistance with opportunities to make students' reasoning and language use visible.***

Keywords: Artificial intelligence, English teachers, Classroom practices, Writing assessment, Teacher judgment.

1. Introduction

Artificial intelligence (AI) has become an important resource in higher education teaching. More recently, generative AI (GenAI) tools have offered opportunities for lesson planning, facilitating brainstorming, preparing examination questions, and providing individualized feedback (Belkina et al., 2025). In the field of foreign language learning, GenAI has potential to support dialogue practice through chatbots, vocabulary enhancement, and grammar improvement (Law, 2024; Lee et al., 2026). These possibilities raise a practical question for teachers: how can AI-generated material and feedback be incorporated into activities that remain directed toward students' language learning?

This question is relevant to English teachers' work with varied language-learning needs and the demands of classroom instruction and assessment. Research in China has highlighted these demands in English teaching for non-English-major students (Du, 2012; Wen, 2014). AI may offer useful assistance where teachers must prepare activities and respond to substantial amounts of student writing. However, the availability of generated content does not determine how that content should be used, how students should interact with it, or what teachers can infer from the resulting work. A polished text may be convenient to read while providing limited evidence of the writer's independent language use.

Research on AI in English language teaching has examined both its applications and the challenges accompanying their use (Jiang, 2022; Law, 2024; Lo et al., 2024). For classroom practice, these two concerns need to be considered together. The same tool that helps teachers provide a rapid response may also contribute to uncertainty about the source and meaning of students' answers. Likewise, allowing access to devices for AI-assisted learning creates a need to manage attention during the activity. Understanding these arrangements requires attention to teachers' accounts of ordinary instructional decisions, alongside the outputs that AI can produce.

This article examines how English teachers describe organizing AI-assisted teaching and responding to the tensions it creates. It focuses on the connection between practical uses of AI and teachers' judgments about the resulting activities and student work. The aim is to document situated arrangements, including the decisions through which teachers connect platform capabilities with classroom purposes. It also examines where these arrangements encounter difficulties, particularly in writing assessment. The study addresses two research questions: (1) How do English teachers incorporate AI into lesson preparation, classroom activities, and assessment? (2) What classroom and assessment tensions do they report, and how do they respond?

The article reports the classroom practice and assessment strand of a wider qualitative project, which also examines teachers' professional roles and learning. Its contribution is a focused account of task-specific tool combinations, teacher-guided activities, and the assessment uncertainties associated with AI-assisted student work.

2. AI-Assisted English Teaching and Classroom Contexts

Studies of AI in language education describe a range of uses across teaching and learning activities. Reviews of ChatGPT and other generative applications cover opportunities for interaction, content generation, and language support (Law, 2024; Lee et al., 2026; Lo et al., 2024). Work on classroom writing also considers how students and teachers can collaborate with AI during the writing process (Su et al., 2023). Together, these studies provide a basis for examining how AI assistance is embedded in instructional tasks, rather than treating access to a tool as a complete pedagogical arrangement.

Embedding AI in a task involves decisions about the purpose of the activity, the contribution expected from students, and the interpretation of the output. A generated outline may provide material for discussion; a generated response may

supply feedback for further evaluation. These uses place different demands on teachers and students. They also make the distinction between support for learning and evidence of learning consequential. The present study examines this distinction through teachers' accounts of how they use AI and the problems they encounter when interpreting completed assignments.

To understand the practical challenges of AI integration, it is necessary to consider English teachers' working context. Teachers must respond to differences in student proficiency while balancing curricular demands for general English and more specialized English learning (Du, 2012; Wen, 2014). In the present study, participants described using several accessible AI platforms, including DeepSeek, Kimi, and Doubao. Their accounts concerned specific teaching purposes rather than a single shared institutional application. An examination of classroom practice therefore needs to consider how teachers combine tools and connect activities across preparation, classroom interaction, and assessment.

AI use can also introduce demands of its own. Literature on AI in EFL education identifies opportunities alongside practical and ethical challenges, including questions about the reliability and use of generated content (Jiang, 2022; Omidvar & Meihami, 2025). For teachers, an output must be considered in relation to the instructional purpose it serves. In this study, the relevant concerns included the plausibility of AI responses, the management of devices during lessons, and the difficulty of evaluating writing that has already been generated or revised with AI. These concerns guide the article's focus on how teachers arrange human and AI contributions within particular activities.

3. Methods

The present study draws on a wider qualitative interview project. Data were collected from 11 English teachers from three universities (Universities A, B, and C) offering English lessons to undergraduates in southern mainland China. Universities A and B were regional public universities, while University C was a private institution. Eleven English teachers, identified as T1–T11, were purposively sampled to participate on a voluntary basis. Selection followed the principle of maximum variation, including teachers at different stages of their careers and from different institutional affiliations (Merriam & Tisdell, 2015; Patton, 2014). Selection was based on their willingness to discuss AI integration experiences, regardless of their current level of use. Their reported experience teaching English to undergraduate students ranged from 1 to 30 years, and their positions ranged from assistant lecturer to associate professor. A companion manuscript on professional roles and learning draws on the same participants, interviews, and artifact collection. The present article focuses on classroom practices and assessment tensions.

In-depth semi-structured interviews formed the primary data source. Individual interviews were conducted over a period of one month, with each initial interview lasting 60–90 minutes. They were conducted in Chinese, the interviewees' and interviewer's native language, to facilitate participants' expression of their AI integration experiences. Follow-up questions were asked through several means: T1, T2, and T11 participated in second individual interviews lasting less than sixty minutes, while T3, T4, T5, T6, T7, and T10 responded via text or voice messages through WeChat. T8 and T9 did not receive follow-up questions because the interviewer considered their initial responses sufficient. All interviews were audio-recorded and transcribed verbatim.

Besides interview data, the wider study collected documents and artifacts from English teachers and their students to contextualize the accounts (Merriam & Tisdell, 2015). These materials included lesson plans incorporating AI tools, teaching activities developed with AI, screenshots of AI interactions, student submissions involving AI, and AI-generated digital personas. The findings presented here center on teachers' accounts of classroom practices and assessment tensions.

The wider project adopted thematic analysis combining inductive and deductive elements, applying boundary-crossing mechanisms (Akkerman & Bakker, 2011) while remaining open to emerging themes. Two researchers read the interview transcripts to locate relevant episodes. Coding was conducted iteratively, with codes examined and refined through comparison with documents and artifacts. Coding discussions and consensus were used to establish consistency. This article reports the classroom-practice strand developed within that analysis, particularly accounts of coordination and reflection. The findings bring these accounts together around tool combinations and guided activities, feedback and evaluation, and practical tensions.

Data collection began following research ethics approval and included informed consent and anonymity protection through participant codes and institutional pseudonyms. The reporting focuses on teachers' experiences and interpretations; no student outcome measures or independent assessment of AI-system accuracy were included.

4. Findings

The classroom accounts reported here show how English teachers integrated AI tools into daily teaching and reflected on the opportunities and constraints involved. Four themes capture the classroom strand: selecting tools for different purposes, developing guided activities, coordinating feedback and evaluation, and managing practical and assessment tensions.

4.1 Selecting and Combining Tools for Teaching Tasks

English teachers' coordination efforts involved integrating multiple AI platforms, managing time allocation, and developing assessment protocols. Through these efforts, teachers developed routines for using AI in daily teaching. Initially, this coordination concerned combining platforms and digital tools for particular tasks. T1 explained: "I prefer DeepSeek for language processing and use Kimi for collecting teaching materials."

T7 similarly described using two applications for different aspects of language learning: "I use Doubao intelligent agent to produce texts for students to practice oral dialogue and use DeepSeek to answer students' questions."

These arrangements extended from practical tool combinations to pedagogical design. Teachers allocated different tasks to AI platforms according to their teaching purposes. T7, for example, connected generated texts with oral dialogue practice and used another tool to answer questions. Preparing a text and responding to questions thus formed distinct contributions within the teacher's arrangement of resources.

The accounts show how teachers coordinated platforms rather than relying on a single application for all teaching needs. The teacher connected these resources with the activity and determined how students would encounter the outputs. Tool selection formed one component of the reported teaching routines, alongside the organization of classroom interaction and evaluation.

4.2 Incorporating Generated Materials Into Guided Activities

Coordination also involved combining AI-generated content with classroom interaction. As teachers incorporated AI-enhanced activities, they organized what students would do with the outputs. T5 described a sequence of idea generation and pair discussion: "I ask students to use AI to generate ideas, then discuss in pairs."

In T5's writing class, AI-generated outlines provided structured content for a task on making online friends. The teacher described guiding students' critical thinking and quality evaluation. The activity connected the generation of material with further consideration of its content, as in the sequence of generation and discussion described above.

Teachers also reflected on AI's affordances for teaching innovation. T4 described the moment when AI provided a creative idea for an activity: "The idea that AI provided was so creative, so brilliant, totally unexpected!"

T4's response shows an appreciation of ideas that the teacher had not anticipated. The perceived benefit concerned extending the material available for lesson design.

These practices indicate how teachers actively created routines that connected AI assistance with existing classroom work. Through organizing generated ideas, discussion, and activity materials, they developed feasible arrangements for AI use. The outputs became part of a teacher-guided activity whose purpose extended beyond obtaining a response from the system.

4.3 Coordinating Feedback, Classroom Evaluation, and Writing Review

Coordination extended to classroom feedback and after-class assessment. T3 described using immediate responses to identify students who might need attention during a lesson: "AI-enabled immediate feedback. When I hand out some questions to test students, I can see who answered incorrectly right away and monitor these students in class, which can promote higher attention."
This account links timely feedback with the teacher's monitoring of students during class. Information about incorrect responses allowed the teacher to direct attention to particular students. T3 regarded this timing as useful for classroom participation.

T5, who also taught writing, described the benefits of AI-supported assessment in relation to a substantial teaching workload: "As I teach 6 classes, each with 60-70students…AI is a good way that I can immediately get a complete score description, which helps me identify the problems with less effort."

The reported benefit concerned obtaining a complete score description quickly and using it to identify problems with less effort. T5's account helps explain why assistance with feedback was attractive when teaching several classes. The output formed a resource for the teacher's examination of writing.

T3 had also developed an assessment arrangement combining AI use, AI-assisted detection and evaluation, and classroom analysis:

I increased my spot checks this semester. I took the articles out for analysis, using AI for detection and evaluation. Then I asked the class to analyze and evaluate together. In this way, they can have an interaction with AI and with their writing.

The arrangement combined spot checks with opportunities for students to analyze and evaluate writing together. Through balancing AI use and teacher-student evaluation across classroom and after-class work, T3 developed a routine for examining student texts and discussing them with the class.

4.4 Managing Unreliable Outputs, Device Use, and Uncertain Authorship

Teachers' reflections also showed concerns about AI outputs, classroom supervision, and academic integrity. T3 expressed apprehension about the reliability of generated content: "Many of AI outputs are quite illusory or unreal…so how to dialectically apply it, think about it, and to judge its rationality is a major challenge for us."

This reflection emphasizes the continuing need to judge the rationality and suitability of AI output. Obtaining a response did not resolve whether that response could serve the teaching purpose. Teachers therefore faced the practical demand of evaluating the material available to them.

This concern was accompanied by classroom-management difficulties: when students used devices in class, it became hard to supervise their participation. T4 reflected on this dilemma: "To use AI in class, you must let them use phones. But when they use phones, you cannot judge whether they are using phones to study or to entertain."

T5 supplied a concrete example of this difficulty:

Once I assigned an activity to finish in the learning platform in class, I went down to check and saw a student had opened Douyin (Chinese TikTok). I got very angry and said that you should open the learning app, not Douyin!

The concern was also evident in T5's account of a student opening an entertainment application during an assigned activity. Allowing access to digital tools created a need to distinguish participation in the task from unrelated device use.

Concerns about academic integrity and fairness were particularly evident in writing assessment. T5 discussed the difficulty of identifying students' language problems in an AI-assisted final text:

If they do not write the article themselves first, but use AI to write, polish, and revise before submitting, when I see the final version, it's no longer their original writing—I have no way to analyze what their actual problems in writing are.

This account shows how AI assistance created a challenge for evaluating students' actual language proficiency. T5 wanted to identify the problems requiring feedback, but the submitted version obscured the contribution made by the student. T1 also raised a concern about unfair rewards: "Those using AI get higher scores...dishonesty brings benefits."

These reflections show teachers examining both the affordances and constraints of AI integration. They valued AI's assistance while questioning what a completed text could reveal about the student's work. T1's statement reports a perceived fairness problem. Across the writing accounts, teachers sought a basis for interpreting students' contributions and identifying the learning needs that required attention.

5. Discussion

The findings connect the practical appeal of AI with the continuing work of organizing instruction and interpreting evidence. Reviews identify uses of AI across language-learning tasks and writing support (Law, 2024; Lee et al., 2026; Lo et al., 2024). The present accounts show how teachers put such possibilities into practice by assigning different tools to different tasks, incorporating generated content into activities, and combining rapid responses with classroom evaluation. These were situated human–AI arrangements: the tools supplied content or feedback, while teachers determined where and how those contributions entered instruction.

One implication is that tool capability and classroom use need to be considered together. The combination of DeepSeek, Kimi, and Doubao reflected teachers' choices about language processing, resource collection, and dialogue practice. T5's sequence of generation followed by pair discussion further showed how a tool output could become material for student interaction. Guidance for teachers can build on this relationship by addressing complete activities: what the tool contributes, what students are expected to do with that contribution, and what the teacher examines afterward.

A second implication concerns the relationship between feedback efficiency and assessment interpretation. T5 valued the reduced effort involved in obtaining a detailed response to writing, while T3 described information that could be used promptly during a lesson. These perceived benefits help explain the attraction of AI assistance under a substantial teaching workload. At the same time, a response generated quickly may still require evaluation for relevance and accuracy. The teachers' concerns therefore place judgment inside the use of AI-supported feedback, rather than treating it as an issue resolved by the availability of a response.

Writing made the distinction between an assisted product and evidence of learning especially visible. T5 wanted to identify students' actual language difficulties but found that a final text could obscure them after AI generation and revision. This concern complements work on collaboration with AI in writing classrooms (Su et al., 2023) by drawing attention to what the teacher needs to learn from the writing process. The practical issue is how to preserve opportunities for students to demonstrate their thinking and language use when AI assistance is permitted. Clear task instructions about the purpose and extent of assistance, together with occasions to discuss choices or show intermediate work, follow from this concern as options for teaching design. Their effects would require separate investigation.

The reported use of AI-assisted detection also requires careful interpretation: this study did not test whether such outputs reliably identify AI-generated writing. T3's account nevertheless shows how classroom analysis and spot checks created further occasions to examine work with students. Assessment guidance should explain how evidence from the learning process will be considered, including how students can account for AI assistance. This addresses the diagnostic difficulty identified by T5 while keeping the interpretation of automated output within the teacher's evaluative responsibilities.

Finally, AI-assisted activity involved familiar classroom concerns in a changing resource environment. Access to phones enabled participation in digital tasks while also making off-task activity possible. T4 and T5 described this issue at the level of classroom management, and T3 raised a parallel need to examine the plausibility of generated content. These accounts support practical attention to task organization, the visibility of student participation, and discussion of outputs. Such attention connects the opportunities for AI-supported language learning described in the literature with the conditions under which teachers can interpret and respond to classroom work (Jiang, 2022; Omidvar & Meihami, 2025).

6. Conclusion and Limitations

This study examined classroom practices and assessment tensions reported by 11 English teachers using AI. Teachers described combining platforms for different purposes, incorporating generated materials into guided activities, and connecting AI-supported responses with classroom and writing evaluation. These arrangements were accompanied by concerns about unreliable content, device use, fairness, and the difficulty of diagnosing language needs from extensively assisted writing. The findings suggest that useful guidance should connect AI assistance with clear instructional purposes and opportunities for students to make their contributions visible.

The study is limited to a small, voluntary sample from three universities in one region and to the classroom strand of a wider qualitative project. The findings presented here center

on teacher accounts. It provides no direct measure of learning improvement, the accuracy of feedback or detection, or the relative effectiveness of the reported arrangements. Student perspectives and classroom observations would help examine how these practices are experienced and enacted. Further research could also investigate whether process-based opportunities to explain and revise AI-assisted writing improve the evidence available for diagnosis and feedback.

Acknowledgments

This work was supported by Guangdong Planning Office of Philosophy and Social Science (Grant Number: GD24XJY42) and Department of Education of Guangdong Province (Grant Number: 2023GXJK118). The funded projects are "Sustainable Career Development of Translation Teachers in the Guangdong–Hong Kong–Macao Greater Bay Area in the New Era" (GD24XJY42) and "Exploring Innovative Models of Translation Education in Guangdong, Hong Kong, and Macao in the Context of Guangdong's Path to Chinese Modernization" (2023GXJK118).

References

[1] Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research, 81*(2), 132–169. https://doi.org/10.3102/0034654311404435

[2] Belkina, M., Daniel, S., Nikolic, S., Haque, R., Lyden, S., Neal, P., Grundy, S., & Hassan, G. M. (2025). Implementing generative AI (GenAI) in higher education: A systematic review of case studies. *Computers and Education: Artificial Intelligence, 8*, Article 100407. https://doi.org/10.1016/j.caeai.2025.100407

[3] Du, H. (2012). College English teaching in China: Responses to the new teaching goal. *TESOL in Context*, Special Edition S3, 1–13. https://tesol.org.au/wp-content/uploads/2019/10/TiC_S3_hui_du.pdf

[4] Jiang, R. (2022). How does artificial intelligence empower EFL teaching and learning nowadays? A review on artificial intelligence in the EFL context. *Frontiers in Psychology, 13*, Article 1049401. https://doi.org/10.3389/fpsyg.2022.1049401

[5] Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open, 6*, Article 100174. https://doi.org/10.1016/j.caeo.2024.100174

[6] Lee, S., Choe, H., Zou, D., & Jeon, J. (2026). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments, 34*(1), 335–359. https://doi.org/10.1080/10494820.2025.2498537

[7] Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S.-Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments, 11*, Article 50. https://doi.org/10.1186/s40561-024-00342-5

[8] Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

[9] Omidvar, S., & Meihami, H. (2025). Exploring the 'what' and 'how' of opportunities and challenges of AI in EFL teacher education. *Computers and Education: Artificial Intelligence, 9*, Article 100443. https://doi.org/10.1016/j.caeai.2025.100443

[10] Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). SAGE.

[11] Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57*, Article 100752. https://doi.org/10.1016/j.asw.2023.100752

[12] Wen, Q. (2014). The controversy between general English and specialized English in university English teaching: Problems and solutions. *Foreign Languages and Their Teaching*, (1), 1–8.