



ChatGPT or “CheatGPT”? EFL Teachers’ perceptions of scoring AI-assisted English writing through the Lens of assessment literacy

JIANFENG ZHANG¹ 

¹Shanghai International Studies University, China

Abstract

The rapid integration of artificial intelligence (AI) into education has reshaped teaching and learning practices, bringing unprecedented challenges to traditional approaches to writing assessment. Grounded in the three-dimensional framework of teacher assessment literacy, which encompasses the conceptual, praxeological, and socio-emotional dimensions, this study explores nine Chinese EFL teachers’ perceptions of assessing AI-assisted writing assignments. Specifically, it focuses on issues of academic integrity, rubric compatibility, and criterion development in the AI era. Adopting a qualitative design, data were collected through semi-structured interviews to examine teachers’ beliefs, practices, and emotional stances toward AI-mediated assessment. The findings reveal that while teachers acknowledge AI’s potential to enhance learner engagement and higher-order thinking, they express persistent concerns about academic integrity and the inadequacy of existing scoring rubrics. Their attitudes toward AI-assisted writing are influenced by varying levels of digital literacy and pedagogical commitment, with most maintaining traditional scoring orientations that prioritize linguistic accuracy and originality. The study highlights the need to strengthen EFL teachers’ assessment literacy, particularly digital and ethical competencies—through professional development initiatives. It also underscores the necessity of updating institutional guidelines to ensure valid, fair, and pedagogically aligned writing assessment in the AI era.

Keywords: English writing, AI, academic integrity, scoring perceptions, EFL teachers

Introduction

English writing is one of the most challenging tasks for English as a foreign language (EFL) learners. The integration of artificial intelligence (AI) has been posited to enhance students’

writing in terms of both language and content quality, as well as foster higher-order thinking, given its ability to provide personalized feedback (Fariani et al., 2023; Liang et al., 2024) and self-regulated learning (Song & Song, 2023; Teng, 2025). However, this technological advancement presents significant challenges to the traditional paradigm of writing assessment, particularly for educators. It may compromise academic integrity and render existing rubrics obsolete (Escalante et al., 2023; Rasul et al., 2024). These issues are exacerbated in the EFL context, where learning is confronted with heightened linguistic barriers, and where teaching and assessment strategies inherently emphasize language proficiency. As noted by Alharbi (2023), the potential for writing assignments refined by AI has called into question the legitimacy of language-centeredness, sparking widespread concern among learners and practitioners.

For EFL teachers, assessing AI-assisted writing remains a complex technical conundrum (Kahveci, 2023) that is beyond linguistic concerns (clearness, coherence, content completion, richness of expression, flow of logic), scoring consistency and fairness, as well as the holistic and analytical feedback (East & Slomp, 2024; Salmani Nodoushan, 2014; Shabani & Panahi, 2020). To date, the academic community has shown a greater interest in the role of teachers as instructors rather than as evaluators within AI-assisted writing curricula (Ali et al., 2024). This oversight definitely leads to the insufficient understanding of teachers’ attitudes towards assessing AI-assisted student writing.

Considering an EFL perspective is essential to broaden pedagogical understanding and enhance teaching efficacy, particularly as EFL writing instructors grapple with both the universal challenges posed by AI and their unique linguistic dilemmas. This study interviewed nine Chinese university EFL teachers to unpack their perceptions towards the assessment of student AI-assisted English writing.

Literature Review

Three-Dimensional Teacher Assessment Literacy Framework

The term assessment literacy was first proposed by Stiggins (1991), who defined it from the perspective of knowledge and skills, emphasizing that assessment literacy refers to assessors’ ability to determine the reliability of assessments, understand assessment content, methods, and issues, and recognize how to avoid potential problems in assessment practices. From a sociocultural perspective, Gee (2004) reconceptualized assessment literacy as a dynamic and context-dependent social activity. Looney et al. (2017) further incorporated socio-emotional factors into the construct of assessment literacy and defined it as teacher assessment identity. Taken together, the definition of assessment literacy should be considered from a multidimensional perspective. It refers to the competence and professional quality that enable teachers to possess appropriate assessment beliefs, knowledge, and skills, and to apply assessment results effectively in the design and implementation of assessment plans.

Considering that language teaching involves both linguistic knowledge and the development of communicative competence, Fulcher (2012) systematically described a framework for language assessment literacy, defining it as language teachers’ ability to design and develop language tests, demonstrate knowledge of test procedures, and understand the concepts and principles underpinning assessment practices, while situating such knowledge, skills, and principles within broader historical, social, political, and philosophical contexts. With the increasing prevalence of the Assessment for Learning (AfL) paradigm, the connotation of assessment literacy has expanded from a “technical and procedural level” to a more holistic conception that integrates teaching, assessment, and learning.

In response to this conceptual evolution, Pastore and Andrade (2019) proposed a three-dimensional framework of teacher assessment literacy, comprising the conceptual, praxeological, and socio-emotional dimensions, thereby forming an organic chain of “cognition-action-emotion.” Within the conceptual dimension, teachers are expected to understand what assessment is, why assessment is conducted, what is assessed, and how assessment is implemented (Popham, 2018). This dimension focuses on teachers’ understanding of assessment theories, models, and principles, including the purposes of assessment (e.g., summative, diagnostic, formative), the range of assessment methods (from classroom observations to large-scale standardized testing), and the assurance of assessment quality. The praxeological dimension conceptualizes assessment literacy as teachers’ ability to integrate assessment practices with instruction, reflecting their capacity to transform theoretical assessment knowledge into pedagogical action (DeLuca et al., 2016). The socio-emotional dimension, by contrast, views assessment as a social practice, emphasizing the emotional and ethical dimensions of assessment activities (Boon 2011; Gülcan, 2015). Teachers are therefore required to be aware of their professional role as assessors, to attend to issues of trust, responsibility, and rights, and to protect students’ data privacy. This dimension also entails teachers’ ethical sensitivity to potential risks such as teaching to the test, cheating, and inequitable assessment practices (Xu & Brown, 2016).

Overall, the three-dimensional framework proposed by Pastore and Andrade (2019) contributes to the systematization and contextualization of assessment literacy and provides a clear and comprehensive structure for teacher education and professional development.

English Writing in the AI Era

The development of English writing competence is crucial for both academic and professional success. Writing tasks demand deep engagement from students, thereby enhancing their English proficiency, sharpening their critical thinking, facilitating the organization of ideas, and deepening their understanding of complex issues (Flowerdew & Peacock, 2001). Moreover, English writing is essential for career advancement, as it not only enhances global communication but also enables individuals to handle specific tasks such as drafting proposals, preparing presentations, and communicating with clients (Lillis & Curry, 2024).

The explosive development of artificial intelligence has revolutionized the landscape of English writing. AI technologies have significantly bolstered human writing productivity through streamlined writing processes and optimized prompt output (Papakonstantinidis, 2024). These tools spare writers the need for manual language checks, provide instant revision suggestions, and notably enhance the readability of texts (Liu et al., 2019). Furthermore, AI can facilitate the ideation process by generating content or expanding on ideas for a given topic (Yan et al., 2024). This allows writers to focus on more demanding tasks that require greater creativity and in-depth analysis.

However, the integration of AI into educational settings also presents several challenges. The availability of ready-made answers can lead to over-reliance on AI tools, which may result in a decline in students’ writing skills. Excessive dependence on AI may also lead to reduced originality, ethical dilemmas, and diminished cognitive skills (Niloy et al., 2024; Teng, 2024). From the educators’ perspective, teachers express distrust in AI tools, such as automated writing evaluation (AWE) systems which may fail to accurately assess writing quality due to biases against diverse linguistic styles (Boulanger, 2024). Additionally, teachers may be reluctant to adapt their teaching methods to an AI backdrop, fearing that AI could distract students from traditional teaching focuses (Chhabra & Sameer Babu, 2024). These issues intensify academic concerns regarding writing instruction in the AI era, particularly in relation to teachers’

attitudes and practices concerning safeguarding academic integrity, applying scoring rubrics, and upgrading assessment criteria.

Academic Integrity in English Writing

Academic integrity is the ethical cornerstone for the sustainable cultivation of writing skills, yet the prevailing conception of it appears to be insufficiently considered in the face of AI technological advancement (Yeo 2023). Macfarlane et al. (2014) delineates academic integrity as the truthful representation of one’s own work, which includes acknowledging the contributions of others through proper citation and employing accurate data. The emphasis on plagiarism reflects a traditional focus on preserving the integrity of intellectual exchange among scholars. However, current studies have not fully addressed the new challenges posed by the AI revolution, which threatens to bypass established norms of presentation and citation, thereby infringing upon the core values of originality, truthfulness, and transparency (Song & Song, 2023).

The incorporation of AI into the writing process challenges the authenticity of originality due to the ambiguous delineation between human and machine authorship (Rasul et al., 2024). Relying heavily on AI assistance may impede the development of linguistic proficiency, writing skills, and critical thinking, with the most significant risk being the erosion of students’ interpersonal ability to articulate their thoughts in their own voice (Mirón-Mérida et al., 2024). A resolution to the issues of originality and truthfulness in writing is contingent upon achieving transparency in AI-assisted writing. However, the successful detection of AI-generated content remains elusive, as conventional anti-plagiarism tools, which rely on matching existing texts, are ill-equipped to keep pace with the rapid evolution of AI algorithms (Weber-Wulff et al., 2023; Gustilo et al., 2024). This technological shortcoming has a direct impact on the efficacy and fairness of writing assessments. Educators must, therefore, evolve their understanding and innovate in assessment practices to ensure that the evaluation of original language and critical thinking is not overshadowed by assessments of AI literacy and digital competence (Huang, 2024; Wise et al., 2024).

Assessing English Writing

Assessing English writing is a multifaceted endeavor which necessitates both assessment criteria and scoring rubrics. To ensure its accuracy, consistency, and objectivity, an effective assessment criterion for English writing encompasses several key dimensions. The most highlighted among these is “content and ideas” which emphasize the clarity, relevance and depth of the argument as well as the provision of ample supporting details and evidence. “Organization and structure” are also crucial, assessing the logical sequencing of ideas and the overall coherence of the writing. Educators attach equal significance to “language use” in writing assessment, examining the accurate and appropriate application of vocabulary, grammar, and syntax. Lastly, “conventions” are designed to evaluate the writer’s ability to employ an appropriate tone and style, as well as to adhere to specific formatting requirements (Aull, 2015; Geijerstam et al., 2021; Mo & Troia, 2017; Neumann, 2014).

Rubrics, a more detailed and structured tool defined variably according to their intended purposes, play a key role in writing instruction. For students, a rubric acts as a navigational tool, guiding them toward enhanced academic performance (Andrade & Boulay, 2003). Educators utilize rubrics as instructional guidelines for student learning (Brookhart, 2013), as a benchmark for assessing writing (Arter & McTighe, 2001), and as a scaffold for providing feedback (Panadero & Jonsson, 2013). Regardless of whether they are used for evaluation, self-assessment, or feedback, rubrics remain a fundamental component of writing instruction.

However, the traditional components of a writing assessment rubric—criteria, descriptors, and scoring levels—may not be legitimized when AI is involved in writing. The components of rubric encompass several aspects, including idea and content, structure and coherence, language and academic style, in addition to the consideration of student attitudes and engagement (Allagui, 2024). With AI assistance capable of significantly improving the quality of writing assignments by refining language, optimizing structure, and enriching content (Kim et al., 2025), generative AI tools have transformed writing from a solely human endeavor into a digital-human collaboration (Wang, 2024). This transformation may render traditional rubric assessments obsolete if evaluators adhere strictly to evaluative paradigms that exclusively value human efforts. To enhance the effectiveness and accuracy of writing rubrics, educators must evolve their assessment practices, potentially integrating AI literacy into assessment criteria, monitoring the learning process, and placing greater emphasis on critical thinking (Wang & Wang, 2024).

In summary, the emergence of AI-assisted writing has profoundly reshaped the landscape of English writing instruction and assessment (Hao et al., 2024). While AI tools have enhanced writing fluency, organization, and lexical sophistication, they have also blurred the boundaries between human and machine authorship, challenging the traditional notions of originality and authenticity. This duality raises pressing questions about how academic integrity can be upheld and how writing quality should be evaluated when human and AI contributions are intertwined. For teachers, these developments necessitate a re-examination of assessment literacy, particularly their ability to interpret writing performance, design valid assessment rubrics, and make ethically informed judgments in AI-mediated contexts. The interplay between teachers’ assessment beliefs, their evaluative practices, and their ethical considerations regarding AI use constitutes an emerging and underexplored area in applied linguistics and language assessment research.

Given this evolving educational landscape, it becomes essential to explore how English as a Foreign Language (EFL) writing teachers perceive and respond to the integration of AI technologies in writing assessment. Specifically, this study aims to examine teachers’ attitudes toward academic integrity, their perceptions and practices in rubric-based assessment, and their expectations for adapting assessment approaches to ensure fairness, validity, and pedagogical alignment in the AI era.

Accordingly, the present study addresses the following research questions:

1. What are the perceptions and stances of EFL writing teachers toward academic integrity in the context of AI-assisted writing?
2. How do EFL teachers perceive and implement scoring rubrics when assessing AI-assisted writing tasks?
3. In what ways do EFL teachers believe writing assessment should evolve to remain valid, fair, and pedagogically meaningful in the AI era?

Method

Participants

Qualitative research, which typically involves smaller sample sizes than quantitative studies (Creswell & Clark, 2007), forms the basis of this study. The participants of the study are nine EFL teachers from a Chinese university renowned for its foreign language education programs. They were purposefully selected due to their intimate engagement with writing instruction and assessment. The group consists of seven native Chinese speakers and two native English

Table 1 Demographic information.

Participant	Age	Gender	Native language	Education level	Teaching years	Teaching areas
Teacher B	35	Male	English	PhD	10	Writing Academic writing
Teacher C	47	Male	Chinese	PhD	20	Writing Listening
Teacher H	57	Male	Chinese	Postgraduate	32	Writing, Interpretation, Literature
Teacher F	37	Male	Chinese	PhD candidate	12	Writing Comprehensive English
Teacher J	45	Female	English	Postgraduate	22	Writing Academic writing
Teacher L	47	Male	Chinese	PhD	19	Writing Comprehensive English Interpretation
Teacher Q	28	Female	Chinese	PhD candidate	2	Writing Listening
Teacher X	53	Female	Chinese	PhD	28	Writing Comprehensive English Learning strategy
Teacher Y	44	Female	Chinese	PhD	21	Writing Comprehensive English

speakers, all of whom hold a master’s degree or higher. Each participant teaches English writing to at least one class of approximately twenty students, with their teaching experience ranging from twenty to thirty-two years. The demographic information of participants is as followed.

Data Collection

Interviews were conducted via WeChat phone calls in November 2024. We employed semi-structured interviews with each participant, in accordance with Corbetta’s approach which allows for “a clearer idea about the questions to be asked, but without strict concern for the exact wording or order” (2003: 579). At the onset of each interview, the interviewer clarified the objectives and procedures to ensure that participants were fully informed and understood that their responses would be treated with strict confidentiality. Personalized schedules were arranged for each participant, who were asked to prepare for the online interview at their convenience.

The interview protocol was designed to elicit the teachers’ perceptions of the use of AI tools in student writing assignments, their views on assessing academic integrity in AI-assisted writing, and their perspectives on rubric-based evaluation of AI-assisted writing. Sample questions included, “*How do you perceive students’ use of AI in writing? How do you assess their AI-assisted writing? Would you please describe the criteria for assessing AI-assisted writing, if any?*”. The semi-structured format of the interviews provided rich qualitative data, while the flexible questioning allowed for a smooth expression of the teachers’ attitudes. Each interview lasted between fifteen minutes and half an hour, was audio-recorded, transcribed, and fully documented for subsequent analysis.

Data Analysis

The qualitative analysis of the nine interviews was conducted using the comparative method espoused by Strauss and Corbin (1997). The analytical process was structured into several sequential phases:

Transcription and proofreading: The audio recordings of the interviews were transcribed with “IFLYTECK meeting recorder”, a speech-to-text tool, and subsequently underwent rigorous proofreading to ensure accuracy and fidelity to the original dialogue.

Comparative analysis and categorization: The transcripts were analyzed on a question-by-question basis to synthesize a comprehensive view of the data. They were then organized into categories that captured the diverse experiences and attitudes of the teachers towards each specific aspect of the subject matter. This categorization was aligned with the key concerns and elements identified in the literature review. Throughout this process, a flexible approach was adopted to adjust predetermined categories as needed, in accordance with Marton’s (1986) advocacy for openness to discovering new shared elements.

Inductive inference and keyword quantification: An inductive approach was applied to the analysis of the interviews. The frequency of recurring keywords was quantified to establish subcategories within the prepared topics. This quantification was performed through Wordsmith, a corpus analysis tool, with the results undergoing manual review and verification to ensure reliability.

Re-examination and refinement of classifications: The established classifications were re-examined to identify and establish connections among the subcategories. This scrutiny facilitated further adjustment and refinement, ensuring that the analytical framework was both coherent and comprehensive.

This meticulous approach to qualitative analysis allowed for a nuanced understanding of the teachers’ perspectives on the use of AI in student writing assignments, the assessment of academic integrity in AI-assisted writing, and the rubric-based evaluation of such writing.

Findings

Finding 1: EFL Teachers’ Perspectives on Academic Integrity in AI-assisted Writing

While all teachers express a deep concern for academic integrity in writing, their comprehension of the concept appears to be rudimentary and has not kept pace with technological progress. The majority were unable to delineate specific scenarios in which AI might compromise

academic integrity. Upon being queried about the acceptability of using AI to generate ideas, construct outlines, enrich content, or refine language, most interviewees, with the exception of Teacher Y, exhibited hesitation.

From the perspective of the **conceptual dimension** of teacher assessment literacy (Pastore & Andrade, 2019), this confusion about the boundaries of academic integrity reflects a fundamental conceptual gap. Teachers’ understandings of authenticity and originality remain grounded in pre-AI assessment traditions, emphasizing human authorship and linguistic accuracy rather than critical awareness of algorithmic mediation. This conceptual deficiency constrains their ability to develop coherent ethical guidelines for AI-integrated assessment.

Teacher J admitted uncertainty, while Teacher B deemed seeking ideas and drafting outlines acceptable, akin to drawing inspiration from reading. However, they were opposed to using AI for enriching or polishing assignments, arguing that proofreading not conducted by students could mask their linguistic deficiencies. Other teachers rejected these practices altogether, emphasizing the importance of student originality.

Unlike traditional referencing through reading, AI tools may provide ready-made answers to assignments. Even without verbatim plagiarism, students might consciously or unconsciously adopt others’ ideas, whether from human authors or generative AIs. Such borrowing violates academic integrity standards, as AI’s ideas are derived from existing human work. AI consults, rewrites, and assembles these ideas but fails to provide proper attribution. It’s less ChatGPT than CheatGPT. This chain of errors should not continue. (Teacher C)

All interviewees acknowledged the necessity of adopting effective measures to detect AI-assisted breaches of academic integrity in student English writing assignments. Some experienced teachers (such as Teacher C and Teacher H) expressed confidence in their ability to identify writing that are not commensurate with students’ language competence or cognitive abilities, attributing this to their years of teaching experience.

Although coming from different parts of China, students have been learning English in some similar ways. In terms of writing, they have memorized, or were asked to memorize, many stylized expressions, idioms and sentence structures... Only a very small number of students have extensive reading. These confine their way of thinking and the depth of writing. If someone’s writing assignment dramatically deviates from students’ average performance and doesn’t align with his or her classroom performance, I think I might have caught a “mole”. (Teacher H)

In contrast, younger teachers (such as Teacher F and Teacher Q) lamented their inability to uncover AI-assisted misconduct, suggesting that technological advancement enable students to produce assignments that perfectly align with teachers’ perceptions and expectations.

Students don’t have to present the best assignment. As long as their assignment is better than what they can achieve by their own, they win hide-and-seek. AI prompts may go like: You’re an average student at [XXX] university. You are asked to write a composition on [XXX]. Please write an article according to the requirements. (Teacher F)

Finding 2: EFL Teachers’ Adherence to Traditional Scoring Rubrics in an AI Context

The majority of interviewees of the study asserted their ability to detect students’ use of AI tools in their writing assignments. They proposed comparing students’ performance on out-of-class assignments with in-class exams to identify significant discrepancies in their writing, such as language style and proficiency. However, this would only be effective at the end of the semester and would not precisely indicate the extent of AI involvement in the assignments. Teacher L suggested,

When the Internet access is cut off in an examination room, students could rely on nothing but their own writing skills. It’s very difficult for students to cheat in a writing exam, unless they know the topic in advance. Usually the topic has certain relations with their assignments. In the examination, they would habitually use those vocabulary and syntax they’re adept at. The stylistic differences, if exist, will never be ignored. (Teacher L)

With regard to AI detection tools, teachers’ perspectives diverged. All were aware of such software, with six holding a positive view that these tools could safeguard students from AI interference and ensure their constructive engagement. However, only one of these six teachers had personally experimented the use of a detection tool, while the other five attributed the lack of such tools in writing assessment to the absence of financial support from the school for premium memberships.

Teachers F and Y were skeptical about the reliability of AI detection, questioning the universal effectiveness of one detector against the multitude of evolving algorithms. Teacher B, placing more trust in his students than in software, tested popular AI detectors with his own pre-AI papers, finding the AI generation likelihood ranging from 30 to 70 percent, which led to his disillusionment and subsequent disuse of any AI detector or scoring platform. He emphasized the importance of fostering mutual trust between teachers and students, stating:

I go through their writings and still cannot recognize whether they fully relied on AI. It’s problematic, but I need to promote mutual trust between myself and the students. (Teacher B)

Interviewees suggested setting a criterion for AI generation rate below 20-30 percent, akin to the requirement of traditional plagiarism checks.

Only Teacher Y, keen on technological and methodological revolutions, set favorable requirements for AI application in writing assignments, while the others continued the writing assessment methods that they were inherently taught. These eight teachers discouraged students from using AI in any form, prioritizing language proficiency development over academic integrity concerns. They required original, human-authored assignments, with no AI drafting, enrichment, or polishing before submission. Teacher F explained that this approach ensured educational equity and the validity of teachers’ scoring, citing the example of Doubao and Volcengine, two AI models developed by ByteDance, to illustrate the potential damage to educational equity. The former is a free AI service available to all, while the latter offers more powerful functions through premium membership. In contrast, Teacher Y embraced AI in writing assignments, in collaboration with an educational technology specialist, which deepened her understanding and inclusiveness towards integrating AI into writing instruction. She observed consistent progress in students’ English competence despite the integration of AI.

If it has been the trend, if we have no way to stop it or get away from it, why not embrace it? On the other hand, we could urge student to compare their writings with the AI-polished texts. Definitely, students may learn something from such instant also customized feedback. (Teacher Y)

Teacher Y taught students to use AI prompts for inspiration seeking, structure optimization, and language polishing, encouraging them to practice digital skills and submit both original and AI-edited work for review. She explained that AI involvement in her syllabus was more research-targeted than teaching-oriented.

Teachers’ scoring attitudes and practices were not strictly congruous with their openness to AI application in writing. Regardless of their stance on AI, all interviewed teachers assessed homework based on traditional writing rubric components, such as idea, content, structure, coherence, language etc. It was understandable that Teacher C, who scrutinized AI with academic rigor, rated obviously AI-generated or polished writings as the lowest on the scale. However, even Teacher Y did not agree that digital literacy should be included in the rubric, suggesting that the focus should remain on content relevance, content efficiency, organization, and language quality, since they were teaching a writing course rather than a computer course.

AI automatic scoring was initially rejected by all teachers but was tentatively adopted by two for reference purposes, aiming to facilitate the identification of language errors. They manually reviewed each assignment and scored them according to long-established rubrics. Teacher X had a strongly negative view of AI scoring, finding the assessment results and ratings from different platforms confusing and contradictory. She mentioned,

Sometimes it’s just a game of prompts and algorithms. Large language models still cannot fully understand human language. They are reorganizing our words in a mathematical way. We have to score assignments in person, if fair evaluation is expected. (Teacher X)

On the other hand, all teachers acknowledged the benefits of AI scoring details, such as automatically highlighting misspellings, awkward syntax, and grammatical errors, which could greatly assist in targeted assessment. They suggested that while AI was not equal to scoring tasks, it could streamline the scoring process by providing more supporting evidence for focused assessment.

I have so many students’ writing assignments to review, revise and score. They seem to be a mission impossible, even for a native speaker of English. With AI, I no longer need to spend so much time correcting their language mistakes. Of course, the AI outcomes are just for reference. I have my judgement. (Teacher J)

Viewed through the **praxeological dimension** of teacher assessment literacy, this reliance on conventional rubrics demonstrates a lag between teachers’ assessment knowledge and their applied practice. Although many teachers acknowledge the inevitability of AI involvement, few have transformed this awareness into adaptive evaluation strategies. The persistence of static rubrics signals an underdeveloped capacity to enact and modify assessment procedures in response to evolving technological realities.

Finding 3: EFL Teachers’ Varied Perceptions on the Necessity to Update Scoring Criterion in Pace with the AI Age

The motivation of teachers to update their scoring rubrics for aligning with the demands of targeted practice and evaluation was found to be more hindered by their information technology literacy than by their level of teaching seniority. Despite having a shorter teaching history in writing courses, Teacher Q, who is indifferent to AI, showed little interest in adapting the existing writing syllabus and rubric to accommodate the new AI context. Adhering to the principle that fostering students’ language competence through active engagement is paramount, she saw no necessity in following the AI trend by revising teaching methods, pedagogical focus, and assessment rubrics. This stance was supported by the fact that students’ exam performance, which was devoid of technological assistance, significantly outperformed their assignment work over the past two years.

... AI may be applied to their writing assignments to some extent, but their language progress has been remarkable. Everything is on track. Why to change? (Teacher Q)

In contrast, Teachers Y and X, who were actively engaged with AI applications and had many years of teaching experience, explicitly expressed their interest in experimenting with new AI-assisted teaching models and consequently upgrading the assessment rubric to reflect these changes. Their preliminary ideas for reforming the criteria included:

Firstly, enhancing whole-process monitoring and evaluation. Although some interviewees were reluctant to adapt writing assessment to the AI era, all acknowledged the inevitability of AI and its extensive application in knowledge acquisition. Given the uncertainty of technological advancements regulating educational AI use through generative content detection and learning process supervision, Teacher X proposed shifting the focus of evaluation from outcome to process.

Since all students used AI in their writing assignments, evaluating their language performance, which can be easily enhanced with machine assistance, is less meaningful. Instead, we should assess their active participation and creative development of ideas. (Teacher X)

Teacher X also suggested that peer review practices could facilitate whole-process monitoring and evaluation, promoting mutual stimulation, monitoring, and assessment.

Second, reallocating grade weights. Most teachers, such as Teacher C, Teacher F, and Teacher H, noted that classroom performance in closed-book examinations was the most reliable indicator of students’ language proficiency in the absence of effective tools to detect AI-generated content. They maintained a teaching orientation that prioritizes language proficiency as a critical concern in their teaching and assessment. Increasing the weight of final exam grades could serve as a reminder to students that their own real language proficiency deserves attention and effort. Teacher Y proposed that the proportion of writing assignments in total grades could be lowered to 20 percent.

Third, delegating language assessment to advanced large language models (LLMs): Teacher F, Teacher H and Teacher X held the idealistic expectation that AI would take full responsibility for assessing and improving language performance, allowing non-native EFL teachers to focus more on instructing students to develop creative ideas and achieve critical draft based on AI

results. Teacher X suggested that students would be inspired by the diverse results presented by AI models, through which their language sensitivity and adeptness would mature. This would position teachers to guide learners in developing critical thinking for producing insightful writings by engaging with AI through competition, debate, and negotiation. Consequently, the central task of writing scoring would shift from evaluating students’ performance to assessing the extent to which human creativity surpasses machine generation.

All in all, from the socio-emotional dimension of teacher assessment literacy, teachers’ ambivalence toward updating scoring criteria reveals emotional and identity-based tensions in adapting to AI integration. Anxiety, distrust, and uncertainty—rather than pedagogical limitations—constitute the primary barriers to innovation. The absence of emotional readiness and professional support systems reflects the neglect of socio-emotional literacy, which is vital for sustaining teachers’ confidence and reflective engagement in AI-mediated assessment.

Discussion

Educators’ reluctance to endorse AI tools in writing assignments reflects deep concerns about the potential erosion of students’ intrinsic motivation. Their responses indicate a fear that overreliance on AI could foster a passive learning environment, in which critical thinking and language proficiency are subordinated to post-AI editing tasks. At the same time, teachers acknowledge the inevitability of AI integration, as evidenced by reports of its widespread use in student assignments. To balance technological involvement with active engagement, teachers may collaborate with AI to create enjoyable learning experiences, design interactive content, and provide individualized support (Huang et al., 2023; Wei, 2023). This aspect warrants particular attention, as research suggests that students value prompt, targeted AI feedback (Huang & Teng, 2025) while instructors perceive assignment review and revision as a substantial burden (Yu, 2021).

The tension between recognition and resistance also highlights a conceptual gap in teachers’ assessment understanding. Conceptions of assessment remain anchored in traditional notions of “authentic effort” and “manual authorship,” neglecting the epistemic shift toward human-AI collaboration. Such limitations restrict educators from reframing assessment integrity in broader terms, including transparency, critical mediation, and ethical authorship (Popham, 2009; 2018).

Adherence to conventional scoring rubrics underscores the continued centrality of teacher effort in curriculum evolution, particularly amid educational technology integration (Hao et al., 2024). The observation that five of six teachers supported AI detection tools despite never having used them raises questions about their commitment to pedagogical enhancement. Workload considerations were frequently cited as a rationale for limited engagement with AI-based assessment innovations, suggesting a preference for technologies that reduce labor rather than enhance instructional quality. Teacher Q exemplified this pattern: although reporting a lack of expertise in educational technology, they could clarify misconceptions about automated scoring mechanisms, revealing latent technical acumen.

This gap between awareness and practice indicates underdeveloped operational assessment literacy. Teachers recognize AI’s presence but fail to integrate it meaningfully into feedback and assessment design, reflecting inertia in translating theoretical understanding into adaptive practice (DeLuca et al., 2016). Addressing this gap requires professional development that fosters procedural reflection, contextual sensitivity, and practical enactment.

Ethical considerations are central to this endeavor. Conflicting signals between technological inadequacy and demonstrated technical understanding underscore the need to prioritize eth-

ical education before substantive pedagogical innovation. Ethical reflection is fundamental to professional conduct, sharpening awareness of moral dilemmas inherent in educational practice and guiding educators toward higher-quality teaching (Gülcan, 2015; Boon, 2011). Emotional hesitation further compounds this issue, as teachers’ anxiety about AI stems from a lack of trust, confidence, and alignment with AI as a pedagogical partner. Strengthening socio-emotional competence is thus critical for fostering empathy, resilience, and ethical self-regulation in technology-mediated assessment.

Digital literacy emerges as a crucial determinant of teachers’ willingness to adapt assessment practices in the AI era. Teachers with limited AI knowledge or experience are more likely to adhere to traditional language-centered instruction, prioritizing vocabulary, grammar, and lexicogrammatical features (Coxhead & Byrd, 2007). Literature majors or instructors with minimal AI engagement, such as Teachers C, H, and L, exemplify this trend. Overcoming such resistance requires targeted professional development in digital literacy, emphasizing recognition of technological competence and effective training strategies (Huang, 2024).

Collectively, conceptual confusion, practical inertia, and emotional hesitation constitute a structural literacy gap in the AI era. Closing this gap demands integrated professional development that cultivates conceptual clarity, procedural adaptability, and emotional resilience, thereby transforming assessment literacy into a holistic competence capable of aligning human judgment with intelligent technology (Xu & Brown, 2016).

Conclusion

This study examines the attitudes of nine EFL teachers towards the assessment of AI-assisted writing assignments from the perspective of teacher assessment literacy. The findings indicate that the participants possess a realistic understanding of the technological shift and its ramifications for educational practices. Their reactions, both active and passive, are influenced by their personal information technology literacy and professional dedication. The majority of participants maintained a conservative stance, adhering to conventional scoring methods, primarily due to a pedagogical emphasis on language instruction and personal preferences for convenience. This suggests that EFL teachers’ innovation in writing assessment has not kept pace with contemporary demands. By meticulously analyzing the interview transcripts, this research offers policymakers, school administrators, and educators a detailed portrait of teachers’ attitudes toward scoring, current assessment practices, and potential adaptations to AI-assisted evaluation.

The findings of this qualitative interview study provide valuable implications for teacher education and professional development. The three-dimensional conceptualization of assessment literacy highlights the necessity of cultivating teachers’ conceptual understanding, practical competence, and socio-emotional awareness in an integrated manner. Teacher education programs are therefore encouraged to design context-sensitive training modules that not only strengthen teachers’ theoretical knowledge and assessment techniques but also foster reflective awareness of the ethical and emotional dimensions of assessment practice. Moreover, the framework offers a reference for policymakers and school leaders to support teachers’ ongoing professional learning through collaborative assessment cultures and identity-oriented development pathways.

Nevertheless, several limitations should be acknowledged. As a qualitative study, the findings are contextually bound and derived from a limited sample, which may restrict their generalizability. The participants’ perspectives were shaped by their specific institutional and cultural settings, and thus may not fully represent the diversity of teachers’ assessment experiences in

other contexts. Future research could employ mixed-methods or longitudinal designs to triangulate findings and further explore how teachers’ assessment literacy evolves across different educational stages and sociocultural environments.

Funding

The work was supported by the SISU Education Reform Grant [41000662/054] and Shanghai Key Courses Program.

Author Bio

Jianfeng Zhang is an English language instructor at School of Education, Shanghai International Studies University. He obtained his Ph.D. degree in English Literature from the same institution. His teaching expertise encompasses courses in English language and narratology, and his research interests are primarily concentrated on narratological studies within educational contexts.

References

- Alharbi, W. (2023). AI in the foreign language classroom: A pedagogical overview of automated writing assistance tools. *Education Research International*, (1), 4253331. <https://doi.org/10.1155/2023/4253331>
- Ali, O., Murray, P. A., Momin, M., Dwivedi, Y. K., & Malik, T. (2024). The effects of artificial intelligence applications in educational settings: Challenges and strategies. *Technological Forecasting and Social Change*, 199, 123076. <https://doi.org/10.1016/j.techfore.2023.123076>
- Allagui, B. (2024). A scaffolding intervention to improve self-efficacy in source-based argumentative writing. *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2024.1454104>
- Andrade, H. G., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *The Journal of Educational Research*, 97(1), 21–30. <https://doi.org/10.1080/00220670309596625>
- Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom*. Sage Publications.
- Aull, L. (2015). Connecting writing and language in assessment: Examining style, tone, and argument in the U.S. Common Core standards and in exemplary student writing. *Assessing Writing*, 24, 59–73. <https://doi.org/10.1016/j.asw.2015.03.002>
- Boon, H. (2011). Raising the bar: Ethics education for quality teachers. *Australian Journal of Teacher Education*, 36(7), 104–121. <https://doi.org/10.14221/ajte.2011v36n7.2>
- Boulanger, D., & Kumar, V. S. (2024). Explainable AI and awe: Balancing tensions between transparency and predictive accuracy. In Mark D. Shermis & Joshua Wilson (Eds.), *The Routledge International Handbook of Automated Essay Evaluation* (pp. 445–468). Taylor and Francis. <https://doi.org/10.4324/9781003397618-26>
- Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Association for Supervision and Curriculum Development.
- Chhabra, P., & Sameer Babu, M. (2024). An explorative study on the use of artificial intelligence in Oman high school education. In Paypal Bansal & Rajeev Kumar (Eds.), *Artificial Intelligence and Communication Techniques in Industry 5.0* (pp. 55–72). CRC Press. <https://doi.org/10.1201/9781003494027-4>
- Corbetta, P. (2003). *Social research theory, methods and techniques*. SAGE Publications.
- Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16, 129–147. <https://doi.org/10.1016/j.jslw.2007.07.002>
- Creswell, J. W., & Clark, V. L. P. (2007). *Designing and conducting mixed methods Research*. Sage Publications, Inc.
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251–272. <https://doi.org/10.1007/s11092-015-9233-6>

- East, M., & Slomp, D. (2024). The ethical turn in writing assessment: How far have we come, and where do we still need to go? *Language Teaching*, 57(2), 262–273. <https://doi.org/10.1017/S0261444823000034>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1), 57. <https://doi.org/10.1186/s41239-023-00425-2>
- Fariani, R. I., Junus, K., & Santoso, H. B. (2023). A systematic literature review on personalised learning in the higher education context. *Technology, Knowledge and Learning*, 28(2), 449–476. <https://doi.org/10.1007/s10758-022-09628-4>
- Flowerdew, J., & Peacock, M. (2001). The EAP curriculum: Issues, methods, and challenges. In J. Flowerdew & M. Peacock (Eds.), *The EAP curriculum: Issues, methods, and challenges* (pp. 177–194). Palgrave Macmillan.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language assessment quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- Gee, J. 2004. *Situated language and learning: A critique of traditional schooling*. Routledge.
- Geijerstam, Å. A., Folkeryd, J. W., & Liberg, C. (2021). Linguistically based scales for assessment of young students’ writing. *Writing and Pedagogy*, 13(1–3), 227–265. <https://doi.org/10.1558/wap.21504>
- Gülcan, N. Y. (2015). Discussing the importance of teaching ethics in education. International Conference on New Horizons in Education, INTE 2014, 25–27 June 2014, Paris, France, 174, 2622–2625. <https://doi.org/10.1016/j.sbspro.2015.01.942>
- Gustilo, L., Ong, E., & Lapinid, M. R. (2024). Algorithmically-driven writing and academic integrity: Exploring educators’ practices, perceptions, and policies in AI era. *International Journal for Educational Integrity*, 20(1), 3. <https://doi.org/10.1007/s40979-024-00153-8>
- Hao, Z., Fang, F., & Peng, J.-E. (2024). The integration of AI technology and critical thinking in English major education in China: Opportunities, challenges, and future prospects. *Digital Applied Linguistics*, 1, 2256. <https://doi.org/10.29140/dal.v1.2256>
- Huang, A. Y. Q., Lu, O. H. T., & Yang, S. J. H. (2023). Effects of artificial Intelligence–Enabled personalized recommendations on learners’ learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education*, 194, 104684. <https://doi.org/10.1016/j.compedu.2022.104684>
- Huang, F. (2024). Examining foreign language teachers’ information literacy: Do digital nativity, technology training, and fatigue matter? *Asia-Pacific Education Researcher*, 33, 901–912. <https://doi.org/10.1007/s40299-023-00797-z>
- Huang, J., & Teng, M. F. (2025). Peer feedback and ChatGPT-generated feedback on Japanese EFL students’ engagement in a foreign language writing context. *Digital Applied Linguistics*, 2, 102469. <https://doi.org/10.29140/dal.v2.102469>
- Kahveci, H. (2023). The positive and negative effects of teacher attitudes and behaviors on student progress. *Journal of Pedagogical Research*, 7(1), 290–306. <https://doi.org/10.33902/JPR.202319128>
- Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students’ perspectives on generative AI-assisted academic writing. *Education and Information Technologies*, 30(1), 1265–1300. <https://doi.org/10.1007/s10639-024-12878-7>
- Liang, J., Huang, F., & Teo, T. (2024). Understanding Chinese university EFL learners’ perceptions of AI in English writing. *IJCALLT*, 14(1), 1–16.
- Lillis, T., & Curry, M. J. (2024). Multilingualism in academic writing for publication: Putting English in its place. *Language Teaching*, 57(4), 501–520.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *Information Systems Research*. arXiv e-prints. <https://doi.org/10.48550/arXiv.1907.11692>
- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442–467. <http://dx.doi.org/10.1080/0969594X.2016.1268090>
- Macfarlane, B., Zhang, J., & Pun, A. (2014). Academic integrity: A review of the literature. *Studies in Higher Education*, 39(2), 339–358. <https://doi.org/10.1080/03075079.2012.709495>
- Marton, F. 1986. “Phenomenography: A research approach to investigate different understandings of reality.” *Journal of Thought*, 21(3), 28–49. <http://www.jstor.org/stable/42589189>

- Mirón-Mérida, V., & García-García, R. (2024). Developing written communication skills in engineers in Spanish: Is ChatGPT a tool or a hindrance? *Frontiers in Education*, 9. <https://doi.org/10.3389/educ.2024.1416152>
- Mo, Y., & Troia, G. A. (2017). Similarities and differences in constructs represented by U.S. States’ middle school writing tests and the 2007 national assessment of educational progress writing assessment. *Assessing Writing*, 33, 48–67. <https://doi.org/10.1016/j.asw.2017.06.001>
- Neumann, H. (2014). Teacher assessment of grammatical ability in second language academic writing: A case study. *Journal of Second Language Writing*, 24, 83–107. <https://doi.org/10.1016/j.jslw.2014.04.002>
- Niloy, A. C., Akter, S., Sultana, N., Sultana, J., & Rahman, S. I. U. (2024). Is ChatGPT a menace for creative writing ability? An experiment. *Journal of Computer Assisted Learning*, 40(2), 919–930. <https://doi.org/10.1111/jcal.12929>
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Papakonstantinidis, S., Kwiatek, P., & Spathopoulou, F. (2024). Embrace or resist? Drivers of artificial intelligence writing software adoption in academic and non-academic contexts. *Contemporary Educational Technology*, 16(2). <https://doi.org/10.30935/cedtech/14250>
- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84, 128–138. <https://doi.org/10.1016/j.tate.2019.05.003>
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory and Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>
- Popham, W. J. (2018). *Assessment Literacy for Educators in a Hurry*. ASCD.
- Rasul, T., Nair, S., Kalendra, D., Balaji, M. S., Santini, F. de O., Ladeira, W. J., Rather, R. A., Yasin, N., Rodriguez, R. V., Kokkalis, P., Murad, M. W., & Hossain, M. U. (2024). Enhancing academic integrity among students in GenAI Era: A holistic framework. *The International Journal of Management Education*, 22(3), 101041. <https://doi.org/10.1016/j.ijme.2024.101041>
- Salmani Nodoushan, M. A. (2014). Assessing writing: A review of the main trends. *Studies in English Language and Education*, 1(2), 119–129. <https://doi.org/10.24815/siele.v1i2.1831>
- Shabani, E. A., & Panahi, J. (2020). Examining consistency among different rubrics for assessing writing. *Language Testing in Asia*, 10(1), 12. <https://doi.org/10.1186/s40468-020-00111-4>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Stiggins, R. J. (1991). Assessment literacy. *Phi delta kappan*, 72(7), 534–539.
- Strauss, A., and J. M. Corbin. 1997. *Grounded theory in practice*. Sage Publications.
- Teng, M. F. (2024). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies*, 6(3), 36–57. <https://doi.org/10.58304/ijts.20240304>
- Teng, M. F. (2025). Metacognitive awareness and EFL learners’ perceptions and experiences in utilising ChatGPT for writing feedback. *European Journal of Education*, 60(1), e12811. <https://doi.org/10.1111/ejed.12811>
- Wang, C. (2024). Exploring students’ generative AI-assisted writing processes: Perceptions and experiences from native and nonnative English speakers. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-024-09744-3>
- Wang, C., & Wang, Z. (2025). Investigating L2 writers’ critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing*, 67. <https://doi.org/10.1016/j.jslw.2025.101187>
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltyniek, T., Guerrero-Dib, J., Popoola, O., Sigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19(1). <https://doi.org/10.1007/s40979-023-00146-z>
- Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, 14. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1261955>
- Wise, B., Emerson, L., Van Luyn, A., Dyson, B., Bjork, C., & Thomas, S. E. (2024). A scholarly dialogue: Writing scholarship, authorship, academic integrity and the challenges of AI. *Higher Education Research & Development*, 43(3), 578–590. <https://doi.org/10.1080/07294360.2023.2280195>

- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58(1), 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Yan, L., Greiff, S., Teuber, Z., & Gasevic, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, 8(10), 1839–1850. <https://doi.org/10.1038/s41562-024-02004-5>
- Yeo, M. A. (2023). Academic integrity in the age of Artificial Intelligence (AI) authoring apps. *Tesol Journal*, 14(3), e716. <https://doi.org/10.1002/tesj.716>
- Yu, S. (2021). Feedback-giving practice for L2 writing teachers: Friend or foe? *Journal of Second Language Writing*, 52. <https://doi.org/10.1016/j.jslw.2021.100798>