The Application of Artificial Intelligence Automated Evaluation Systems in English Writing Instruction: Scoring Accuracy, Feedback Effectiveness, and User Experience

*Tong Zou**

School of English Language and Culture, Guangdong University of Foreign Studies South China Business College, Guangzhou, Guangdong, 510545, China

Abstract. With the deepening application of artificial intelligence (AI) in education, automated evaluation systems have become important aids in English writing instruction. However, disputes persist regarding their scoring accuracy, feedback effectiveness, and user experience. This paper reviews the application of AI automated evaluation systems in English writing instruction, focusing on scoring accuracy, feedback effectiveness, and user experience. Findings show that AI systems perform well in evaluating linguistic forms, demonstrating high consistency with human raters, especially in grammar and vocabulary. Nevertheless, they remain inadequate in assessing content depth, logical structure, and creativity. Regarding feedback, AI systems can enhance students' writing proficiency if combined with scaffolded designs (e.g., model essays, checklists, multiple rounds of revision). In terms of user experience, students value the immediacy and interactivity of AI feedback but trust it less than teacher feedback on content-related aspects. The conclusion points out that AI systems should be positioned as playing an auxiliary role in formative evaluation, leverage their advantages in efficiency and personalization through human-machine collaboration, and pay attention to teacher training, system optimization, and ethical norms, so as to promote English writing instruction toward a more efficient, inclusive, and human-centered direction.

1 Introduction

As information technology grows quickly and schools push digital change in education, people around the world now agree on the need to modernize education with technology. Education quality has always been at the center because it supports both social progress and personal growth. In this case, new tools such as artificial intelligence (AI) are changing the education system deeply. They also show strong promise in teaching writing, speaking, and reading.

In today's world, English is an important tool for communication and study, yet its teaching still needs new models and methods. Writing instruction, which is a central part of

* Corresponding author: 22401651220@e.gwng.edu.cn

developing English skills, has long faced problems such as heavy grading for teachers, slow feedback, and limited personal guidance. To address these problems, AI-based Automated Essay Scoring (AES) systems, like Pigai and ChatGPT, have been introduced. These systems are simple to use and easy to access. They depend on natural language processing and deep learning to give quick scores and many kinds of feedback on students' writing. They not only reduce the heavy grading work of teachers but also give learners personal and step-by-step help in writing, which improves the efficiency of both teaching and learning [1, 2].

However, even though AI systems show many clear benefits, there are still big gaps and much debate about scoring accuracy, feedback depth, user experience, and other areas, which need careful review and deeper study. Using recent empirical research from both inside and outside the country, this paper gives a full review of how AI automated evaluation systems are used in English writing teaching. It looks at three main parts: scoring accuracy, feedback usefulness, and user experience, together with related talks, and aims to give theory support and method guidance for improving teaching practice and pushing system innovation.

2 Key aspects of AI AE system application

2.1 Scoring accuracy

Scoring accuracy is mainly judged by how close system scores are to human scores (e.g., exact match rate, correlation numbers), by reliability (whether results stay steady across tasks and groups), and by validity (whether scores match the idea of writing) [3]. Recent studies have further examined bias types, such as preference for longer texts (length bias), overfitting to specific genres, and group differences caused by uneven training corpora. Han Fangting et al. found that Pigai' s scores were generally higher (mean = 85.07) with smaller variance, while teachers' scores were lower (mean = 76.95) with larger variance [3]. ChatGPT's scores fell in between (mean = 82.07) and showed significant correlations with both (r = 0.611, 0.569). Other research has supported this conclusion. Escalante et al., through rigorously controlled experiments, discovered that GPT-4-generated feedback produced no significant difference from human feedback in improving students' writing performance, suggesting that AI has approached human-level proficiency in promoting language development [4].

Nevertheless, AI scoring still has notable limitations. While studies show high consistency with human raters under clear rubrics, its performance in assessing content depth (e.g., originality, evidence quality) and argumentative logic remains unstable [3]. Teachers also worry that AI fails to capture creativity, critical thinking, and authorial voice, reflecting its weakness in deep semantic understanding [5]. AI often makes writing look the same and sometimes judges non-native writers' language in the wrong way. Mun found that ChatGPT cuts grammar mistakes but does not make writing more unique or richer in words. This shows a risk of sameness [6].

Evidence shows that AI Automated Evaluation (AE) works better for formative use than for high-stakes tests [1]. In high-stakes cases, teachers say AI should only give support and not make the final call [5]. One way is "dual-channel scoring" with AI doing the first check and humans reviewing it. Another way is "multiple evidence" with final drafts, process drafts, and oral talks. Wu gave a "dynamic assessment model" that uses process data and shows a real case of multidimensional checks [7]. Systems with alerts for doubt and error checks can also make sure humans step in on time.

In conclusion, in formative writing assessment, AI scoring should act as an early check tool. Systems should not only give one score. They should also record score changes from the first draft to the last draft. This can push students to revise many times with AI help. The key value lies in generating a "progress curve" that reflects whether students adopt feedback,

correct errors, and achieve improvement. At the same time, AI can display high-confidence results in areas such as grammar and vocabulary, while suggesting "teachers should check" for weaker aspects like content innovation and logical structure. This transparent strategy can not only build user trust, but also improve the practical reference value of the scoring results. These innovative applications can improve the accuracy of AI scoring in classroom practice and better integrate it into formative assessment systems to support learning. Future development should not aim to replace teacher scoring with AI entirely but rather to establish a human-AI collaborative evaluation paradigm, leveraging complementary strengths to enhance both the accuracy and the educational value of assessment.

2.2 Feedback effectiveness

Successful AI feedback is typically specific, hierarchical, and trackable. Firstly, it is specific and actionable, pointing to executable revision moves (e.g., "merge two paragraphs to strengthen the causal chain") rather than offering vague judgments [8]. Secondly, it has a clear hierarchy, covering multiple levels from language form (vocabulary/grammar) to discourse structure (cohesion/logic) and content ideas (arguments/evidence), while aligning with course rubrics [9]. Thirdly, it is trackable. By tracking and showing the trajectory of progress, it can help students build self-efficacy. A case study by Zhang Yue demonstrated this point: one student, after making 150 revisions using Pigai, improved his score from 30 to 73.5, highlighting the importance of continuous feedback [8].

The effectiveness of AI feedback in enhancing learning outcomes often depends on the implementation of specific scaffolded designs and revision processes. When teachers provide clear writing requirements, exemplary essays, and concrete revision checklists (draft–feedback–revision–re-evaluation), students can better utilize AI feedback to improve their writing quality [9]. For example, when writing an essay on "How Artificial Intelligence Changes Learning," teachers can first provide a model essay demonstrating clear arguments, sufficient evidence, and coherent logic, along with a checklist specifying the standards to be met. The AI system can then offer each student targeted suggestions based on the checklist: if the argument is unclear, AI might prompt, "Clearly state the three main impacts of AI on learning in the first paragraph"; if examples are insufficient, it might suggest, "Add concrete cases, such as online learning platforms or intelligent teaching systems"; if paragraph transitions are awkward, it could recommend, "Try using conjunction such as 'therefore' or 'however' to improve logical flow." Through such specific and actionable guidance, students can accurately understand what needs to be revised and how to improve, thereby genuinely enhancing their writing proficiency.

When providing feedback, AI systems may sometimes offer suggestions that appear correct on the surface but actually alter the original meaning, or cause all students' s writing styles to become similar [3]. To avoid these issues, instructors can provide AI with specific prompts and requirements, guiding it to follow certain strategies. For example, when revising a reading response, if AI suggests, "Add more emotional description here," a student can ask, "Why should I add this?" The AI would then explain, "This helps the reader better experience your feelings when reading this part of the story," and provide an example, such as: "You could write, 'When I saw the protagonist overcome the challenge, I couldn't help feeling happy for him.'" Additionally, AI can focus on one key issue at a time. For instance, today it might guide students on how to make sentences more fluent, and tomorrow focus on improving paragraph cohesion. This step-by-step approach is akin to learning to walk—progress gradually without rushing. Moreover, students can be encouraged to record their reflections after revisions, like "I followed the suggestion to rearrange sentence order because it reads more smoothly, but I kept the original opening because it better expresses my genuine feelings."

The author believes that the most important goal is to cultivate students' independent thinking. Similar to using navigation software, students can refer to AI suggestions, but must ultimately decide the direction and style of their writing themselves. Teachers can guide students to reflect on questions such as: Does this suggestion align with the meaning I want to convey? Will this revision compromise my personal style? In this way, students can learn writing skills while maintaining their unique voice.

2.3 User experience

Students generally show high acceptance of AI writing systems, particularly valuing their instant feedback and interactive features. Shen Zhihua reported that 92% of students found Pigai helpful for writing, and 90% were most satisfied with its real-time feedback functionality. Some studies further support these findings [2]. For instance, Rahman et al. found that students using Grammarly experienced significant improvements in both writing confidence and interest, and they particularly appreciated the detailed explanations and immediate revision suggestions provided by the AI [10].

The system's "score progression" mechanism has been shown to effectively enhance writing motivation [8]. Students receiving feedback from ChatGPT not only improved their writing accuracy but also experienced a noticeable reduction in anxiety, further demonstrating the positive affective impact of AI systems [6]. However, students' trust in AI feedback remains lower than that in teacher feedback, particularly regarding content and logical structure [3]. Escalante et al. revealed the underlying reasons for this trust gap: students who prefer human feedback value the sense of engagement in face-to-face interactions and the opportunity for immediate clarification [4].

For teachers, AI systems shift the focus of their work from "grading burden" to "instructional design and management." Romadhoan's interview study found that teachers generally believe AI cannot replace human evaluation of creativity and critical thinking in writing, but they acknowledge its value in improving scoring efficiency and reducing workload [5]. Teachers are still responsible for developing rubrics, designing workflows, and managing academic integrity [1].

AI systems still face several challenges in practical application: students often exhibit insufficient trust in AI feedback, particularly regarding content and logical structure; teachers need to adapt to a new role, shifting from graders to learning process designer; and there are ongoing concerns regarding academic integrity, as well as the accuracy and applicability of feedback. The author suggests several approaches to address these issues.

Firstly, AI can serve as an "intelligent learning companion," not only identifying errors but also providing encouraging feedback, such as, "This example is great! Adding one more piece of data would make it even better," which can increase students' willingness to accept suggestions. Secondly, three feedback modes can be offered: a concise mode (highlighting only key errors), a detailed mode (providing revision suggestions), and a learning mode (explaining the reasons for errors and providing example sentences). Students can select the mode according to their proficiency, which not only reduces time costs by focusing effort on important aspects but also helps them clearly understand their weaknesses. Furthermore, relevant authorities like governments and schools, can develop and update various AI assistants for teachers, such as developing more assistants that can automatically generate class writing reports and quickly display common questions of the whole class, so as to facilitate teachers to give more efficient and targeted guidance to students.

2.4 Other aspects

In terms of instructional models, AI Automated Evaluation (AI AE) systems can be deeply integrated with various teaching approaches, such as flipped classrooms, writing workshops, and project-based learning. A successful example is Professor Zhang Wenxia's "Five-Step Writing Method" at Tsinghua University. This model, which involves pre-writing preparation, AI evaluation, peer review, teacher feedback, and self-editing, has significantly improved teaching outcomes, increasing students' average essay scores by 20% [9]. The author believes that an ideal AI-supported writing instruction model should follow the principle of "AI handles standardization, teachers guide creativity." In detail, AI systems give feedback on basic parts like language accuracy and structure rules, while teachers guide students in critical thinking, creativity, and personal writing style. This way of dividing work not only makes teaching more efficient but also helps stop the sameness that AI feedback may bring. For instance, in argumentative writing, AI can check if arguments are clear and if the evidence is enough, while teachers can help students think about how original their arguments are and how deep their supporting evidence is.

When it comes to ethics and academic honesty, it is important to improve source citation and protect data privacy, and also to use process-based proof (e.g., draft changes, oral reports) to stop ghostwriting [1]. Han Fangting et al. showed that 10% of students admitted they used AI directly to write essays, so teachers need to stress academic honesty more [3]. Romadhoan pointed out that AI writing needs both technical checks and teaching guidance [5]. The author thinks that besides using technical tools to spot AI writing, it is even more important to build students' sense and skill of using AI in the right way. Schools can hold lectures and courses on AI writing, teaching students how to use AI tools well without losing their own thinking. At the same time, testing systems should change to pay less attention to standard answers and more to thought process and originality. For example, writing scores can give more weight to "depth of thinking" and "personal insight," so students will learn to use AI wisely instead of depending fully on AI text.

On the topic of equity and accessibility, researchers stress the need to fix gaps in devices and language resources, and to support educational fairness with tools like offline materials and bilingual interfaces [11]. The use of Pigai in Chaoyang District, Beijing, shows that AI systems can give equal learning chances to schools with few resources and help promote balance in education [8]. The author advises using a "differentiated support" plan. For areas with fewer resources, besides giving basic tech devices, it is also key to build AI writing tools that fit the cultural backgrounds and learning styles of local students. Such tools should notice and value different cultural ways of expression instead of using only one rule. In addition, it is suggested to set up regional AI writing resource-sharing platforms so that teachers in different schools can share ideas and materials for AI-aided writing teaching.

Looking to future development, research indicates that in the medium to long term, the effective use of intelligent teaching systems can sustainably reduce teachers' workload. The author thinks AI writing evaluation systems will move toward being "growth companions." In the future, systems will not only check writing results. They will also track students' writing steps and give clear advice. For example, if a system finds a student often struggles with logic, it can give useful study materials and practice right away. This data-based and personal help can guide each student to reach their full writing skill and get real personal teaching.

The author points out that teacher training and support are very important. Many teachers still feel worried and face problems when using AI tools, so they need steady training and technical help. Schools can set up platforms for practice and sharing to drive the deeper use of AI in writing classes. Education offices should also make clear rules and standards to ensure that AI use in schools really improves fairness and quality.

3 Conclusion

The AI Automated Evaluation (AI AE) System has gradually become part of English writing classes. It acts as an important tool to improve teaching efficiency and provide personal learning support. The system works well in areas like automatic scoring, fast feedback, and motivating writing. The system still has clear limits in understanding deep meaning, checking logical order, and assessing creative ideas. It may also cause problems such as biased scores and making language too uniform. Future work in this area should focus on several key directions. Teacher training should be improved to help teachers use AI tools and plan lessons. This can change their role from just giving grades to guiding the learning process. Technical improvements should be made by using more advanced natural language processing models. This can help the system better recognize text structure, follow logic, and support creative thinking. Applied research should also grow, looking at how AI feedback affects writing over time for students at different levels and from different backgrounds. This can help create more varied ways to use AI in teaching. In the end, AI systems should work as "intelligent partners" with teachers. They can form a new human-AI way of teaching. This approach uses AI's fast data processing and instant feedback. Teachers still focus on helping students think critically, be creative, and get emotional support. Together, they can make English writing teaching more efficient, personal, and human-focused. Using these systems well needs careful planning and a focus on people. Assessment methods like "dual human-AI evaluation with multidimensional evidence" can help students reach their full writing ability and improve overall language skills.

References

1. M. Liu, C. Luo, G. Yuan, Innovative approaches to reducing teachers' workload in the era of smart education: A study on automated test generation and personalized writing. J. Teach. Educ. **8**, 48–55 (2021)
2. Z. Shen, An application study of Pigai in college English writing instruction. J. Hebei Radio TV Univ. **24**, 65–69 (2019)
3. F. Han, J. Mao, S. Ling, A comparative study on the effectiveness of AIGC-based automated evaluation in college English writing. Chin. For. Educ. **43**, 12–18 (2025)
4. J. Escalante, A. Pack, A. Barrett, AI-generated feedback on writing: Insights into efficacy and ENL student preference. Int. J. Educ. Technol. High. Educ. 20, 57 (2023)
5. M.G.E. Romadhoan, Has AI cracked the code on English proficiency assessments? J. Qual. Res. Lang. Educ. **15**, 112–130 (2024)
6. C. Mun, EFL learners' English writing feedback and their perception of using ChatGPT. J. Engl. Teach. Movies Media **25**, 78–95 (2024)
7. X. Wu, Dynamic evaluation of college English writing ability based on AI technology. J. Intell. Syst. **31**, 889–901 (2022)
8. Y. Zhang, Intelligent correction of English essays. Artif. Intell. **5**, 86–94 (2019)
9. S. Xue, A study on college English writing correction models based on large language models. Front. Mod. Educ. **6**, 19–21 (2025)
10. N.A.A. Rahman, et al., Assessing writing abilities using AI-powered writing evaluations. J. Asian Behav. Stud. **8**, 33–47 (2023)
11. P. Yang, Research on constructing a multimodal college English writing competence evaluation system with big data and AI technology. Mod. Educ. Technol. **33**, 75–80 (2023)