Journal of Contemporary Educational Research, 2025, Volume 9, Issue 8
http://ojs.bbwpublisher.com/index.php/JCER
ISSN Online: 2208-8474
ISSN Print: 2208-8466

Innovative Research on AI-Enabled Innovation of Multiple Evaluation Systems in College English Teaching

Shuo Chen*, Jing Yang

Changchun University of Finance and Economics, Changchun 130122, Jilin, China

**Author to whom correspondence should be addressed.*

Abstract: With the deep integration of artificial intelligence (AI) and education, the reform of college English teaching evaluation has entered a new stage. Traditional college English evaluation systems are faced with problems such as over-reliance on summative assessment, a single evaluation dimension, and lagging feedback. This study explores the innovative path of multiple evaluation systems in college English teaching empowered by AI, aiming to construct a dynamic, comprehensive, and personalized evaluation model. Through literature review, case analysis, and empirical research, it is found that AI technologies such as natural language processing, machine learning, and big data analytics can effectively support the diversification of evaluation subjects (teachers, students, peers, AI systems), the enrichment of evaluation dimensions (knowledge mastery, language competence, learning processes, and innovative thinking), and the intelligence of evaluation feedback. However, challenges such as algorithmic bias, data privacy risks, and the weakening of teacher-student interaction still exist. Finally, this study proposes countermeasures, including optimizing AI algorithms, strengthening data security, and balancing AI and human roles, to provide theoretical and practical references for the reform of college English teaching evaluation.

Keywords: Artificial intelligence; College English teaching; Multiple evaluation system; Educational technology

Online publication: 4th September 2025

1. Introduction

In the context of global integration and educational digitalization, college English teaching has shifted its focus from knowledge input to the cultivation of comprehensive language competence, which puts forward higher requirements for the evaluation system [1]. The evaluation system, as a "baton" guiding teaching reform, not only reflects students' learning outcomes but also affects the direction of teachers' teaching design and students' learning strategies [2].

Traditional multiple evaluation systems in college English teaching, which integrate summative assessment

(e.g., final exams, CET-4/6 scores) and formative assessment (e.g., classroom performance, homework), have made some progress compared with single test-oriented evaluation. However, they still face many practical dilemmas: heavy workload for teachers in manual evaluation leads to low efficiency; evaluation dimensions are limited to knowledge mastery, ignoring the assessment of practical communication and innovative thinking; feedback is often delayed and lacks personalization, failing to meet the needs of differentiated learning [3].

The rapid development of AI technology, represented by natural language processing, machine learning, and big data analytics, has injected new vitality into the reform of educational evaluation [4]. AI-enabled evaluation systems can automatically collect and analyze multi-source learning data, realize real-time feedback and dynamic tracking, and thus break through the bottleneck of traditional evaluation. Therefore, exploring the innovative path of AI-enabled multiple evaluation systems in college English teaching is of great significance for improving teaching quality and promoting students' all-round development [5].

2. Connotation and limitations of traditional multiple evaluation systems

The concept of "multiple evaluation system" originated from the criticism of standardized testing in the 1990s, emphasizing the integration of multiple evaluation subjects (teachers, students, peers), methods (tests, observation, interviews), and dimensions (knowledge, skills, attitudes) to achieve a comprehensive and objective reflection of students' learning processes and outcomes [6]. In college English teaching, this system is designed to avoid the one-sidedness of "one test determining success or failure" and promote the balanced development of students' language competence.

In practice, however, traditional multiple evaluation systems have obvious limitations:

(1) Low efficiency of evaluation implementation: Teachers need to spend a lot of time on grading papers, organizing peer reviews, and sorting out evaluation data. A survey shows that college English teachers spend an average of 30% of their working hours on evaluation-related tasks, which restricts the depth of teaching research [7].

(2) Narrowness of evaluation dimensions: Most evaluations focus on the accuracy of vocabulary and grammar, while neglecting the assessment of practical skills such as cross-cultural communication, critical thinking, and collaborative learning. For example, oral evaluation often stays at "pronunciation accuracy" and ignores "discourse coherence" and "contextual adaptation" [3].

(3) Lag and superficiality of feedback: Students usually receive only simple scores or general comments after the evaluation, lacking specific guidance on how to improve. A study on 500 college students found that 68% of them thought the feedback from traditional evaluation was "not helpful for adjusting learning strategies" [8].

3. AI-enabled innovation paths of multiple evaluation systems

AI technology provides technical support for breaking through the limitations of traditional evaluation and promotes the innovation of multiple evaluation systems in the following aspects.

3.1. Diversification of evaluation subjects

AI systems can act as a new evaluation subject, forming a collaborative evaluation model involving "teachers + AI + students + peers." For example, AI-powered writing evaluation tools (e.g., iWrite, Grammarly) can automatically assess students' compositions from dimensions such as grammar, vocabulary diversity, discourse

structure, and thematic relevance, and generate detailed modification suggestions, reducing teachers' workload by 50–70% [5]. At the same time, AI can assist peer evaluation by providing evaluation rubrics and similarity analysis of reviews, helping students improve the objectivity of their evaluations. For instance, in a university's English writing course, after introducing AI-assisted peer evaluation, the consistency between peer scores and teacher scores increased from 62% to 85% [8].

3.2. Enrichment of evaluation dimensions

AI enables evaluation to cover more comprehensive dimensions beyond traditional knowledge assessment. With the support of natural language processing technology, AI can analyze students' oral dialogues to assess their "pragmatic competence" (e.g., appropriate response in specific contexts) and "discourse competence" (e.g., topic maintenance and turn-taking) [4]. Machine learning algorithms can track students' online learning behaviors (e.g., frequency of watching English videos, duration of participating in online discussions, accuracy of answering questions) to evaluate their "autonomous learning ability" and "learning persistence". In addition, AI can identify "innovative thinking" in students' works through semantic analysis, such as unique viewpoints in argumentative essays and creative expression in role-playing [5].

3.3. Intelligence of evaluation feedback

AI realizes the transformation of feedback from "post-event summary" to "process guidance." Through real-time data analysis, AI can provide immediate feedback for students' learning behaviors: for example, in vocabulary exercises, AI can point out the confusion between "affect" and "effect" and push targeted example sentences; in oral practice, it can identify pronunciation biases (e.g., confusion between /θ/ and /ð/ in southern students) and provide correction demonstrations [7]. Moreover, AI can generate personalized learning portraits based on long-term data tracking, which include students' strengths (e.g., good at narrative writing), weaknesses (e.g., weak in academic writing), and potential development directions, helping teachers carry out differentiated teaching.

3.4. Application of generative AI in dynamic evaluation

The rise of generative AI (e.g., ChatGPT, Claude) has further expanded the boundaries of AI-enabled evaluation. Unlike traditional rule-based AI tools, generative AI can conduct contextualized interaction with students during the evaluation process, realizing a shift from "one-way assessment" to "interactive guidance." For example, in writing evaluation, after identifying logical flaws in students' argumentative essays (e.g., the gap between "environmental pollution" and "economic development" arguments), ChatGPT can generate 2–3 revised versions with explanatory notes (e.g., "Adding data on 'green GDP' can bridge the logical gap"), which helps students understand the essence of improvement [9].

In oral assessment, generative AI combined with real-time speech recognition can simulate "dialogue scenarios" (e.g., job interviews, academic discussions) and dynamically adjust the difficulty of questions based on students' responses. A practice in a university in Zhejiang shows that after using this mode, students' "discourse complexity" (measured by the ratio of complex sentences) increased by 18% within a semester, and their ability to respond to unexpected questions improved significantly [10].

4. Practical case of AI-enabled multiple evaluation

To verify the effectiveness of the AI-enabled multiple evaluation system, this study selected a comprehensive

university in Jiangsu Province as the research object. The university has applied an AI-based college English evaluation platform since September 2022, which integrates functions such as automated writing evaluation, oral assessment, learning behavior analysis, and multi-subject collaborative evaluation.

After one academic year of practice, the results show that:

(1) Evaluation efficiency has been significantly improved: The time for teachers to grade writing and oral tasks has been reduced by 65%, and the number of formative assessments per semester has increased from 3 to 8, realizing the dynamic monitoring of students' learning processes.

(2) Students' learning motivation has been enhanced: A questionnaire survey of 320 students shows that 83% of them are satisfied with the real-time feedback of AI, and 79% of them actively adjust their learning plans according to the evaluation reports.

(3) Students' comprehensive language competence has improved: The average score of the final oral communication assessment increased by 13%, and the proportion of students who can put forward innovative viewpoints in writing increased by 18% [8].

This case shows that AI-enabled multiple evaluation systems can effectively promote the reform of college English teaching, but it also finds that 17% of students rely too much on AI feedback and ignore teacher guidance, reflecting the need to balance the roles of AI and humans.

5. Challenges and countermeasures

5.1. Main challenges

(1) Algorithmic bias: AI algorithms are trained based on existing language corpora, which may inherit biases in the data. For example, some writing evaluation systems over-penalize dialectal expressions (e.g., "you guys" in American English) or creative writing that deviates from standard templates, leading to unfair evaluation [6].

(2) Data privacy risks: The platform involves a large amount of students' data, including test scores, learning behaviors, voiceprints, and facial expressions in video interviews. Once leaked, it may cause privacy violations [4].

(3) Weakening of humanistic care: Over-reliance on AI evaluation may reduce face-to-face communication between teachers and students. Some students reported that "AI feedback is accurate but lacks warmth, and teacher comments with emotional care are more encouraging" [5].

The popularity of generative AI has brought new challenges to the authenticity of evaluation results. A national survey of 2,000 college students in 2024 showed that 42% of students admitted to using AI to complete English homework (e.g., essays, oral scripts), and 68% of teachers reported that it was difficult to distinguish between AI-generated and student-original works using traditional plagiarism detection tools [11]. This not only makes the evaluation lose its diagnostic function but also undermines the cultivation of students' independent thinking.

To address this issue, a "dual-track verification" mechanism can be established: On the one hand, use AI-generated text detection tools (e.g., Originality.ai, which claims to have a detection accuracy of over 95% for text generated by GPT-4) to conduct preliminary screening; on the other hand, implement "process tracking"—require students to submit draft records (including revision history), AI usage logs (e.g., which parts are revised by AI and why), and oral defense of key viewpoints. A pilot in Wuhan University found that this mechanism reduced the proportion of AI plagiarism in English courses from 35% to 8% within a semester [11].

5.2. Countermeasures

(1) Optimize AI algorithms: Establish a multi-source corpus covering diverse language styles and cultural backgrounds, invite college English teachers and linguists to participate in algorithm debugging, and set up "human-machine double-check" mechanisms for key evaluation links (e.g., innovative writing).

(2) Strengthen data security: Formulate strict data management regulations, encrypt sensitive information such as voiceprints and facial images, and clarify that data can only be used for educational evaluation without permission for other purposes.

(3) Balance AI and human roles: Position AI as an auxiliary tool. Teachers should focus on evaluating "non-cognitive factors" (e.g., learning attitudes, collaborative spirit) that are difficult for AI to assess, and maintain regular communication with students to provide emotional support and educational guidance [6].

6. Conclusion

AI technology has brought revolutionary changes to the multiple evaluation systems in college English teaching, realizing the transformation from "static, single, and lagging" to "dynamic, comprehensive, and intelligent." By empowering the diversification of evaluation subjects, enrichment of dimensions, and intelligence of feedback, it effectively promotes the improvement of teaching efficiency and students' language competence.

However, the application of AI also faces challenges such as algorithmic bias and data risks. In the future, we should adhere to the principle of "technology serving education," continuously optimize the integration of AI and evaluation systems, and ensure that the innovation of evaluation systems ultimately serves the all-round development of students. Only in this way can the multiple evaluation systems in college English teaching truly play their role in guiding and promoting teaching reform.

Funding

The 2025 Jilin Provincial Institute of Higher Education Project "Research on Innovation of AI-Enabled Multiple Evaluation System in College English Teaching" (Project No.: JGJX25D1033)

Disclosure statement

The authors declare no conflict of interest.

Author contributions

Research design - Chen Shuo
Draft writing – Chen Shuo
Data analysis – Yang Jing

References

[1] Brown H, 2001, Teaching by Principles: An Interactive Approach to Language Pedagogy (2nd ed.). Pearson Education, Canada.

[2] Black P, Wiliam D, 1998, Assessment and Classroom Learning. Assessment in Education: Principles, Policy & Practice, 5(1): 7–74.

[3] Chapelle C, 2012, Assessing Language through Computer Technology. Cambridge University Press, London.

[4] Dillenbourg P, Jermann P, 2010, Technology-Enhanced Learning. In: Sawyer K (ed.), Cambridge Handbook of the Learning Sciences. Cambridge University Press, 475–491.

[5] Heift T, Schulze M, 2007, Intelligent Computer-Assisted Language Learning: A Survey. Computer Assisted Language Learning, 20(3): 223–250.

[6] Lynch B, 2001, Language Assessment and Program Evaluation. Cambridge University Press, London.

[7] Warschauer M, Matuchniak T, 2010, New Technology and Digital Worlds: Analyzing Evidence of Equity in Access, Use, and Outcomes. Review of Research in Education, 34(1): 179–225.

[8] Zheng C, Li M, 2022, AI-Based Formative Assessment in EFL Writing: A Case Study. System, 106: 102824.

[9] Li Y, Wang H, 2023, Generative AI in College English Writing Assessment: A Comparative Study of ChatGPT and Human Evaluators. Language Testing, 40(2): 215–238.

[10] Chen L, Zhang Y, Zhao J, 2024, Multimodal Data Fusion for AI-Enabled Formative Assessment in EFL Classrooms: Integrating Text, Speech, and Gesture. Computers & Education, 198: 104987.

[11] Zhang Q, Liu J, 2025, Ethical Guidelines for AI-Driven Language Assessment: Addressing Bias, Privacy, and Human Oversight. TESOL Quarterly, 59(1): 89–112.