Dean&Francis

ISSN 2959-6149

The Role and Impact of Artificial Intelligence in Personalized English Teaching Evaluation

Tian Xia

Department of Translation, Lingnan University, Hong Kong, 999077, China
xiatian8426@outlook.com

Abstract:

The widespread application of artificial intelligence (AI) in English teaching evaluation has drawn extensive attention due to its advantages which AI can boost evaluation efficiency and offer personalized feedback. However, some deficiencies still exist, standing for technical function limitations, potential negative impacts on the teaching process, and constraints on implementation conditions. This study concentrates on the application of AI in English teaching evaluation to analyze its technical function shortcomings, potential negative impacts on the teaching process, and capability constraints during implementation. The study finds that AI can enhance evaluation accuracy via multi-dimensional data integration, diversified corpus expansion, and an AI-human hybrid model. Besides, AI application capabilities of teachers can be improved through hierarchical training and continuous learning mechanisms. Based on these findings, the study proposes improvement strategies from technical optimization and teacher training aspects for promoting the in-depth integration of AI and English teaching, advancing educational equity, and improving teaching quality.

Keywords: Artificial Intelligence (AI); English Teaching Evaluation; Technical Optimization; Teacher Training; Educational Equity

1. Introduction

The acceleration of globalization and internationalization is driving a surge in the demand for cross-cultural communication. Therefore, the importance of English is gradually becoming more prominent as an international language, so more learners are requiring more personalized and efficient language learning plans in English learning [1].

While teaching evaluation is essential for improving learning outcomes. It is difficult to meet the differentiated needs of various learners in different proficiency levels since traditional evaluation has problems with delayed feedback. Additionally, traditional evaluation lacks personalization. It is difficult for teachers to cover every error pattern of each learner.

The breakthroughs in Artificial Intelligence Technology offer the possibility of solving the mentioned problems. According to Chen's research, Natural Language Processing models based on the Transformer architecture such as BERT-4 can achieve an accuracy of 94.8% in detecting grammatical errors [2]. Besides, adaptive learning platforms such as Pearson Test of English AI can track learning progress in real time [3].

However, there are still shortcomings in the application of AI technology. On the one hand, AI cannot evaluate semantic meaning sufficiently. The recognition accuracy rate of deep language phenomena like metaphors and idioms is less than 50% [4]. On the other hand, over-reliance on AI may lead to the degradation of learners' autonomous error correction capability. An experiment from Okonji and Ashabua showed that learners who had long used AI feedback had about 18% higher error rate in unassisted writing than the control group [5].

This study mainly focuses on the role and positive impact of AI can supplementing personalized English language teaching evaluation to explore the specific role, positive influence, existing limitations, and optimization suggestions of AI in personalized evaluation to assist teaching strategies in the education field. In addition, references for enhancing the accuracy and effectiveness of evaluation can be further provided by promoting the deep integration of AI technology and personalized English teaching evaluation.

2. The Effect and Positive Impact of AI in Personalized English Language Teaching Evaluation

2.1 Providing a Personalized Diagnosis to Accurately Identify the Shortcomings in Capabilities

AI can achieve almost full coverage of dimensions and precise individual diagnosis through multimodal technology. In the writing evaluation, the system based on GPT-4 can identify grammatical errors, syntactic problems, and label high-frequency errors in combination with historical data. For example, the proportion of incorrect word segmentation usage among learners in the past participle learning reached 62% [2]. Additionally, the pragmatic adaptability related to polite expression in business scenarios or academic scenarios can even be adapted to variants such as Indian English and Southeast Asian English to avoid standard accent bias [4].

The feedback delay in traditional evaluation leads to the solidification of learners' incorrect memories while AI can achieve submitting and receiving feedback. For example, the Grammarly Business version can mark errors in real time during text input and provide "error reasons + modification examples", such as visualizing the difference in usage between the word phrase despite and the word despite [5]. AI builds a capability growth map through learning and

2.2 Positive Influence on Learners and Teachers

2.2.1 Positive influence on learners

For English learners, the application of AI in personalized language learning such as precise weakness diagnosis and real-time progress tracking, can achieve the dual improvement of learning efficiency and learning motivation. AI-generated targeted practice recommendations allow learners to concentrate on their learning weaknesses and avoid ineffective repetitive training. After analyzing the result of AI real-time detection of learners' oral anxiety and making adaptive adjustments to the difficulty of practice, Tajik ultimately concluded that the efficiency of learners' oral practice increased by 37% meanwhile the frequency of active participation rose from 2.1 times per week to 4.3 times. The empirical evidence supports this effect [4].

AI systems that provide personalized learning plans, like step-by-step task lists for overcoming specific language weaknesses and progress-tracking features, play a key role in the enhancement of English learners' autonomous learning ability. AI enables learners to independently control the frequency and content of evaluations. It breaks the time and space limitations of traditional classrooms. Ali found that 70% of Swedish high school ESL learners can boost their self-planning by 38% by using AI-generated paths to tackle article errors, practice complex sentences, and simulate talks. Additionally, 62% students can adjust plans via AI feedback, like adding subjunctive mood exercises [6]. Tikiz Erturk and Kurt Taspinar noted that AI Learning Logs record error fixes, which allows learners to review weak points and set goals [7]. This closed loop of feedback to review to planning mode significantly enhances the pertinence of self-study.

2.2.2 Positive influence on teachers

In teaching contexts, the application of AI in automated evaluation significantly reduces the workload of teachers spent on repetitive assessments such as marking vocabulary quizzes or verifying sentence structure correctness. It indeed frees up time for teachers to engage in advanced teaching design and further differentiates instructional strategies for learners with different language weaknesses or design project-based learning modules that connect English learning to real-life needs. AI dramatically re-

ISSN 2959-6149

duces teachers' mechanical working hours by automating repetitive evaluation tasks. Specifically, Chen's research indicates that AI can automatically complete tasks such as scoring multiple-choice listening questions with an accuracy rate of 98.2% and correcting grammatical errors, including subject-verb agreement, singular and plural nouns, to verify writing formats like citation formats and paragraph spacing [2]. Furthermore, the AI Evaluation Report automatic generation function mentioned by Okonji & Ashabua provides convenience for teachers. The system can automatically summarize the top 3 class errors aimed at article misuse, tense confusion, and prepositional collocation errors. The efficiency of teaching decision-making has increased by 40% because teachers can quickly grasp the overall situation of the class according to learner progress rankings without manual statistics [5].

The error patterns analysis and the construction of learner profiles from AI provide teachers with an objective and detailed teaching basis. Alhusaiyan's review indicates that AI can statistically analyze the distribution of errors at the class level [8]. Ali's classroom experiments also demonstrated that the coverage rate of differentiated teaching in the class could increase from 45% to 82% after teachers adjusted their teaching by referencing data from AI recording [6]. Through AI, advanced challenge tasks such as complex text analysis can be designed for average students, and weak skill enhancement packages can be designed for students with weak oral skills but strong writing skills. This has changed the previous situation where only top students or underachievers could be considered which significantly improved the accuracy of teaching.

3. The Existing Limitations

3.1 The Limitations of Technical Functions

However, the limitations of technical functions of AI still exist in English teaching evaluation, manifest in multiple aspects, which can be performed in inaccurate judgment of deep language logic, cultural and temporal biases in data support even the insufficient adaptability to non-standard or specialized language scenarios. The evaluation accuracy of AI in high-level capabilities such as semantic understanding and cultural adaptability is limited. Moreover, judgment biases are very likely to occur in complex language scenarios.

In writing evaluations, AI can efficiently detect basic errors such as grammar and spelling, but it lacks a good grasp of the logic of deep-seated arguments and the coherence of content. The research conducted by Okonji and Ashabua demonstrated that the accuracy rate of AI in evaluating the correlation between evidence and thesis in argumentative essays is only 65% [5]. AI often misjudges paragraphs with concisely expressed examples as lacking sufficient evidence. Furthermore, it is more difficult for AI to identify the rationality of literature citations or the accuracy of research method descriptions in academic writing. For instance, AI is unable to distinguish the essential difference between an incorrect citation format and the content of the citation being irrelevant to the topic.

In oral evaluation, AI's recognition of cross-cultural pragmatic ability is even weaker. The Tajik's experiment found that Arab learners tend to use euphemistic and roundabout sentence patterns, such as I wonder if to make requests in oral expression. However, the probability that this is misjudged by AI as a redundant expression is as high as 42%. Moreover, due to Japanese learners' cultural habit of using humble interjections such as maybe or perhaps are often marked by AI as having an ambiguous attitude, which ultimately leads to a disconnection between the evaluation results and the actual communication effects [4].

In addition, the training data of AI evaluation models have problems such as cultural bias and lagging timeliness. Alhusaiyan's systematic review discovered that the current mainstream AI evaluation systems are mostly trained based on English corpora from Europe and America. This has poor adaptability to variants such as African English and Southeast Asian English. The accuracy rate of evaluating intonation and rhythm in Nigerian English is also 23% lower than that in standard American English. There are often misjudgments in handling characteristic words such as "lah" in Singaporean English [8].

3.2 The Limitations of Students' Subjective Initiative in Learning

Long-term reliance on AI feedback may lead to a weakening of both learners' autonomous error correction ability and language creativity, which would ultimately result in a learning predicament that depends on technology. Okonji and Ashabua's following experiment revealed that EFL learners who used AI grammar tools for six consecutive months had an 18% higher error rate in the AI-assisted writing test than the control group that did not use AI [5]. Moreover, their frequency of actively consulting grammar books, dictionaries, and other learning resources has decreased by 70%. Learners gradually lose the ability to independently analyze the reasons for their mistakes and only rely on the directly provided correction answers by AI. Furthermore, the standard answer orientation of AI can limit language creativity. Ali's survey of ESL learners in Swedish high schools indicated that to avoid AI marking grammatical risks73% of the learners tend to deliberately abandon the use of complex sentence structures such

as attributive clauses and subjunctive moods. Instead, they would choose simple sentences which leads to monotonous writing content and stagnant expression ability. In oral practice, learners may also reduce their experimental expressions for fear of being deducted points by AI, so they would be reluctant to use newly learned vocabulary or sentence patterns, which ultimately affects the advancement of their language skills [6].

3.3 The Constraints of Implementation Conditions

The popularization of AI evaluation tools is limited by hardware resources such as devices, networks, and funds. The digital divide in some regions has led to an even more unbalanced distribution of educational resources. Alshewitter et al. found that 25% of rural college students in Jordan encountered difficulties in oral evaluations of AI. The instability of the network can lead to a deviation of up to 30% in the assessment results. Furthermore, 18% of the students lack personal equipment and practice only once a week, while urban students can practice 4.2 times a week [9]. This resource disparity not only affects spoken English but also means that learners with professional AI tools could make much faster progress in overall language proficiency, thereby deepening the educational gap in all aspects of language learning.

The insufficient ability of teachers to integrate AI with teaching has also become a key bottleneck restricting the application of AI evaluation. The training content did not cover core abilities such as how to interpret AI data or how to adjust teaching based on error patterns. When AI shows that the proportion of past perfect tense errors in a class is 75%, most teachers only know how to talk about the past perfect tense. They do not use AI to further analyze whether the errors are concentrated in the collocation of time adverbials or subject-verb agreement. The insufficient teaching pertinence becomes more apparent.

Alhusaiyan's review also pointed out that the current training is mostly one-off and short-term courses. These courses cannot keep up with the update pace of AI technology. For instance, the GPT model is updated every six months to include new functions such as real-time dialogue evaluation and interdisciplinary writing guidance [8]. Most teachers still use the old operation methods because of the lack of subsequent training, so that the technical value cannot be fully exploited. In addition, the compatibility issues of AI evaluation tools also affect the implementation effect. Tikiz Erturk & Kurt Taspinar found that there was a problem of data non-interoperability between the teaching management systems and AI evaluation tools used in schools. Teachers need to manually input the AI evaluation data into the teaching system, which adds extra workload and may lead to data errors due to operational mistakes [7].

4. Recommendations

4.1 Optimizations of AI Technology in Teaching Evaluation

The limitations of AI evaluation can be reduced by integrating multimodal data such as text, speech, and vision to construct a comprehensive evaluation model in semantic understanding and emotional expression.

In verbal evaluation, the Amazon Alexa Speak system developed by Tajik has achieved the determination of the appropriateness of learners' emotional expression through data by combining speech features with facial micro-expressions and eye contact states. For instance, when the system detects that learners speak too fast and frown, it will assess whether their emotional transmission is sincere in combination with the context of the conversation. In this way, the accuracy rate of pragmatic fit in oral evaluation can be increased by 32% [4].

In the writing evaluation, Okonji & Ashabua proposed the introduction of the argument relevance algorithm. The data of learners during the writing process can be combined to track the traces of revision, the duration of thinking, etc., to distinguish logical loopholes from expression omissions, thereby avoiding the situation where temporary clerical errors are misjudged as logical flaws [5]. In addition, a contextual knowledge base can be introduced to assist evaluation for complex scenarios such as literary texts and cross-cultural dialogues. For instance, AI can invoke the Literary Metaphor Database, which contains the cultural connotations of common images, such as roses symbolizing love in poetry evaluation, to combine the text semantics and cultural background to determine whether the use of metaphors is appropriate. Linking the cultural custom library, such as the differences in etiquette expressions among different countries, is helpful to assess cross-cultural writing. For instance, the lengthy greetings of Arab learners can be avoided from being misidentified as redundant content. Thus, the deviation rate of cross-cultural pragmatic evaluation was reduced by 28% [8].

To address the issues of cultural bias and outdated data in AI evaluation models, it is recommended to build a cross-cultural English evaluation corpus that covers the entire world and is dynamically updated. This corpus should cover English variants from over 100 regions and label language usage norms in different cultural contexts.

To balance the efficiency of evaluation and the depth of high-level ability assessment, it is necessary to clarify

ISSN 2959-6149

the division of responsibilities between AI and teachers. Establishing a collaborative evaluation mechanism can be effective. Chen proposed a 4:6 division of labor model. AI is responsible for the basic ability assessment of repetitive tasks such as grammar error correction, pronunciation accuracy scoring, and automatic marking of objective questions to reduce the mechanical labor pressure on teachers. This accounts for 40% in the whole evaluation. Teachers focus on high-level ability assessment which accounts for the remaining 60% including dimensions that are difficult for AI to accurately judge, such as the logical depth of writing content, the cultural adaptability of oral expression, and the appropriateness of creative language use [2]. Specifically, AI can mark grammatical errors in compositions. Teachers can further assess whether the errors affect the understanding of the content and determine whether students need targeted teaching intervention. Meanwhile, Alhusaiyan emphasized that AI needs to add the function of uncertainty annotation. For the content with an evaluation result credibility lower than 70% like ambiguous emotional expression or unfamiliar cultural scenarios, it should be clearly marked that the evaluation credibility is low. It is recommended to manually recheck the marking and provide a summary of the evaluation basis to help teachers quickly grasp the key points of the recheck [8].

4.2 Advanced Teaching Training Supplement with AI Technology

Differentiated training systems should be designed for teachers with different teaching experiences and technical foundations. Since the rapid iteration of AI technology, it is necessary to establish a long-term training methodology to prevent teachers from giving up using it due to technological disconnection. To address this issue, a teacher AI learning community can be established to promote experience sharing and problem-solving.

5. Conclusion

In conclusion, this study explored how AI impacts personalized English teaching evaluation; meanwhile, it has some limitations. AI demonstrates clear strengths. It indeed enables personalized diagnosis of learners' weaknesses via multimodal technology and provides real-time feedback. These assistances boost learning efficiency and participation, thus reducing teachers' repetitive workload while supporting differentiated instruction. However, its key limitations persist. AI has technical flaws in judging deep language logic and cultural biases in data. This weakened learner autonomy and creativity due to over-reliance and implementation barriers from uneven resource access. To address these, optimizing AI technology and enhancing teacher training are critical. These measures will drive deeper AI-English teaching integration to advance educational equity and improve overall teaching quality.

References

[1] Dekhakhena A. AI-Powered Personalized Learning in EFL Acquisition: Exploring Adaptive Instruction and Feedback Systems. Journal of Studies in Language, Culture and Society (JSLCS), 2025, 8(1): 111-131.

[2] Chen Y. Transforming English Language Education with AI-Driven Deep Learning Models for Scalable Adaptive and Inclusive Assessment. International Journal of Information and Communication Technology, 2025, 26(12): 15–31.

[3] Hussain SM, Jabeen Satti SM, Khan Z, Safia U-e. AI-Powered Personalized Learning: Advancing Language Education in the Digital Era. Journal of Social Signs Review (JSSR), 2024, 2(4): 730–740.

[4] Tajik A. Integrating AI-Driven Emotional Intelligence in Language Learning Platforms to Improve English Speaking Skills through Real-Time Adaptive Feedback. Preprints.org, 2025.

[5] Okonji NN, Ashabua DA. Enhancing Writing Skills through Automated Feedback: An Exploration of Artificial Intelligence and Natural Language Processing in Language Learning. Education For Today, 2025, 21(1).

[6] Ali AM. An Intervention Study on the Use of Artificial Intelligence in the ESL Classroom: English Teacher Perspectives on the Effectiveness of ChatGPT for Personalized Language Learning. University of Malmö, 2023.

[7] Tikiz Ertürk G, Kurt Taspinar H. Applications of Artificial Intelligence (AI) in EFL Classrooms. In: Digital Pedagogy in the 21st Century: Emerging Technologies in Foreign Language Classrooms, 2023: 27–44.

[8] Alhusaiyan E. A Systematic Review of Current Trends in Artificial Intelligence in Foreign Language Learning. Saudi Journal of Language Studies, 2025, 5(1): 1–16.

[9] Alshewiter KM, Ahmad Mohammad FSS, Shawaqfeh AT, Khasawneh AJ, Alqudah H, Khasawneh YJA, Khasawneh MAS. Improving the Learning of Language Proficiency at Tertiary Education Level through AI-Driven Assessment Models and Automated Feedback Systems. Migration Letters, 2024, 21(2): 712–726.