WHIOCE PUBLISHING PTE. LTD.
PROVIDING
FIRST-CLASS SCIENTIFIC INFORMATION
FOR TOP SCHOLARS

2026 Volume 4, Issue 2
Educational Innovation Research
ISSN: 3029-1844(Print) ISSN: 3029-1852(Online)

AI-Empowered Teaching Evaluation for "Integration of Basic and Professional English" in Business Schools

Jingwei Zhu*
Jiangsu Maritime Institute Nanjing 211119, Jiangsu, China

**Author to whom correspondence should be addressed.*

Abstract: With the deepening integration of basic and professional English teaching in business schools, the traditional evaluation system has obvious limitations (static indicators, subjective grading, delayed feedback), which fail to measure students' comprehensive English application capabilities and meet the demand for compound business talents. This proposal focuses on the reform of the "Integration of Basic and Professional English" teaching evaluation, exploring the application of AI technology by constructing the "4D Evaluation Model" and developing the EconEvalAI platform to realize multimodal analysis, real-time diagnostics and dynamic benchmarking. It details the implementation framework, technological innovations, and solutions to potential challenges, aiming to transform summative evaluation into a continuous improvement-oriented mechanism. Expected outcomes include shortened evaluation cycles, improved evaluation reliability and student satisfaction, providing a feasible reference for the digital transformation of English teaching evaluation in business schools.

Keyword: AI-empowered teaching evaluation; Integration of Basic and Professional English; Business schools; Teaching reform; Multimodal analysis; 4D Evaluation Model; Intelligent evaluation platform; Cross-cultural communication evaluation

Online publication: February 26, 2026

1. Background and Rationale

1.1. Current Challenges in Evaluation

Against the backdrop of the deep integration of basic and professional education, the teaching of "Integration of Basic and Professional English" in business schools has shifted to a "language + professional" model, focusing on students' ability to apply English in professional business scenarios. However, the traditional assessment system, formed in the pure language teaching era, has prominent limitations that hinder its effectiveness, mainly reflected in the following aspects:

1.1.1. Static Metrics that Deviate from Professional Practice

The traditional system over-relies on standardized tests like TOEIC and written exams, which only measure basic language skills but fail to assess students' ability to use professional terminology and solve practical business problems. This leads to the "high scores but low ability" phenomenon, as a high TOEIC score does not guarantee proficiency in business negotiations or contract drafting. Liu, H., & Wang, L. (2021) [1]confirmed that the traditional system lacks pertinence to professional scenarios, making it hard to reflect actual learning effects.

1.1.2. Subjectivity Bias in Manual Grading

Practical links like oral negotiations rely on manual grading, which is inconsistent due to teachers'different backgrounds and standards. Asurvey of 50 business English teachers shows the inter-rater reliability (IRR) of oral grading is only 0. 68, far below the ideal 0.85. Chen, Y., & Li, J. (2023) noted that this bias restricts the effectiveness of formative assessment[2].

1.1.3. Delayed Feedback that Fails to Guide Real-Time Improvement

Students often receive feedback weeks or months after assessments, making it impossible to timely correct deficiencies. This breaks the "learning-practice-evaluation-improvement" loop, reducing evaluation's guiding value.

1.2. AI Transformation Potential

The rapid development of AI technology (NLP, ASR, machine learning) provides a solution to the traditional evaluation's deficiencies, enabling transformation from "static, subjective, delayed" to "dynamic, objective, real-time". Its potential is reflected in three aspects:

1.2.1. Multimodal Analysis for Comprehensive Evaluation

AI integrates speech, text, and video data to evaluate speech fluency, professional term accuracy, and cross-cultural sensitivity simultaneously. For example, in simulated negotiations, it can analyze term use, speech fluency, and even body language, fully reflecting students' integrated capabilities[3].

1.2.2. Real-Time Diagnostics to Guide Timely Improvement

AI provides instant feedback during practice, such as correcting term misuse or adjusting negotiation strategies. This helps students and teachers adjust in real time, forming an effective improvement loop and enhancing teaching effect.

1.2.3. Dynamic Benchmarking for Personalized Development

Machine learning analyzes students' historical data to set personalized benchmarks, focusing on progress rather than absolute levels. This mobilizes learning enthusiasm and provides targeted guidance, consistent with Mitrović, A., & Suraweera, P. (2022)[4]'s intelligent tutoring concept.

2. Reform Objectives

2.1. Core Goals

The core goal is to establish an AI-empowered, scientific, and dynamic evaluation system to accurately measure students' capabilities and promote teaching reform. Specific goals are as follows:

2.1.1. Develop an AI-Driven Evaluation Framework

Construct the "4D Evaluation Model" (Basic Language Skills, Professional Application Capability, Cross-Cultural Communication Competence, Learning Progress), with weights 30%, 40%, 20%, and 10% respectively. It integrates process and result evaluation, referring to Liu, H., & Wang, L. (2021) and Baker, R. S., & Inventado, P. S. (2022)'s research[3].

2.1.2. Build the EconEvalAI Platform

Develop an intelligent platform with core modules: ASR with professional term highlighting, negotiation strategy sentiment analysis, comparative performance dashboards, adaptive assessment, and privacy protection (referring to Zhang, Q., & Chen, Z. (2024)[5].

2.2. Expected Outcomes

Table 1. Clear baseline and target data are set to reflect reform effects:

Indicator	Baseline	Target
Evaluation turnaround time	14 days	≤ 2 hours
Inter-rater reliability (IRR)	0.68	≥ 0.89
Student satisfaction with feedback	3.1/5	≥ 4.5/5

These outcomes align with teaching reform goals, aiming to make evaluation a tool for student development and teaching quality improvement.

3. Implementation Framework

3.1. AI Evaluation System Architecture

To ensure the stable operation and effective implementation of the AI-empowered teaching evaluation system, the EconEvalAI platform adopts a four-layer architecture, which is designed based on the principles of modularization, scalability, and security.

3.2. Evaluation Process Redesign

To adapt to the AI-empowered "4D Evaluation Model" and the EconEvalAI platform, the traditional evaluation process is comprehensively redesigned, breaking the original "assessment-feedback" two-stage process, and establishing a "formative evaluation-summative evaluation-diagnostic evaluation-iteration improvement" four-stage closed-loop evaluation process. This new process emphasizes the whole learning process and continuous improvement, and its specific comparison between the traditional approach and the AI-empowered approach in each stage is as follows (**Table 2**):

Stage	Traditional Approach	AI-Empowered Approach
Formative	Teacher-written comments based on weekly homework and occasional practice, with delayed feedback	AI generates "micro-feedback" on term usage errors, grammar mistakes, and strategy inappropriateness during practice negotiations, oral presentations, and professional document writing, with real-time feedback
Summative	Human-scored role plays and written examinations at the end of the semester, with single evaluation standards and subjective bias	System evaluates 360° performance using video/audio analysis of final negotiations, professional document drafting, and cross-cultural communication scenarios, combining quantitative and qualitative evaluation
Diagnostic	One-size-fits-all tests at the beginning and end of the semester, which can only reflect the overall level but not individual gaps	Adaptive assessment adjusts difficulty based on real-time learner performance, and generates personalized diagnostic reports

The redesigned evaluation process integrates formative evaluation, summative evaluation, and diagnostic evaluation, and realizes the organic combination of process evaluation and result evaluation. Through real-time feedback and personalized iteration improvement, the evaluation process truly becomes a powerful tool to promote students' learning and improve teaching quality, which is fully consistent with the core concept of "Integration of Basic and Professional English" teaching.

4. Technological Innovation

4.1. Core AI Components

The core of the AI-empowered teaching evaluation system lies in the two core AI components: the Professional Term Recognition Engine and the Cross-Cultural Competence Analyzer. These two components are independently developed based on the characteristics of "Integration of Basic and Professional English" teaching in business schools, trained on a large number of business English corpora and cross-cultural communication cases, and have high professionalism and accuracy. Their specific functions and technical characteristics are as follows:

4.1.1. Professional Term Recognition Engine

This engine is the core component for evaluating the Dimension of Professional Application Capability in the "4D Evaluation Model", mainly responsible for identifying and evaluating the accuracy and appropriateness of students' use of professional terms in business scenarios. To ensure the professionalism and accuracy of the engine, it is trained on a diverse and extensive dataset of more than 50,000 annotated business documents, including SEC filings, WTO dispute resolutions, international trade contracts, financial reports, marketing plans, and legal documents. These documents cover various professional fields such as international trade, finance, marketing, and law, which are closely related to the professional directions of business schools. In the training process, the engine uses advanced NLP algorithms to extract and annotate professional terms, establish a professional term corpus with context information, and realize the accurate recognition of professional terms in different business contexts[4].

4.1.2. Cross-Cultural Competence Analyzer

This analyzer is the core component for evaluating the Dimension of Cross-Cultural Communication Competence in the "4D Evaluation Model", mainly responsible for detecting cultural taboos and sensitivities in students' cross-cultural communication, and evaluating their cross-cultural communication ability. In the globalized business environment, cross-cultural communication ability has become an essential quality for compound business talents, and any inappropriate cross-cultural communication behavior may lead to business failure. The analyzer is designed to help students avoid cross-cultural communication misunderstandings and improve their cross-cultural adaptation ability. Specifically, the analyzer can detect inappropriate language expressions, behaviors, and strategies in students' cross-cultural communication[6].

4.2. Privacy Protection Mechanisms

In the process of AI-empowered teaching evaluation, a large amount of students' personal information and learning data (such as speech recordings, video materials, and evaluation results) will be collected and processed, which involves the important issue of data privacy protection. To ensure the security of students' data and comply with relevant data protection laws and regulations such as the GDPR and China's "Personal Information Protection Law", the EconEvalAI platform adopts a variety of strict privacy protection mechanisms, which are designed with reference to the research results of Zhang, Q., & Chen, Z. (2024) on privacy-preserving AI in educational evaluation[5]. The specific mechanisms are as follows:

4.2.1. On-Device Processing

Speech data, video materials, and other sensitive data generated by students during the assessment process are processed and analyzed locally on the student's tablet or computer, without ever leaving the device. This eliminates the risk of data leaks or breaches when the information is transmitted over the internet or stored on remote servers.

4.2.2. Synthetic Data Training

In the process of model development and training, AI-generated negotiation transcripts, speech recordings, and other synthetic data are used instead of real student recordings. Synthetic data is generated by AI algorithms based on real

business English corpora and cross-cultural communication cases, which has the same characteristics and rules as real data but does not involve any real student information. This not only effectively preserves student privacy but also allows the research team to have more control over the data used to improve the system's performance.

5. Evaluation Metrics

5.1. Quantitative Assessment

Quantitative assessment metrics are designed based on the four dimensions of the "4D Evaluation Model, with clear calculation standards and strong operability, mainly including the following indicators:

5.1.1. Term Application Score (TAS)

This metric is used to evaluate the accuracy and appropriateness of students' use of professional terms in business scenarios, belonging to the Dimension of Professional Application Capability. It is a weighted calculation that assesses the accurate and appropriate use of terminology in professional contexts. The specific calculation formula is: TAS = (Correct_Terms / Total_Terms) × Context_Relevance_Factor. Among them, Correct_Terms refers to the number of professional terms used correctly by students, Total_Terms refers to the total number of professional terms used by students, and Context_Relevance_Factor refers to the relevance coefficient of professional terms to the business scenario (ranging from 0.8 to 1.2, with higher relevance corresponding to a higher coefficient). The TAS score ranges from 0 to 100, and a higher score indicates a higher level of professional term application ability. The data required for this metric is collected and calculated by the Professional Term Recognition Engine of the EconEvalAI platform.

5.1.2. Strategic Flexibility Index (SFI)

This metric is used to evaluate students' ability to adapt and pivot their negotiation tactics in response to changing circumstances, belonging to the Dimension of Professional Application Capability. It is a measure of an individual's ability to adjust negotiation strategies according to the opponent's performance and changes in the negotiation situation. This is achieved through the use of AI analysis to detect subtle shifts in speech patterns, sentiment tendencies, and expression logic that indicate a change in strategy. The SFI score ranges from 0 to 10, and the specific calculation is based on the frequency of strategy adjustment, the rationality of adjustment, and the effect of adjustment.

5.2. Qualitative Validation

Qualitative validation is used to supplement quantitative assessment, comprehensively understand the actual effect of the evaluation system and students' comprehensive ability, and avoid the one-sidedness of quantitative evaluation. Qualitative validation mainly includes enterprise partner reviews, student feedback, and teacher evaluations, which are collected through interviews, questionnaires, and on-site observations. The specific content is as follows:

5.2.1. Enterprise Partner Reviews

Invite cooperative business enterprises (such as international trade companies, financial institutions, and multinational corporations) to evaluate the English application ability and professional literacy of students who have participated in the AI-empowered evaluation. Enterprise partners evaluate students' performance in real cross-border business projects, including their professional term application ability, cross-cultural communication ability, and problem-solving ability. A typical evaluation comment from Alibaba Global Talent Program is: "The students who were evaluated using EconEvalAI demonstrated a 41% higher adaptability in our cross-border M&A projects compared to traditionally assessed peers. They were able to quickly grasp the nuances of different market conditions and regulatory environments, and their analytical skills and problem-solving abilities were highly valued."

5.2.2. Student Feedback

Collect feedback from students who have participated in the AI-empowered evaluation through questionnaires and focus group interviews, understanding their learning experience, views on the evaluation system, and the help of feedback on their learning improvement. A typical feedback from a 2026 pilot participant is: "The AI feedback showed me exactly when I misused 'force majeure' in contract negotiations, which my professor never caught. It helped me understand the nuances of legal language and improve my drafting skills."

6. Faculty Development Plan

6.1. Training Programs

Teachers are the core implementers of the AI-empowered teaching evaluation reform. To ensure the smooth implementation of the reform and improve the effect of the evaluation system, it is necessary to strengthen the training and support for teachers, improve their AI literacy and ability to use the AI evaluation system. Based on the actual needs of teachers.

6.2. Support Infrastructure

To provide continuous support for teachers in the process of using the AI evaluation system, a complete support infrastructure is established, including an Evaluation Technology Helpdesk and a shared repository of AI-annotated resources:

6.2.1. Evaluation Technology Helpdesk

The helpdesk is staffed by professional AI engineers and instructional designers, providing 24/7 online and offline support for teachers. It is responsible for solving the technical problems encountered by teachers in the use of the EconEvalAI platform, answering their questions about AI evaluation principles and operation methods, and collecting their suggestions for platform improvement.

6.2.2. Shared Repository of AI-Annotated Resources

Create a shared repository of AI-annotated business negotiation videos, professional documents, and cross-cultural communication cases for teacher calibration and exchange. Teachers can access these resources to better understand the AI evaluation standards, improve their ability to interpret AI evaluation results, and share their teaching experience and best practices with each other.

7. Challenges and Solutions

During the implementation of the AI-empowered teaching evaluation reform, it is inevitable to encounter various challenges. To ensure the smooth progress of the reform, this proposal puts forward targeted solutions for potential challenges (**Table 3**):

Challenge	Solution
Algorithm bias in cultural scoring	Regularly audit training data with diverse cultural consultants (including experts from different countries and regions) to ensure the algorithm is not biased against any particular cultural group or context. Update the training data regularly to cover more diverse cross-cultural scenarios.
Student resistance to AI evaluation	Implement a "human-in-the-loop" appeals process for controversial scores. This would allow students to appeal their AI-generated scores to a human reviewer (professional teachers or industry experts) if they believe the score is inaccurate or unfair. At the same time, conduct publicity and guidance for students to help them understand the advantages and scientificity of AI evaluation.
High implementation costs	Phase deployment starting with high-enrollment courses (e.g., Intl Business English). This would allow the college to test and refine the system before full-scale implementation across all courses, reducing the initial investment risk. In addition, seek cooperation with enterprises and research institutions to share the research and development costs.

This reform proposal establishes a data-driven, culturally-aware, and real-time evaluation system that transforms business English assessment from a summative judgment tool into a continuous improvement engine. It is closely aligned with the demands of globalized digital economies and the teaching reform of "Integration of Basic and Professional English" in business schools, and is expected to provide strong support for the cultivation of high-quality compound business talents.

Funding

This paper is a partial research achievement of the Education and Teaching Reform Project of Jiangsu Maritime Institute, titled Reform and Practice of Higher Vocational English Teaching for the Shipping Economics and Management Professional Cluster under the Background of Basic-Professional Integration.

Disclosure statement

The author declares no conflict of interest.

References

[1] Chen Y, Li J, 2023, AI-empowered formative assessment in business English teaching: A case study. Journal of English for Academic Purposes, 58: 101289.

[2] Baker R S, Inventado P S, 2022, Artificial intelligence in education. Annual Review of Psychology, 73: 671-699.

[3] Liu H, Wang L, 2021, Integration of basic and professional English: A curriculum reform perspective in business schools. English Language Teaching, 14(11): 102-115.

[4] Mitrović A, Suraweera P, 2022, Intelligent tutoring systems for language learning: A systematic review. Computers & Education, 188: 104428.

[5] Hofstede G, 2020, Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations (4th ed.). SAGE Publications.

[6] Zhang Q, Chen Z, 2024, Privacy-preserving AI in educational evaluation: Techniques and applications. Journal of Educational Technology & Society, 27(2): 156-168.

[1] Liu H, Wang L, 2021, Integration of basic and professional English: A curriculum reform perspective in business schools. English Language Teaching, 14(11): 102-115.

[2] Chen Y, Li J, 2023, AI-empowered formative assessment in business English teaching: A case study. Journal of English for Academic Purposes, 58: 101289.

[3] Baker R S, Inventado P S, 2022, Artificial intelligence in education. Annual Review of Psychology, 73: 671-699.

[4] Mitrović A, Suraweera P, 2022, Intelligent tutoring systems for language learning: A systematic review. Computers & Education, 188: 104428.

[5] Zhang Q, Chen Z, 2024, Privacy-preserving AI in educational evaluation: Techniques and applications. Journal of Educational Technology & Society, 27(2): 156-168.

[6] Hofstede G, 2020, Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations (4th ed.). SAGE Publications.