

Developing An AI-Enhanced Authentic Oral Assessment Model (AI-AOAM) For Evaluating EFL Speaking Performance among ESP Hotel Management Majors at a Public University in China: A Proposal

Liao Liling

¹Faculty of Tourism and E-Commerce, Baise University, China

²Universiti Malaysia Sabah (UMS), Malaysia

Asmaa AlSaqqaf (Corresponding Author)

Centre for the Promotion of Knowledge and Language Learning (PPIB)

Universiti Malaysia Sabah (UMS), Malaysia

Received: November 12, 2025 Accepted: November 27, 2025 Published: November 30, 2025

doi:10.5296/jsel.v13i1.23378 URL: <https://doi.org/10.5296/jsel.v13i1.23378>

Abstract

This paper proposes a conceptual framework for developing an AI-Enhanced Authentic Oral Assessment Model (AI-AOAM) to assess English oral performance among hotel management students enrolled in English for Specific Purposes (ESP) courses at a public university in China. Traditional oral assessments in Chinese higher education are often summative, teacher-centred, and disconnected from authentic communicative contexts, particularly those relevant to professional fields such as hospitality. Such limitations raise concerns about the validity, reliability, and pedagogical value of current practices in assessing EFL speaking performance. Grounded in a pragmatic research paradigm, this study employs a mixed-method sequential explanatory design. The qualitative phase involves structured interviews with ESP instructors to explore current assessment methods and challenges, while the quantitative phase collects data through authentic speaking proficiency tests administered to students. The AI-AOAM is constructed around three key assessment dimensions: accuracy, coherence, and appropriateness of spoken English, each reflecting both theoretical foundations and industry-specific communication demands. To ensure validity, the model incorporates expert feedback from both the language education sector and the hospitality industry and undergoes correlational analysis

with standardised English test scores. The proposed model aims to offer a more practical, context-sensitive, and occupation-oriented alternative to traditional assessment practices. Findings from this study are expected to provide valuable insights for language educators, curriculum developers, and policymakers, contributing to more effective and authentic oral assessment practices in ESP programs and improving the overall quality of EFL instruction in China's tertiary education system.

Keywords: AI, English for Specific Purposes (ESP), EFL Speaking Performance, Hotel Management Students and Oral Authentic Assessment

1. Introduction

In the field of English for Specific Purposes (ESP), the assessment of oral proficiency has become increasingly central as language instruction shifts from general language learning to communicative, occupation-specific approaches that reflect the real demands of professional environments (Nhi Nguyen & AlSaqqaf, 2023; Douglas, 2000; & Basturkmen, 2010). For students in disciplines such as hotel management, oral communication is not merely a linguistic skill but a professional competency that directly affects service quality, guest satisfaction, and employability. Therefore, effective assessment must go beyond linguistic accuracy to evaluate how well learners perform communication tasks that are integral to their future workplace roles.

Despite this growing recognition, in many Chinese higher education contexts, oral English assessment still tends to follow summative, decontextualized, and teacher-centred formats (AlSaqqaf et al., 2024; Cheng, 2008; Jin & Yang, 2021). Common practices include memorized dialogues, classroom recitations, or end-of-term oral exams that emphasize grammatical correctness over authentic communication. Although these approaches yield measurable outcomes, they often fail to capture students' ability to use English meaningfully in unpredictable, real-world hospitality situations. Consequently, such methods undermine both the construct validity of oral assessment and its pedagogical value (Luoma, 2004), leaving graduates underprepared for the communicative demands of the workplace.

Scholars have long emphasized the need for authentic assessment models that integrate professional contexts and foster the development of practical communicative competence (Fulcher, 2010; Norris et al., 2002). In the hospitality industry, this includes tasks such as handling guest check-ins, providing concierge recommendations, resolving complaints, and managing intercultural interactions, all of which require not only fluency but also pragmatic appropriateness, cultural awareness, and problem-solving skills (Lockwood, 2015). An assessment system that fails to replicate these workplace interactions risks widening the gap between classroom learning and professional practice.

This study addresses that gap by proposing an AI-Enhanced Authentic Oral Assessment Model (AI-AOAM) for evaluating EFL oral performance among Chinese hotel management students in ESP programs. The model is developed within a pragmatic paradigm and employs a mixed-methods sequential explanatory design (Creswell & Plano Clark, 2017), integrating qualitative insights from teacher interviews with quantitative performance data from authentic oral tasks.

By incorporating artificial intelligence tools, specifically the FiF speech-analysis app and automated feedback mechanisms, the model aims to enhance reliability, standardization, and efficiency in oral assessment (Choi & Chung, 2021; Xi, 2021).

Grounded in both linguistic theory (Weir, 2005) and hospitality-specific communication needs (Lockwood, 2015), the AI-AOAM emphasizes three key dimensions of oral performance: accuracy, coherence, and appropriateness. In doing so, it aims to create an assessment system that is not only authentic and fair but also professionally aligned and pedagogically valuable for Chinese tertiary ESP contexts. By bridging the divide between theory and practice, as well as between classroom learning and industry expectations, this research seeks to contribute to more effective oral English assessment practices in higher education.

2. Literature Review

Authentic assessment has increasingly been recognized as a paradigm shift in both language testing and pedagogy. At its core, it emphasizes evaluating a learner's ability to apply language in meaningful, real-world, or professionally relevant contexts rather than in artificial or isolated testing situations (Brown & Abeywickrama, 2019; O'Malley & Pierce, 1996). Traditional assessments—often grammar-based written exams, discrete-point tests, or scripted oral tasks—tend to measure rote knowledge and linguistic accuracy in decontextualized ways. By contrast, authentic assessment privileges performance-based tasks that reflect communicative demands outside the classroom, requiring learners to mobilize linguistic, pragmatic, and strategic resources simultaneously. In this sense, authentic assessment is not only a measure of language knowledge but also a measure of communicative competence, adaptability, and problem-solving, qualities that are vital in professional and intercultural environments.

The importance of authentic assessment is particularly pronounced in English for Specific Purposes (ESP), where learners' ultimate success depends on their ability to perform domain-specific communicative tasks. Nowhere is this more evident than in the hospitality and hotel management sector. Oral communication forms the cornerstone of guest service: staff are expected to greet and check in guests, explain services, handle bookings, provide recommendations, and resolve complaints or emergencies. Each of these interactions requires not only fluency and accuracy but also cultural sensitivity, service-oriented politeness, and the ability to manage unpredictable situations (Douglas, 2000; Lockwood, 2015). In such a dynamic industry, oral communication is not an “add-on” skill but the very medium through which service quality and guest satisfaction are delivered.

However, assessment practices in many EFL contexts, including China, remain misaligned with these professional realities. Oral evaluations often rely on memorized dialogues, scripted role-plays, or discrete grammar-based drills. While such formats may capture certain linguistic features, they rarely assess learners' ability to negotiate meaning, adapt speech to unexpected scenarios, or employ appropriate cultural and interpersonal strategies (Fulcher, 2010). As a result, students may achieve high scores in examinations but remain ill-prepared for the complexity of authentic workplace communication. Empirical evidence supports this concern. A recent industry survey revealed that 71.4% of hotel employers in China perceive graduates' oral English skills as inadequate, particularly in situations requiring improvisation, conflict

resolution, or cross-cultural awareness (Zhu, 2025). This disconnect underscores the urgent need for assessment models that extend beyond grammatical correctness to incorporate pragmatic, professional, and intercultural dimensions of communication. Scholars have argued that such models should evaluate not only accuracy and fluency but also lexical richness, interactional strategies, appropriateness, and service orientation within tasks that replicate real workplace encounters (Ma'fiah & Sumardiono, 2023).

Recent advances in artificial intelligence (AI) provide powerful tools to reimagine oral performance assessment in ways that better meet these demands. AI applications such as automated speech recognition (ASR), natural language processing (NLP), and machine learning-based scoring systems are now capable of capturing nuanced features of spoken language, pronunciation accuracy, fluency, coherence, and lexical diversity (Choi & Chung, 2021; Xi, 2021). Unlike human raters, AI systems offer consistency, immediacy, and scalability, reducing subjectivity and minimizing the resource burdens often associated with oral testing. When embedded into authentic, professional-specific tasks, AI technologies not only enhance reliability but also improve construct validity by more closely mirroring the communicative practices learners must handle in real professional environments.

An illustrative example of this integration is the FiF platform, a mobile-based oral training and assessment tool that has gained wide adoption in Chinese higher education. FiF leverages ASR technology to deliver interactive speaking tasks that closely simulate everyday communication. Students engage in role-play exercises or situational dialogues, record responses, and receive instant diagnostic feedback on pronunciation, intonation, grammar, and fluency. The platform also generates detailed performance reports that allow teachers to monitor progress, identify areas of difficulty, and tailor instruction accordingly. Crucially, FiF's flexibility makes it well-suited for ESP contexts such as hospitality management. For example, instructors can design hotel-specific tasks, check-in procedures, concierge recommendations, complaint handling, or emergency communication that require learners to practice authentic service encounters. In addition, the platform allows integration of teacher feedback and peer evaluation alongside automated scoring, ensuring a balance between human judgment and AI-driven objectivity. This hybrid model strengthens the pedagogical value of assessment by combining the efficiency of automation with the nuance of human oversight.

Despite these promising developments, significant research gaps remain. The majority of studies on AI in language assessment have focused on general English proficiency or standardized exam contexts, such as TOEFL or IELTS, rather than ESP settings like hospitality. Furthermore, little is known about how AI-enhanced assessments operate in Chinese higher education, where systemic challenges, including large class sizes, exam-driven traditions, and uneven digital literacy, may shape their implementation and outcomes. Addressing these gaps is critical not only to improving graduates' communicative readiness but also to advancing scholarly debates on the integration of AI into ESP pedagogy and assessment.

In response, the present study proposes the development of an AI-Enhanced Authentic Oral Assessment Model (AI-AOAM) designed specifically for hotel management students in China. By merging authentic assessment principles with AI-enabled platforms such as FiF, the model

seeks to establish a reliable, valid, and occupation-oriented framework for evaluating oral performance. It aims to bridge the gap between academic instruction and the communicative realities of the hospitality industry, while also contributing to broader understandings of AI's potential to transform language education and assessment.

3. Research Methodology

A mixed-methods exploratory sequential design, combined with Type 2 Design and Development Research (DDR), supports both product creation and iterative refinement (Rita, 2014). Qualitative interviews with ESP instructors identify assessment gaps, informing the AI-AOAM prototype. Quantitative testing, using student performance data, standardized scores, and FiF analytics, validates and adjusts the model. Triangulating teacher insights, student feedback, and empirical results strengthens validity (Denzin, 1978).

Mixed methods research is grounded in the pragmatic paradigm, which emphasizes the practical application of research findings and the value of using multiple approaches to address complex educational and social problems (Creswell & Plano Clark, 2017). Pragmatism is particularly appropriate for language assessment research because it encourages the combination of quantitative and qualitative methods to capture both measurable outcomes and contextual insights. By doing so, researchers can generate a more complete and nuanced understanding of the research problem. As Onwuegbuzie and Leech (2005) argue, many pragmatic studies are anchored in the integration of existing traditions from qualitative, quantitative, and mixed methods designs, reflecting the adaptability and inclusiveness of this paradigm.

In this framework, quantitative research is typically used to measure and analyze numerical data in order to identify patterns, trends, and statistical relationships. Within language assessment, this often involves standardized tests, scoring rubrics, and statistical analyses that evaluate learner performance objectively (Dörnyei, 2007). In contrast, qualitative research seeks to capture participants' perspectives, experiences, and behaviors through non-numerical forms of data such as interviews, observations, and document analysis. Such an approach is particularly valuable in language assessment, where contextual and affective factors, such as learner motivation, classroom dynamics, and perceptions of fairness, play a central role (Richards, 2003).

The strength of mixed methods lies in its ability to integrate these two complementary perspectives. Quantitative data often addresses the “what”, such as whether students' oral performance scores improve, while qualitative data helps to explain the “why”, shedding light on learner and teacher perceptions, challenges, and contextual influences (Leech, 2010). By bringing these strands together, mixed methods research provides a holistic account of complex language assessment phenomena that would remain incomplete if studied from only one perspective.

For this study, a mixed-methods design was selected to effectively address research questions concerning the development, implementation, and evaluation of the AI-Enhanced Authentic Oral Assessment Model (AI-AOAM). Mixed methods have a strong track record in language

assessment research, having been widely applied to evaluate the effectiveness, validity, and reliability of tools such as oral proficiency tests and technology-assisted assessments (Chapelle & Voss, 2016). Typically, such studies integrate quantitative data (e.g., test scores, rubric-based evaluations) with qualitative evidence (e.g., learner feedback, teacher interviews) to examine not only the technical accuracy of an assessment tool but also its pedagogical relevance and user perceptions.

Furthermore, mixed methods are particularly suited for investigating teachers' and learners' perceptions of assessment practices. Quantitative surveys can highlight overall patterns and trends in how assessments are implemented, while qualitative interviews can probe deeper into the beliefs, experiences, and challenges that underlie those patterns (Brown & Abeywickrama, 2019). This combination is especially important when developing context-specific assessment models such as AI-AOAM, where both measurable outcomes and contextual realities must be considered (Norris et al., 1998).

The advantages of a mixed-methods approach are widely recognized in educational research. It allows researchers to overcome the limitations of relying solely on quantitative or qualitative approaches, producing richer and more comprehensive data that strengthens the validity of the findings (Toomela, 2008; Almeida, 2018). Importantly, the flexibility of this approach ensures that research design can be adapted to the specific needs of the study context, while maintaining both rigor and relevance (Creswell & Plano Clark, 2017). In the case of this project, mixed methods not only support the empirical evaluation of AOAM's effectiveness but also provide critical insights into how teachers and students experience its implementation in real classroom settings.

4. Research Setting

The study will take place at a provincial university in southern China, within the School of Tourism and Hotel Management's four-year program. Participants include four ESP instructors with postgraduate qualifications and 84 second-year Hotel Management majors (CEFR A2–B2). All had at least one year of general English. The FiF platform, offering speech recognition and AI feedback, is integrated into assessment tasks simulating authentic hospitality scenarios, combining role-plays and AI-analyzed submissions.

The present study was conducted at Baise University, a public institution in China, with particular emphasis on the context of language learning and assessment within its language education programs. The research specifically targets the hotel management department, where students study English as a Foreign Language (EFL) in parallel with their hospitality and tourism-related coursework. Within this department, English is taught not merely as a general subject but as a professional and applied skill integrated into the training of future hotel professionals.

The study focuses on the classroom and instructional environments where EFL teaching and oral performance assessment are carried out. Consideration is given to both the cultural and linguistic context of the Chinese higher education system and the specific communicative demands faced by students preparing for hospitality careers. Recognizing the institutional

framework, the study also considers relevant university policies, curricular guidelines, and administrative structures that shape the design and implementation of language education. Instruction is further informed by established standards for assessing language proficiency, which guide the integration of ESP-oriented English within professional training.

The geographical site of the study is Baise City, located in the Guangxi Zhuang Autonomous Region of China. Baise is well known for its role in the red tourism industry, which connects historical significance with modern cultural and economic development. The choice of this setting is significant because the city's socio-economic profile underscores the importance of the tourism and hospitality sectors, both of which rely heavily on effective communication in English.

The institutional site is Baise University, the only comprehensive public university in the city. It comprises 21 secondary colleges and offers 57 undergraduate majors. For this study, Baise University was selected as the sole participant institution during both the implementation and evaluation phases because it is representative of the target population of hotel management students. According to Cohen (1992), the representativeness of a sample is critical in determining its effectiveness for research purposes, and Baise University offers a strong alignment with the needs of this study.

Within Baise University, the School of Tourism Management was selected as the focal unit. This school offers two undergraduate majors, tourism management and hotel management, and has an enrolment of approximately 1,217 full-time students. The decision to focus on this school was based on the significant role of English for Specific Purposes (ESP) within its curriculum. Professional English courses are compulsory components of both majors, reflecting the institution's recognition of the importance of ESP in preparing students for professional roles in the tourism and hospitality industries.

5. Conclusion

The development of the AI-Enhanced Authentic Oral Assessment Model (AI-AOAM) is intended to offer a more relevant and practical alternative to traditional ESP oral assessments. Unlike conventional approaches that often rely on scripted dialogues and isolated grammar points, this model incorporates real hospitality scenarios, such as check-in conversations, concierge recommendations, and complaint handling, ensuring that students are assessed in ways that accurately reflect the demands of their future workplace. By integrating AI tools such as automated speech recognition and natural language processing, the model also provides more objective scoring, timely feedback, and detailed insights into learners' strengths and weaknesses. Collectively, these features enhance the validity, reliability, and pedagogical value of oral assessment, making it not only more effective but also more engaging for students.

The design of the AI-AOAM is grounded in a mixed-methods Design and Development Research (DDR) process, which ensures that the model integrates robust theoretical foundations with practical classroom applicability. This methodological approach also enhances the model's adaptability, enabling its application beyond hotel management to other professional fields, such as nursing, aviation, or business, where authentic oral communication

is equally critical.

The impact of this model extends across multiple dimensions of higher education. In curriculum design, it offers a mechanism to align assessment with both learning objectives and industry expectations, thereby strengthening the connection between academic instruction and professional practice. For educators, it underscores the importance of professional development in designing and implementing technology-supported assessments. At the institutional and policy levels, it advocates for a shift toward more authentic, performance-based assessment practices that align with global trends and the increasing emphasis on graduate employability.

Ultimately, the AI-AOAM aims to catalyze a positive transformation in the assessment of oral English within ESP contexts, particularly in the Chinese higher education system. By enhancing the authenticity, professional relevance, and technological integration of assessments, the model supports students in developing not only test-taking abilities but also the genuine communicative competence required for success in the hospitality industry and other service-oriented professions.

References

- Almeida, F. (2018). Strengths and limitations of qualitative and quantitative research methods. *European Journal of Education Studies*, 5(9), 102-112. <https://doi.org/10.5281/zenodo.887089>
- AlSaqqa, A., Yang, Y., & Hu, K. (2024). The Effectiveness of the Incentive Autonomous Learning Strategies (IALS) Module to Improve Chinese Non-English Major Undergraduates' EFL Speaking Proficiency. *Journal of Language Teaching and Research*, 15(3), 843-852.
- Basturkmen, H. (2010). *Developing courses in English for specific purposes*. Palgrave Macmillan.
- Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (2nd ed.). Pearson Education.
- Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education* (pp. 349-364). Springer. <https://doi.org/10.1002/9781118411360.wbcla071>
- Choi, I., & Chung, Y. (2021). *Artificial Intelligence in Language Assessment: A Review of Emerging Trends*. *Language Assessment Quarterly*, 18(2), 121-141.
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
- Chapelle, C. A., & Voss, E. (2016). Evaluation of learning technology in language education. In Farr, F., & Murray, L. (Eds.), *The Routledge Handbook of Language Learning and Technology*. Routledge.
- Cohen, J. (1992). Statistical power analysis. *Current directions in psychological science*, 1(3), 98-101.

- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE. <https://doi.org/10.1080/17439760.2016.1262619>
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge University Press.
- Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). McGraw-Hill.
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge University Press.
- Dornyei, Z. (2007). *Research methods in applied linguistics*. Oxford university press.
- Fulcher, G. (2010). *Practical Language Testing*. Hodder Education.
- Jin, Y., & Yang, H. (2021). *Revisiting speaking assessment in China's university English programs*. *Asian Englishes*, 23(3), 245–260.
- Jilani, S., & Yasmin, M. (2016). Analyzing the Effectiveness of Simulation in Promoting Learner Autonomy: Perceptions of ESP Hotel Management Students. *Psychology*, 7, 1154-1162. <https://doi.org/10.4236/psych.2016.78116>
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26. <https://doi.org/10.3102/0013189X033007014>
- Lee, Y. W., & Sawaki, Y. (2009). Cognitive diagnosis approaches to language assessment: An overview. *Language Assessment Quarterly*, 6(3), 172-189. <https://doi.org/10.1080/15434300902985108>
- Leech, N. L., Dellinger, A. B., Brannagan, K. B., & Tanaka, H. (2010). Evaluating mixed research studies: A mixed methods approach. *Journal of mixed methods research*, 4(1), 17-31. <https://doi.org/10.1177/1558689809345262>
- Lockwood, J. (2015). *Developing an English for specific purpose curriculum for Asian call centres: A language-in-use approach*. Springer. <https://doi.org/10.1016/j.esp.2011.05.002>
- Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.
- Ma'fiah, I., & Sumardiono, S. (2023). English Language Needs for Hotel Service and ESP Program: The Case of Hospitality Vocational Secondary School. *Journal of Innovation in Educational and Cultural Research*, 4(3), 429-437. <https://doi.org/10.46843/jiecr.v4i3.680>
- Nhi Nguyen, T. U., & AlSaqqaf, A. (2023). Impact of a DMGA Scaffolding-Based Module on Improving the EFL Speaking Skills among Vietnamese ESP Learners. *Arab World English Journal*, 14(4), 342-357. <https://dx.doi.org/10.24093/awej/vol14no4.21>
- Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (2002). *Designing second language performance assessments*. University of Hawai'i, Second Language Teaching & Curriculum Center.
- Onwuegbuzie, A. J., & Leech, N. L. (2005). Taking the “Q” out of research: Teaching research

methodology courses without the divide between quantitative and qualitative paradigms. *Quality & Quantity*, 39(3), 267–296. <https://doi.org/10.1007/s11135-004-1670-0>

Richards, K. (2003). *Qualitative Inquiry in TESOL*. Palgrave Macmillan.

Rita, C., Richey, J. D. K., & Nelson, W. A. (2014). Developmental research: studies of instructional design and development. *Handbook of Research for Educational Communications and Technology*. Springer. https://doi.org/10.1007/978-1-4614-3185-5_12

Toomela, A. (2008). Variables in psychology: A critique of quantitative psychology. *Integrative Psychological & Behavioral Science*, 42, 245–265. <https://doi.org/10.1007/s12124-008-9059-6>

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Palgrave Macmillan.

Xi, X. (2021). Automated Speaking Assessment: Lessons Learned and Future Directions. *Annual Review of Applied Linguistics*, 41, 125–144.

Zhu, Y. (2025). *An evaluation of the hotel English ESP module at a Chinese higher vocational college from the needs of learners, teachers and hotel staff* [Unpublished doctoral dissertation]

Acknowledgments

This paper is published under the Knowledge, Attitude and Practice (KAP) Research Group, Centre for the Promotion of Knowledge and Language Learning (PPIB), Universiti Malaysia Sabah (UMS), Malaysia.

Funding

Not Applicable.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Macrothink Institute.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.