

Volume 8, Issue 1, April 2026 p. 67-80

Istanbul / Türkiye

Article Information

Article Type: Research Article

This article has been checked for similarity.

This is an open access article under the CC-BY-NC licence

Article History:

Received

11/12/2025

Received in

revised form

06/03/2026

Accepted

30/04/2026



PERCEPTIONS OF IMPLEMENTING AI-BASED SPEAKING ASSESSMENT TOOLS IN LEGAL ENGLISH EDUCATION: A MIXED-METHODS STUDY AMONG LECTURERS AND STUDENTS AT HANOI LAW UNIVERSITY

Nguyen Hai anh¹

Thu Nguyen Thi Hong²

Abstract

This study investigated perceptions of AI-based speaking assessment tools in legal English education at Hanoi Law University (HLU) in Vietnam, addressing a gap in discipline-specific applications within non-native EFL contexts. Employing a concurrent mixed-methods design, quantitative surveys were administered to 105 second-year legal English students, while semi-structured interviews were conducted with 6 lecturers to explore views on AI's role in evaluating legal discourse. Findings revealed positive perceptions of AI's efficiency (mean = 4.23 on a 5-point scale) and skill improvement potential (mean = 3.98), with perceptions varying significantly by student performance levels (ANOVA: $F(2,102) = 5.24$, $p = 0.007$ for feedback immediacy). Lecturers expressed optimism for AI's immediacy and advocated hybrid approaches. However, results are limited by self-reported data, a single-institution sample, and potential response bias, restricting generalizability. Future research should incorporate objective measures and multi-site designs for broader validation.

Keywords: AI-based assessment, Automated speaking evaluation, English for specific purposes (ESP), Legal English, Higher education, Student perceptions.

¹ Lecturer, Hanoi Law University, Faculty of Legal Foreign Languages, Vietnam. tienganhphaplyhlu@gmail.com. ORCID ID: 0009-0004-9088-3464

² Lecturer, Hanoi Law University, Faculty of Legal Foreign Languages, Vietnam. thuvicnguyen@gmail.com. ORCID ID: 0000-0002-6421-7511

Introduction

The integration of Artificial Intelligence (AI) into language assessment has revolutionized traditional methods, particularly for speaking skills, by offering efficient, scalable, and standardized evaluation through technologies like Automatic Speech Recognition (ASR) and Natural Language Processing (NLP) (Lu & Hua, 2022 b). Globally, AI-driven tools, such as SpeechRater used in TOEFL and IELTS, have enhanced assessment reliability by reducing rater bias and resource constraints (Chapelle & Voss, 2021). However, their application in domain-specific contexts like English for Legal Purposes (ELP) remains underexplored, especially in Vietnam, where the demand for legal professionals with proficient English skills is growing due to global integration (Giang, Loc, & Tuyen, 2025).

In Vietnam's higher education, pedagogical and infrastructural factors influence AI integration in specialized language education (Giang, Loc, & Tuyen, 2025). This study addresses this gap by examining perceptions of AI-based speaking assessment among 105 legal English students and 06 lecturers at Hanoi Law University, using a mixed-methods design. It explores how legal English students and lecturers perceive AI-based speaking assessment tools and how student perceptions vary by academic performance levels. In short, this research aims to inform effective AI integration in Vietnam's legal English curricula, contributing to innovative assessment practices in specialized language education.

Literature Review

The integration of Artificial Intelligence (AI) into language assessment has transformed traditional evaluation methods, particularly for speaking skills. This study synthesizes debates on AI's benefits and applicability in English for specific purposes (ESP), with a focus on legal English, to situate the current study within the broader scholarship. AI-powered assessment tools, leveraging Automatic speech recognition (ASR) and Natural language processing (NLP), offer standardized scoring, immediate feedback, and scalability (Lu & Hua, 2022.a). Studies demonstrate their reliability in large-scale tests like TOEFL and IELTS, where tools like Speech-rater reduce rater bias and resource constraints (Chapelle & Voss, 2021; (Ockey & Neiriz, 2021). For instance, Zhang and Zechner (2020) highlight that automated systems achieve high correlation with human

scores for fluency and pronunciation, making them efficient for high-stakes assessments. These advancements position AI as a valuable tool in English as a Foreign Language (EFL) education, particularly in resource-limited contexts like Vietnam (Giang, Loc, & Tuyen, 2025).

In ESP contexts, AI's applicability is supported by its ability to address discipline-specific communicative demands. Tse and Chan (2023) found that business English learners valued AI's immediacy for professional tasks, a sentiment echoed in legal English education (Lim & Park, 2023). Legal English requires mastery of specialized terminology, legal reasoning, and persuasive discourse, which AI can support through immediate feedback (Bhatia, 2021; Shin & Park, 2024). For example, Lim and Park (2023) note that AI models can evaluate argumentative coherence, leading to valid assessments in high-stakes contexts. This gap is particularly pronounced in Vietnam, where AI integration in higher education is evolving, and empirical studies on legal English assessment are scarce (Giang, Loc, & Tuyen, 2025). Learner proficiency significantly influences perceptions of AI tools. Carless and Lam (2020) demonstrate that lower-proficiency learners find AI-generated feedback helpful, while higher-proficiency learners benefit from autonomous support. This variability aligns with the Technology Acceptance Model (TAM), which posits that perceived usefulness and ease of use drive technology adoption (Davis, 1989; Lai, 2017). In ESP settings, Kohnke and Moorhouse (2022) and Tse and Chan (2023) emphasize that trust in AI's feedback relevance enhances acceptance, particularly when aligned with discipline-specific needs. These insights inform the current study's exploration of how proficiency levels shape perceptions of AI-based assessment among legal English students.

AI's efficiency and objectivity are lauded in domain-specific skills. While studies like (Ockey & Neiriz, 2021) advocate AI's scalability, the collective body underscores the need for integrated analyses like the present one. In ESP, particularly legal English, the lack of domain-specific training data is addressed through customized models and transparent algorithms. This study addresses these gaps by examining perceptions of AI-based speaking assessment among legal English students and lecturers at Hanoi Law University, using TAM to frame user acceptance in a specialized educational context. Therefore, the study is structured around two research questions:

1. How do students perceive the use of AI in oral assessment within legal English courses, and how do these perceptions differ across performance levels?
2. What are legal English lecturers' perspectives on the opportunities of integrating AI into oral assessment practices?

Research Methodology

Research Design

This study employs a mixed-methods research design using a concurrent parallel approach (Creswell & Plano Clark, 2018) to explore the perceptions associated with AI-based speaking assessment tools. It involves both quantitative & qualitative components conducted simultaneously, allowing for a comprehensive triangulation of findings. It captures the multifaceted experiences of English lecturers and second-year students at the Faculty of Legal English, Hanoi Law University. By collecting and analysing both numerical data and narrative insights, the research aims to provide a rich understanding of the pedagogical feasibility, acceptability, and implications of integrating AI-supported tools in oral proficiency assessment.

Participants

The study involved 105 second-year students enrolled in 03 classes (K48) majoring in Legal English. These students were selected because they had just completed the 1st Basic Legal English course, a foundational module delivered in the first semester of their second year. The 3 credit-course spans 15 weeks, designed to build core competencies in legal communication in English, with a strong emphasis on oral presentation and speaking tasks. To stratify students by speaking proficiency, their midterm presentation scores from the course were used as the primary criterion. The midterm presentation is a structured oral performance evaluated by lecturers using a rubric tailored to assess both legal content delivery and spoken communication skills. Scores are assigned on a scale of 1 to 10, and students were categorized into three proficiency groups: high (8.0–10), medium (5.5–7.9), and low (below 5.5).

This stratification was justified on both pedagogical and methodological grounds. The midterm presentation is a targeted measure of students' oral

English ability within a legal context, covering not only general fluency and accuracy but also the ability to use specialized terminology, structure arguments, and respond to audience questions. As it is the most prominent and formal speaking task within the course, and it is assessed by qualified Legal English lecturers, it provides a valid, course-specific proficiency indicator suitable for research purposes. Moreover, it offers consistency across all 03 classes and reflects students' competence at a key point in their academic path.

Data Collection Instruments

Data were collected over a six-week period (weeks 10-15) to ensure students had completed midterm presentations and were familiar with AI-based speaking assessment tools introduced in formative activities. Two instruments were used: a student survey questionnaire and semi-structured lecturer interviews.

Student Survey Questionnaire

A structured, self-administered questionnaire was distributed to 105 students in weeks 10–12 to explore perceptions of AI-based speaking assessment tools based on their experiences with Looora AI (see Section 3.3.3 for tool specifications). The 15-item instrument, adapted from Lu and Hua (2022) and Wang and Harding (2022), comprised 03 sections: perceived benefits (5 items, e.g., feedback immediacy), perceived usefulness (5 items, e.g., skill improvement), and acceptance (5 items, e.g., willingness to use AI tools). Items were rated on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). To ensure accessibility, the questionnaire was translated into Vietnamese, with an English version for bilingual participants, and contextual terms were modified (e.g., “conversation practice” to “legal role-play”). A pilot with 15 K47 students confirmed clarity and cultural appropriateness, leading to minor wording revisions. The questionnaire achieved a Cronbach's alpha of 0.85, indicating good reliability. The full instrument is available in Appendix A or upon request, subject to journal space constraints.

Semi-structured Interviews with Lecturers

Semi-structured interviews were conducted with 06 lecturers teaching the 1st basic legal English course in weeks 13-14. Interview prompts referred specifically to the AI tools students had used (Looora AI) to ground

discussions in shared experiences. The interview guide, developed based on Kohnke and Moorhouse (2022) and aligned with the Technology acceptance model (Davis, 1989), included open-ended questions across three areas: perceptions of AI tools, integration opportunities, and pedagogical recommendations. Questions were piloted with one lecturer to ensure clarity and relevance, resulting in refined prompts (e.g., specifying legal English contexts). Interviews, conducted in Vietnamese with participants' consent, lasted 15–30 minutes and were audio-recorded for accuracy. The interview guide explored three main areas: (1) lecturers' perceptions of AI-based speaking assessment tools, (2) opportunities encountered when integrating such tools in legal English contexts, and (3) pedagogical recommendations for effective implementation.

AI-based Speaking Assessment Tools

To ensure clarity and replicability, this study specifies the AI-based speaking assessment tool with which participants engaged during the 1st Basic Legal English course. Prior to data collection, students and lecturers were introduced to Lora AI (Lora App Limited), an AI-powered English conversation tutor. This single-tool approach was deliberately chosen to focus the investigation on AI's capacity to support legal English-speaking development without dispersing participant experience across multiple platforms. Lora AI offers 04 features particularly relevant to legal English learning: (1) free conversation capability allowing students to discuss legal scenarios and case-related themes; (2) real-time pronunciation and grammar correction for mastering complex legal syntax (e.g., conditional clauses) and terminology (e.g., "jurisdiction," "tort"); (3) flexible learning topics tailored to individual proficiency levels; and (4) immediate feedback simulating genuine conversation, enabling low-anxiety practice of extended legal discourse such as client interviews or persuasive arguments. The tool was systematically introduced during weeks 8-9 through a 90-minute workshop and guided practice sessions. Lora was selected based on three criteria: accessibility via free versions, relevance to extended speaking assessment (rather than discrete pronunciation practice), and alignment with legal English's communicative demands for sustained, coherent discourse.

A key methodological consideration concerns the legal-specific nature of student discourse when using Lora. While the tool provides

sophisticated conversation practice, it was not trained on legal English corpora but on general English datasets. This domain mismatch likely influenced AI scoring accuracy: Loora's algorithms, optimized for everyday conversational patterns, may misinterpret formal legal register features (e.g., archaic terms like "hereby," complex passive constructions) as errors or fail to recognize discipline-appropriate expression. Conversely, the tool might overlook genuine legal discourse errors that deviate from general English norms. This limitation contextualizes students' lower perceptions of AI's ability to accurately reflect legal English skills ($M = 2.85$, Section 5.1), reinforcing the need for domain-specific AI training data as elaborated in future research directions.

Ethical Consideration

This study adhered to rigorous ethical standards to ensure participant welfare and research integrity. Ethical approval was obtained from the Institutional Review Board at Hanoi Law University, confirming compliance with national and institutional guidelines for human subject research. All participants, including 105 second-year legal English students and six lecturers, provided informed consent prior to participation. Consent forms, available in both Vietnamese and English, clearly outlined the study's purpose, procedures, potential risks, and participants' rights, including the right to withdraw at any time without consequences. To protect confidentiality and anonymity, all data were de-identified, with personal information replaced by unique codes and stored securely in password-protected databases accessible only to the researcher. Special attention was given to supporting potentially vulnerable participants, particularly low-performing students, who may experience anxiety or lack confidence in speaking assessments (Carless & Lam, 2020). To mitigate these risks, survey instructions were clear, concise, and available in Vietnamese to enhance accessibility. Participation was explicitly voluntary, and students were reassured that non-participation or withdrawal would not affect their course grades or academic standing, addressing potential power imbalances in the classroom context.

Additionally, during semi-structured interviews, lecturers were encouraged to reflect on their experiences without pressure to disclose sensitive information, fostering a supportive environment. These measures ensured equitable treatment and minimized potential distress, aligning

with ethical principles for educational research (Creswell & Plano Clark, 2018).

Data Analysis

Quantitative data analysis

Descriptive statistics, including means, standard deviations, and frequency distributions, were used to summarize students’ overall perceptions of AI-based speaking assessment tools. Inferential analysis was performed using one-way ANOVA to compare perception scores across three academic performance groups (low, medium, high), categorized based on final course grades. Post-hoc tests were conducted when significant differences were detected to identify specific group contrasts. All analyses were conducted using SPSS software.

Qualitative data analysis

Interview data were transcribed and analysed using thematic analysis (Braun & Clarke, 2006). An inductive coding approach was employed to identify emerging themes related to teachers’ perceptions, observed opportunities, and pedagogical recommendations. To ensure reliability, 02 independent coders cross-checked the coding results, achieving an intercoder agreement rate of 90%.

Results

Quantitative Results

Table 1. Demographics & academic performance of participants (N = 105)

Characteristic	Frequency	Percentage (%)
Gender		
- Male	28	26.7
- Female	77	73.3
Basic Legal English 1 Grade		
- Low (<6)	28	26.7
- Medium (6–7.9)	54	51.4
- High (≥8)	23	21.9

Table 1 presents the demographics and academic performance of the 105 participants. Among them, 28 students (26.7%) were classified as low performers (final grades below 6), 54 students (51.4%) as medium

performers (grades between 6 and 7.9), and 23 students (21.9%) as high performers (grades 8 and above). The sample included 28 male students (26.7%) and 77 female students (73.3%).

Table 2. Students' perceptions of AI-based speaking assessment tools

Survey Item	Mean	SD
AI assessment provides immediate feedback	4.23	0.72
AI assessment is easy to understand	3.65	0.89
AI tools help improve my speaking skills	3.98	0.81
AI assessment reflects legal English skills well	2.85	1.12
I feel confident using AI speaking assessment tools	3.45	0.93

Descriptive statistics were computed to summarize students' perceptions of AI-based speaking assessment tools (see Table 2). Overall, students agreed that AI assessment provided immediate feedback ($M = 4.23$, $SD = 0.72$) and helped improve their speaking skills ($M = 3.98$, $SD = 0.81$). However, perceptions regarding the fairness of AI scoring ($M = 3.12$, $SD = 1.05$) and its ability to accurately reflect legal English skills ($M = 2.85$, $SD = 1.12$) were comparatively lower. The relatively low perception score for 'AI assessment reflects legal English skills well' ($M = 2.85$, $SD = 1.12$) may be attributable to participants' recognition that Loora AI, while providing valuable conversation practice and immediate feedback, was not specifically designed to evaluate the discipline-specific discourse features central to legal English proficiency (see Section 3.3.3).

Table 3. Differences in perceptions by academic performance

Group Comparison	Mean Difference	p-value	Significance
Low vs Medium	-0.35	0.045	Significant ($p < .05$)
Low vs High	-0.85	0.001	Highly Significant ($p < .01$)
Medium vs High	-0.50	0.020	Significant ($p < .05$)

A one-way ANOVA was conducted to examine differences in perception scores across the three academic performance groups. The results revealed statistically significant differences in perceptions of AI assessment providing immediate feedback ($F(2,102) = 5.24$, $p = 0.007$) and AI assessment reflecting legal English skills ($F(2,102) = 3.76$, $p = 0.027$). Post-hoc comparisons using Tukey's HSD were subsequently performed to identify

which groups differed significantly (see Table 3). The analyses indicated that students in the low academic performance group had significantly lower perception scores regarding the immediacy and contextual relevance of AI assessment compared to those in the medium and high groups. Additionally, the high-performing students reported significantly more favourable perceptions than the medium group.

Qualitative results

Table 4. Major Themes from Semi-Structured Interviews with Legal English Lecturers (N = 6)

Theme	Description	Example Quote
Optimism for AI's Efficiency	Lecturers value AI's ability to reduce grading workload and provide immediate feedback, especially in large classes.	"AI tools not only reduce grading workload but also allow us to offer students more immediate, formative feedback."
Limitations in Legal Discourse	AI struggles to evaluate rhetorical and argumentative skills critical to legal English, such as persuasive coherence.	"AI can detect surface-level fluency, but it misses how a student constructs a persuasive argument."
Fairness and Transparency Concerns	Concerns about AI's handling of diverse accents and lack of scoring transparency, impacting student trust.	"Some students feel demoralized because their accent is interpreted as incorrect by the system."
Advocacy for Hybrid Model	Lecturers propose combining AI for routine tasks with human evaluation for complex skills to enhance assessment.	"AI can pre-screen responses for basic errors, freeing us to focus on rhetorical framing."

Note: Themes were derived from thematic analysis of interviews conducted in weeks 13–14 of the semester.

Thematic analysis of semi-structured interviews with six legal English lecturers revealed three key themes regarding AI-based speaking assessment tools (see Table 4). Lecturers expressed optimism about AI's efficiency and objectivity, noted its opportunities in evaluating legal discourse complexities, and advocated a hybrid model combining AI and human evaluation. These themes provide context for the quantitative findings, particularly the lower perceptions among low-performing students (Table 3). These insights underscore the need for transparent communication and supportive pedagogy, such as teacher training to interpret AI feedback, to bridge the attitudinal gap across performance levels (Kohnke & Moorhouse, 2022).

Discussion

This study's findings reveal a nuanced landscape of perceptions regarding AI-based speaking assessment tools among legal English students and lecturers at Hanoi Law University. The positive reception of AI's efficiency ($M=4.23$) and its role in improving speaking skills ($M=3.98$) aligns with prior research praising automated systems for their scalability and objectivity (Chapelle & Voss, 2021; (Ockey & Neiriz, 2021). The findings highlight the potential for tailored AI tools to address the rhetorical and argumentative demands of legal English.

The significant differences in perceptions across performance levels (Table 3) corroborate Carless and Lam (2020), who found that lower-proficiency learners perceive AI-generated feedback positively, while higher-proficiency learners benefit from autonomous support. Qualitative insights from lecturers further underscore AI's ability to assess persuasive elements, aligning with Shin and Park (2024), who advocate for domain-specific training data to enhance AI's applicability in ESP contexts.

To leverage these perceptions, a blended assessment approach is recommended, where AI tools handle routine tasks (e.g., pronunciation, fluency) and human evaluators focus on complex skills like legal reasoning (Wang & Harding, 2022). This hybrid model aligns with lecturers' views, who see AI as a complementary tool rather than a replacement, ensuring comprehensive evaluation of communicative competence. Curriculum developers should integrate AI tools strategically, embedding domain-specific tasks (e.g., legal role-plays) to enhance contextual relevance. For instance, AI could be programmed with legal English corpora to better

evaluate argumentative coherence, addressing student perceptions ($M=2.85$). Teacher training programs should prioritize equipping educators to interpret AI feedback, particularly for low-performing students who report lower confidence (Table 3). Workshops could focus on translating AI's technical outputs into actionable, student-centered guidance, fostering trust and engagement (Kohnke & Moorhouse, 2022). These steps would enhance AI's pedagogical integration, aligning with Vietnam's push for innovative language education (Giang, Loc, & Tuyen, 2025).

Conclusion

This study advances the understanding of AI-based speaking assessment in legal English education by highlighting its potential in a Vietnamese higher education context. It contributes to global EFL and ESP research by demonstrating the need for domain-specific AI tools that address the rhetorical and argumentative demands of specialized disciplines. The findings offer a model for other ESP fields, such as business or medical English, where nuanced communicative competencies are critical (Tse & Chan, 2023). By advocating a hybrid assessment approach, the study provides a framework for balancing AI's efficiency with human expertise, ensuring effective evaluation practices.

The findings have implications beyond Vietnam, informing global efforts to integrate technology into language curricula. Future research should explore multi-institutional and longitudinal designs to validate these findings and develop AI systems trained on legal English corpora, enhancing assessment accuracy (Shin & Park, 2024). Additionally, learner-centered training programs could further support effective technology adoption, fostering inclusive ESP education worldwide (Kohnke & Moorhouse, 2022).

References

- Bhatia, V. K. (2021). *Analysing genre: Language use in professional settings*. Routledge.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Carless, D., & Lam, R. (2020). Integrating assessment for learning, teaching, and student self-regulation. *Assessment & Evaluation in Higher Education*, 45(4), 542–556. <https://doi.org/10.1080/02602938.2019.1631320>
- Chapelle, C. A., & Voss, E. (2021). Validity argument for AI-based language assessment. *Language Testing*, 38(3), 293–312. <https://doi.org/10.1177/0265532220987923>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage Publications.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Kohnke, L., & Moorhouse, B. L. (2022). Artificial intelligence in language education: A systematic review of AI-driven tools for speaking and writing assessment. *Computer Assisted Language Learning*, 35(8), 1892–1915. <https://doi.org/10.1080/09588221.2021.1984312>
- Lai, J. W. M. (2017). Technology acceptance model for the use of AI in education: An updated review. *Educational Technology Research and Development*, 65(6), 1571–1591. <https://doi.org/10.1007/s11423-017-9542-6>
- Lim, S., & Park, H. (2023). Evaluating AI-based assessment tools for professional English: A case study of business and legal English. *Language Testing in Asia*, 13(1), Article 1. <https://doi.org/10.1186/s40468-023-00189-4>
- Lu, X., & Hua, X. (2022.b). Automated speaking assessment: A review of recent advances and challenges. *Computer Assisted Language Learning*, 35(7), 1500–1523. <https://doi.org/10.1080/09588221.2021.1902232>
- Lu, X., & Hua, Y. (2022.a). Students' perceptions of automated feedback in speaking tasks. *Language Testing in Asia*, 12(1), Article 1. <https://doi.org/10.1186/s40468-022-00151-w>

- Giang, N. T., Loc, N., & Tuyen, T. (2025). Using AI-Powered Tools for Enhancing Speaking Skills of EFL Learners. *North American Journal of Psychology*, 27(4).
- Ockey, G. J., & Neiriz, R. (2021). Evaluating technology-mediated second language oral communication assessment delivery models. *Assessment in Education: Principles, Policy & Practice*, 28(4), 350–368. <https://doi.org/10.1080/0969594X.2021.1976106>
- Shin, D., & Park, M. (2024). Assessing professional English competence: AI-based tools in business and legal contexts. *Language Testing in Asia*, 14(1), Article 1. <https://doi.org/10.1186/s40468-024-00203-3>
- Tse, A. C., & Chan, K. (2023). Perceptions of AI-based feedback in ESP classrooms: A case study of business English learners. *English for Specific Purposes*, 70, 125–138. <https://doi.org/10.1016/j.esp.2022.12.003>
- Wang, L., & Harding, L. (2022). Transparency in AI scoring: Teachers' and students' perspectives. *Language Assessment Quarterly*, 19(3), 312–330. <https://doi.org/10.1080/15434303.2022.2051135>
- Zhang, J., & Zechner, K. (2020). Automated speaking scoring systems: Advances and challenges. *ETS Research Report Series*, 2020(1), 1–23. <https://doi.org/10.1002/ets2.12312>