THE USE OF ARTIFICIAL INTELLIGENCE TOOLS IN ASSESSING STUDENTS' SPEAKING PROFICIENCY

Zayniyeva Gulchaman Ilhomjon qizi
Teacher, National university of Uzbekistan named after Mirzo Ulugbek
Department of foreign language and literature
gulchamanzayniyeva@gmail.com
+99897 002 53 71

Abstract

The growing integration of artificial intelligence (AI) into language education has created new opportunities for assessing learners' speaking proficiency. AI-powered speech recognition and automated speaking assessment tools can support the evaluation of pronunciation, fluency, grammatical accuracy, lexical use, and selected features of spoken interaction. This thesis examines the potential educational benefits and limitations of using AI tools to assess students' speaking proficiency, particularly in higher education. It argues that such tools may provide timely feedback, increase opportunities for low-stakes practice, and help teachers manage assessment workloads. However, their results should be interpreted cautiously because recognition accuracy and scoring may vary according to accent, speech rate, recording quality, task type, and the characteristics of the learner. The thesis therefore advocates a human-in-the-loop approach in which AI-generated feedback complements, rather than replaces, teacher judgement. It also highlights the need for transparent criteria, data protection, accessibility, and empirical validation before AI-based scores are used for high-stakes decisions.

Keywords

artificial intelligence, speaking proficiency, automated assessment, language learning, feedback, validity, reliability.

Introduction. Speaking proficiency is a multidimensional component of communicative competence. It involves not only pronunciation and grammatical accuracy but also fluency, vocabulary use, coherence, pragmatic appropriateness, and the ability to respond meaningfully in interaction. Assessing these dimensions can be demanding because teachers must apply consistent criteria while considering the communicative purpose of each task. Recent developments in artificial intelligence have introduced tools that can transcribe speech, analyse selected speech features, and generate immediate feedback. These developments make it important to examine what AI can reliably assess, where its limitations lie, and how it can be integrated responsibly into language education.

Aim and Objectives. The aim of this thesis is to examine the pedagogical potential and limitations of AI tools in assessing students' speaking proficiency. The objectives are to identify speaking features that AI tools may evaluate; consider the potential benefits of automated feedback; discuss issues of validity, reliability, fairness, and privacy; and outline principles for combining AI-supported assessment with teacher evaluation.

Methods. The thesis uses a conceptual and literature-informed approach. It draws on established principles of language assessment and research concerning automated speech processing and AI-supported feedback. The discussion is organised around assessment validity, scoring consistency, pedagogical usefulness, and responsible implementation. As no original experimental data are presented, the observations are framed as an analytical synthesis rather than as claims of measured effectiveness.

Discussion. AI-based tools can offer several practical advantages. Automated speech recognition can produce transcripts, while some applications provide indicators related to pronunciation, pauses, speech rate, or lexical and grammatical features. Immediate feedback may enable students to practise more frequently and identify areas

for improvement outside class. Teachers may also use AI-generated information as supplementary evidence when planning instruction or monitoring progress. These benefits are most plausible in formative assessment, where feedback is used to guide learning rather than to determine consequential outcomes.

Nevertheless, automated assessment does not necessarily capture the full construct of speaking proficiency. A system may identify measurable speech features without adequately evaluating meaning, interactional competence, pragmatic appropriateness, creativity, or the quality of an argument. Speech-recognition errors can also affect subsequent feedback. Performance may vary with accents, dialects, speaking styles, background noise, microphone quality, and the design of the speaking task. Consequently, a numerical score should not automatically be treated as an objective or complete measure of a learner's ability.

Validity and reliability should therefore be examined for the intended learner population and assessment purpose. Where feasible, AI-generated scores should be compared with ratings from trained human assessors using a shared rubric. Teachers should review unexpected or consequential results, explain the assessment criteria, and allow students to question or contextualise automated feedback. Institutions should also communicate how recordings and transcripts are stored, who can access them, and whether they are used for purposes beyond assessment. Clear consent procedures and appropriate data-protection safeguards are essential.

Conclusion. Artificial intelligence tools can support the assessment of students' speaking proficiency by increasing access to practice, providing rapid feedback, and offering supplementary information about selected aspects of spoken performance. Their usefulness, however, depends on the quality of the tool, the assessment criteria, the learner context, and the way results are interpreted. AI should therefore be used as an aid to professional judgement, not as an unquestioned substitute for teachers. A

balanced approach combining validated AI tools, transparent rubrics, human review, and data protection is more educationally defensible than relying on automated scores alone. Future empirical research should compare AI ratings with trained human ratings across different proficiency levels and language backgrounds, while examining whether AI feedback leads to measurable improvement over time.

References

1. Council of Europe. (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment—Companion volume. Council of Europe Publishing.

2. Chapelle, C. A., & Voss, E. (2016). 20 years of technology and language assessment in Language Learning & Technology. Language Learning & Technology, 20(2), 116–128.

3. Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading? Language Testing, 27(3), 291–300.

4. Zechner, K., Higgins, D., Xi, X., & Williamson, D. M. (2009). Automatic scoring of non-native spontaneous speech in tests of spoken English. Speech Communication, 51(10), 883–895.