Next-Generation Oracy Assessment: AI-Driven Solutions for Listening and Speaking

Oracy—encompassing the interrelated skills of listening and speaking—is central to both interpersonal communication and human-computer interaction. Traditionally, assessing oracy has posed significant challenges. These include the need for intensive human involvement, difficulties in achieving scoring reliability, and limited scalability. Despite various theoretical frameworks that support oracy development, existing assessments often fail to reflect the diverse linguistic realities of learners or provide timely feedback that supports language growth.

This presentation explores how generative artificial intelligence (GenAI) can be used to address these limitations and enable a new generation of listening and speaking assessments that are more scalable, efficient, and context-sensitive. Drawing on my forthcoming Routledge volume, *Assessing Listening in the Age of Generative Artificial Intelligence*, I introduce two AI-driven solutions currently under development.

The first is a “dual-AI-in-the-loop” model for automated listening test generation. This framework combines GenAI with human input to streamline task development, significantly increasing production speed while maintaining construct relevance and content control. The second innovation is a “GPT-powered conversational agent” that integrates large language models (LLMs), text-to-speech (TTS), and automatic speech processing (ASP) to assess speaking in real time. This system enables dynamic interaction and provides learners with immediate, personalized feedback on their spoken performance.

These AI-powered tools do more than enhance operational efficiency—they also support a shift in the underlying purpose and structure of language assessment. Building on arguments from my 2023 article in *Language Testing*, “The Vexing Problem of Validity and the Future of Language Assessment,” I propose that next-generation language assessments should move beyond the global, proficiency-focused models that dominate

the field. Instead, it should prioritize local relevance, contextual appropriateness, and learner empowerment.

This reorientation has implications for test developers and institutions: rather than serving solely as producers of standardized assessments, they can facilitate teacher- and learner-led creation of customized testing and learning materials through AI-enhanced platforms. In doing so, they will help realize a more flexible, scalable, and inclusive assessment ecosystem—one that reflects the communicative needs of learners and supports meaningful engagement with spoken language.