AI-Powered Assessment of English Speaking Skills

Marina Canapero*

Riassunto

L'avvento dell'intelligenza artificiale ha trasformato profondamente il panorama della valutazione linguistica, in particolare nell'accertamento delle competenze di produzione orale in lingua inglese. Il contributo esamina i fondamenti teorici, le applicazioni attuali e le implicazioni pedagogiche degli strumenti di valutazione orale basati sull'intelligenza artificiale, facendo riferimento ai più recenti studi nell'ambito della linguistica applicata, delle tecnologie educative e del machine learning. Pur evidenziando i vantaggi offerti da questi sistemi in termini di scalabilità, coerenza e tempestività del feedback, il lavoro sostiene che la loro efficace integrazione nella pratica didattica richieda un'attenta riflessione sulla validità, l'equità dei processi valutativi e sull'insostituibile ruolo del giudizio umano.
Parole chiave: Intelligenza artificiale, Valutazione linguistica, Produzione orale in lingua inglese, Tecnologie educative, Valutazione delle competenze.

Abstract

The advent of artificial intelligence has significantly transformed the landscape of language assessment, particularly in the evaluation of English speaking skills. This article examines the theoretical foundations, current applications, and pedagogical implications of AI-powered oral assessment tools, drawing on recent literature in applied linguistics, educational technology, and machine learning. It argues that, while AI-driven systems offer considerable advantages in terms of scalability, consistency, and immediacy of feedback, their effective

* English language assistant, University of Padova.

Excellence and Innovation in Learning and Teaching (ISSNe 2499-507X), 2026, 1 Special Issue

Doi: 10.3280/exioa1-2026oa23982

Copyright © FrancoAngeli
This work is released under Creative Commons Attribution - Non-Commercial – No Derivatives License. For terms and conditions of usage please see: http://creativecommons.org

integration into educational practice requires careful attention to issues of validity, fairness, and the irreplaceable role of human judgment.
Keywords: Artificial Intelligence, Language Assessment, English Speaking Assessment, Educational Technology, Oral Proficiency Assessment

Article submitted: 12/06/2026; accepted: 03/07/2026

The assessment of oral language proficiency has historically represented one of the most resource-intensive and methodologically complex challenges in language education. Traditional speaking assessments rely heavily on trained human raters, whose evaluations – though informed by established rubrics such as those underpinning the Common European Framework of Reference for Languages (CEFR) – are inevitably subject to inter-rater variability, cognitive fatigue, and implicit bias (Fulcher, 2003). The emergence of Automated Speech Recognition (ASR) and Natural Language Processing (NLP) technologies has introduced a new generation of assessment instruments capable of evaluating spoken English with increasing sophistication and reliability (Litman, 2016).

Theoretical Foundations

AI-powered speaking assessment systems operate primarily through the integration of ASR, prosodic analysis, and NLP algorithms. These tools evaluate a range of linguistic dimensions – including phonological accuracy, lexical diversity, syntactic complexity, discourse coherence and fluency markers such as speech rate and pause distribution – against large corpora of reference data (Zechner et al., 2009). Scoring models are typically trained on datasets annotated by expert human raters, enabling machine learning systems to approximate – and in some domains, replicate – human scoring behaviour with statistically significant accuracy (Yoon et al., 2012).

The theoretical framework underpinning these systems draws substantially from Bachman and Palmer's (1996) model of communicative language ability, which identifies grammatical, discourse, pragmatic, and strategic competences as core components of oral proficiency. Contemporary AI tools increasingly attempt to operationalise these constructs computationally, though pragmatic and interactional dimensions remain particularly resistant to automated measurement (Chapelle & Voss, 2016).

Copyright © FrancoAngeli
This work is released under Creative Commons Attribution - Non-Commercial –
No Derivatives License. For terms and conditions of usage please see: http://creativecommons.org

Current Applications

Several well-established platforms currently deploy AI for speaking assessment at scale. The Duolingo English Test employs machine learning models to evaluate spoken responses in real time, reporting strong correlations with TOEFL and IELTS scores (Cardwell et al., 2022). Pearson's Versant system, one of the earliest automated speaking tests, uses ASR to assess pronunciation, fluency, and sentence mastery across timed spoken tasks (Pearson, 2011). Similarly, ETS's SpeechRater℠ – integrated within the TOEFL iBT – provides automated scoring of integrated speaking tasks, demonstrating criterion validity against human rater scores (Zechner et al., 2009).

A particularly significant development in recent years has been the emergence of conversational AI platforms that move beyond passive assessment into fully interactive oral practice environments. These systems engage learners through realistic avatars and AI tutors capable of sustaining extended spoken dialogues, correcting grammatical inaccuracies in real time, providing targeted pronunciation feedback, and – crucially – identifying and adapting to a wide range of regional and national accents. This accent-sensitive functionality represents a meaningful advance toward more equitable and inclusive oral language assessment, addressing longstanding concerns regarding the privileging of standard Inner Circle English varieties in automated scoring systems (Subtirelu, 2017).

Busuu is among the most pedagogically sophisticated of these platforms, integrating AI-driven speaking exercises with community-based feedback from native speakers. Its oral practice modules provide immediate automated correction of both grammar and pronunciation, with adaptive pathways aligned to CEFR descriptors from A1 through to C2. Of particular institutional relevance, Busuu is currently being offered at scale to university staff across multiple higher education institutions, including as a complementary resource to face-to-face taught language classes at all CEFR levels. This blended deployment model – in which AI-mediated practice supplements rather than replaces formal instruction – represents an increasingly prevalent and evidence-informed approach to institutional language support (Banados, 2006). The platform's capacity to extend learning beyond scheduled classroom hours, while maintaining alignment with instructor-led curricula, positions it as a highly practical tool for busy academic and administrative professionals seeking to develop or consolidate English proficiency.

Speeko, primarily designed as a public speaking and communication coach, employs an AI tutor interface to guide users through structured oral exercises targeting fluency, pacing, clarity, and vocal confidence. The platform's

Copyright © FrancoAngeli
This work is released under Creative Commons Attribution - Non-Commercial –
No Derivatives License. For terms and conditions of usage please see: http://creativecommons.org

feedback engine analyses spoken output across multiple paralinguistic dimensions, offering granular suggestions that extend beyond purely linguistic accuracy to encompass the rhetorical and presentational competences increasingly demanded in professional and academic English contexts. For university staff and researchers required to present at international conferences or engage in English-medium instruction, Speeko's focus on communicative effectiveness represents a valuable complement to more grammar-centred tools.

Praktika deploys immersive, scenario-based conversations with photorealistic AI avatars, simulating authentic communicative situations such as job interviews, academic discussions, and professional meetings. Learners interact verbally with these avatars in real time, receiving post-interaction reports detailing grammatical accuracy, vocabulary range, pronunciation scores, and fluency metrics. The platform's accent recognition capabilities allow it to calibrate feedback sensitively across a broad spectrum of English varieties, reducing the risk of penalising non-native speakers for accent features that do not impede intelligibility (Jenkins, 2000). The psychological dimension of avatar-mediated interaction has also attracted scholarly attention, with preliminary evidence suggesting that the reduced evaluative threat of speaking to a virtual interlocutor may lower affective barriers to oral production, particularly among anxious language learners (Lan, 2020).

Within higher education, institutions including the University of Padova are actively investigating how AI-enhanced tools – including blended deployments of platforms such as Busuu – may support language centre services, particularly for large-scale proficiency screening, staff development, and self-directed learning pathways aligned to institutional language policies.

Pedagogical Implications and Limitations

The pedagogical value of AI-powered speaking assessment lies substantially in its capacity to deliver immediate, granular, and actionable feedback – a dimension of formative assessment shown to significantly enhance language acquisition outcomes (Hattie & Timperley, 2007). Students gain access to detailed diagnostic profiles without the scheduling constraints inherent in human-administered oral examinations.

Nevertheless, critical limitations must be acknowledged. AI systems trained predominantly on standard Inner Circle varieties of English (Kachru, 1985) risk penalising speakers of legitimate World Englishes, raising significant concerns regarding construct validity and assessment equity (Subtirelu, 2017). Furthermore, higher-order communicative competences – including

Copyright © FrancoAngeli
This work is released under Creative Commons Attribution - Non-Commercial –
No Derivatives License. For terms and conditions of usage please see: http://creativecommons.org

interactional fluency, rhetorical effectiveness, and pragmatic appropriateness – remain beyond the reliable reach of current automated systems (Chapelle & Voss, 2016).

Conclusions

AI-powered assessment of English speaking skills represents a compelling and rapidly maturing field with genuine transformative potential for language education. However, its responsible implementation demands a human-AI collaborative model, in which automated tools augment rather than supplant the expertise of qualified language educators. Future research should prioritise the development of more linguistically inclusive training datasets, transparent scoring algorithms, and robust frameworks for integrating AI assessment within CEFR-aligned curricula.

References

Bachman L. F., & Palmer A. S. (1996). *Language Testing in Practice*. Oxford University Press.

Banados E. (2006). A blended-learning pedagogical model for teaching and learning EFL successfully through an online interactive multimedia environment. *CALICO Journal*, 23(3): 533-550.

Busuu (2023). *Busuu for Business: Enterprise Language Learning Solutions*. -- Retrieved from https://www.busuu.com/en/business.

Cardwell A., LaFlair G., & Yoon S. Y. (2022). Validity evidence for the Duolingo English Test. *Duolingo Research Report DRR-22-01*.

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

Fulcher G. (2003). *Testing Second Language Speaking*. Pearson Longman.

Hattie J., & Timperley H. (2007). The power of feedback. *Review of Educational Research*, 77(1): 81-112.

Jenkins J. (2000). *The Phonology of English as an International Language*. Oxford University Press.

Kachru B. B. (1985). Standards, codification and sociolinguistic realism. In: R. Quirk & H. Widdowson (Eds.). *English in the World*. Cambridge University Press.

Lan Y. J. (2020). Immersion into virtual reality for language learning. *Language Learning & Technology*, 24(3): 1-15.

Litman D. (2016). Natural language processing for enhancing teaching and learning. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 4170-4176.

Copyright © FrancoAngeli
This work is released under Creative Commons Attribution - Non-Commercial –
No Derivatives License. For terms and conditions of usage please see: http://creativecommons.org

Pearson (2011). *The Versant English Test: Test Description and Validation Summary*. Pearson Knowledge Technologies.
Praktika (2023). *AI-Powered English Speaking Practice with Avatars*. -- Retrieved from https://praktika.ai.
Speeko (2023). *Speeko: Public Speaking Coach*. -- Retrieved from https://www.speeko.co.
Subtirelu N. C. (2017). Recognizing and addressing language bias in automated speaking assessment. *TESOL Quarterly*, 51(3): 749-757.
Yoon S. Y., Bhat S., & Zechner K. (2012). Vocabulary profile as a measure of vocabulary sophistication. *Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications*, 180-189.
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.

Copyright © FrancoAngeli
This work is released under Creative Commons Attribution - Non-Commercial –
No Derivatives License. For terms and conditions of usage please see: http://creativecommons.org