CENDEKIA: Jurnal Ilmu Pengetahuan
Vol. 6, No. 3, Mei-Juli 2026
e-ISSN : 2774-4183 | p-ISSN : 2774-8030
Online Journal System : https://jurnalp4i.com/index.php/cendekia

ENHANCING EFL ORAL ASSESSMENT: IMPLEMENTING GENERATIVE AI FOR POETRY RECITATION EVALUATION

Muhammad Arief Budiman[1], Davron Aslonqulovich Juraev[2], Nazira Mohubbat Mammadzada[3], Nisa Prasetyoningtyas[4], Nita Wahdatul Mustaghfiroh[5], Prisca Oktaviana Putri[6], Puteri Ayu Marshanda[7]
Universitas PGRI Semarang[1,4,5,6,7], Azerbaijan University of Architecture and Construction[2], State Oil Company of the Azerbaijan Republic[3]
e-mail: ariefbudiman@upgris.ac.id

Diterima: *26/04/2026*; Direvisi: *08/05/2026*; Diterbitkan: *13/05/2026*

ABSTRAK

Integrasi *Generative Artificial Intelligence* (GAI) dalam pendidikan bahasa menawarkan peluang inovatif untuk penilaian, khususnya dalam mengevaluasi keterampilan lisan dan ekspresif. Penelitian ini menyelidiki penerapan penilaian berbantuan GAI dalam kelas Bahasa Inggris sebagai bahasa asing (EFL) tingkat universitas, dengan fokus pada video pembacaan puisi mahasiswa. Sebanyak 69 mahasiswa berpartisipasi dengan mengumpulkan rekaman video pembacaan mereka, yang kemudian dievaluasi menggunakan sistem penilaian berbasis rubrik yang ditingkatkan dengan analisis AI. Hasil kuantitatif menunjukkan bahwa penilaian berbantuan AI konsisten dan sangat selaras dengan penilaian dosen, menunjukkan reliabilitas tinggi dalam menilai pelafalan, kelancaran, intonasi, dan ekspresivitas. Analisis kualitatif terhadap persepsi mahasiswa mengungkapkan bahwa umpan balik dari AI dipandang informatif, memotivasi, dan bermanfaat untuk meningkatkan performa lisan, meskipun beberapa mahasiswa mencatat keterbatasan dalam menangkap nuansa interpretatif. Temuan ini menunjukkan bahwa GAI dapat secara efektif mendukung penilaian lisan EFL, dengan memberikan umpan balik yang cepat dan andal serta melengkapi evaluasi tradisional oleh dosen. Implikasi terhadap praktik pedagogis, desain kurikulum, dan penelitian masa depan dalam pembelajaran bahasa berbantuan AI turut dibahas.

Kata Kunci: *Generative Artificial Intelligence, Penilaian EFL, Pembacaan Puisi, Kemahiran Lisan, Umpan Balik Berbantuan AI*

ABSTRACT

The integration of Generative Artificial Intelligence (GAI) in language education offers innovative opportunities for assessment, particularly in evaluating oral and expressive skills. This study investigates the application of GAI-supported assessment in a university-level English as a Foreign Language (EFL) class, focusing on students' poetry recitation videos. A total of 69 students participated by submitting video recordings of their recitations, which were evaluated using a rubric-based scoring system enhanced with AI analysis. Quantitative results indicated that AI-assisted assessments were consistent and closely aligned with instructor evaluations, demonstrating high reliability in scoring pronunciation, fluency, intonation, and expressiveness. Qualitative analysis of student perceptions revealed that AI feedback was perceived as informative, motivating, and useful for improving oral performance, though some students noted limitations in capturing interpretive nuances. The findings suggest that GAI can effectively support EFL oral assessment, providing timely and reliable feedback while

https://doi.org/10.51878/cendekia.v6i3.10634

CENDEKIA: Jurnal Ilmu Pengetahuan
Vol. 6, No. 3, Mei-Juli 2026
e-ISSN : 2774-4183 | p-ISSN : 2774-8030
Online Journal System : https://jurnalp4i.com/index.php/cendekia

complementing traditional instructor evaluation. Implications for pedagogical practice, curriculum design, and future research in AI-assisted language learning are discussed.
Keywords: *Generative Artificial Intelligence, EFL Assessment, Poetry Recitation, Oral Proficiency, AI-Supported Feedback*

INTRODUCTION

The rapid development of digital technologies has significantly transformed educational practices across the globe, particularly in language education. In recent years, Artificial Intelligence (AI) has emerged as one of the most influential technological innovations in teaching and learning (Black & Wiliam, 1998; Liu et al., 2025; Sadler, 2005). Among these advancements, Generative Artificial Intelligence (GAI) has attracted increasing attention due to its capacity to generate human-like responses, analyze linguistic data, and provide interactive feedback. In higher education, these capabilities offer new opportunities to enhance instructional practices and improve assessment systems. In English as a Foreign Language (EFL) classrooms, assessment plays a critical role in evaluating learners' language development and guiding pedagogical decisions (Budiman et al., 2020). However, traditional assessment methods for oral performance tasks often require substantial time and effort from instructors, particularly when evaluating multiple linguistic and expressive dimensions of speaking. As class sizes increase and educational institutions adopt more technology-mediated learning environments, there is a growing need for innovative assessment approaches that can enhance efficiency while maintaining reliability and pedagogical value.

The evaluation of oral language performance in EFL contexts is grounded in several well-established theoretical perspectives. Early frameworks of speaking assessment proposed by Harris (1969) and later expanded by Brown (2010) emphasize core components of oral proficiency, including pronunciation, fluency, comprehension, and grammatical accuracy. These elements are widely recognized as essential dimensions of communicative competence and remain fundamental in evaluating speaking performance in language classrooms. In addition, formative assessment theory highlights the importance of continuous feedback in supporting learners' progress and improving performance outcomes (Black & Wiliam, 1998). From a sociocultural learning perspective, knowledge development occurs through mediated interactions involving tools and collaborative processes (Vygotsky, 1980). In this context, digital technologies including AI systems can function as mediational tools that facilitate learning by providing timely feedback, scaffolding learners' performance, and supporting reflective learning practices. Therefore, integrating AI technologies into language assessment aligns with these theoretical perspectives by enhancing feedback mechanisms and supporting the development of learners' communicative competence.

Recent research has increasingly explored the role of Artificial Intelligence in education, particularly in supporting adaptive learning and automated assessment. Studies suggest that AI technologies can enhance instructional efficiency by providing immediate feedback, analyzing large volumes of student data, and supporting personalized learning pathways (Holmes et al., 2019; Holmes & Luckin, 2016). In EFL learning environments, AI applications have been widely used in areas such as automated essay scoring, grammar correction, vocabulary learning, and pronunciation training. Recent studies also demonstrate that AI-powered pronunciation tools can significantly improve learners' pronunciation accuracy and increase their motivation in EFL learning environments (Abdelhalim & Alsehibany, 2025). For example, AI-driven writing evaluation systems can identify

https://doi.org/10.51878/cendekia.v6i3.10634

CENDEKIA: Jurnal Ilmu Pengetahuan
Vol. 6, No. 3, Mei-Juli 2026
e-ISSN : 2774-4183 | p-ISSN : 2774-8030
Online Journal System : https://jurnalp4i.com/index.php/cendekia

grammatical errors and provide suggestions for improvement, enabling learners to refine their language use independently. Similarly, speech recognition technologies have been applied to analyze pronunciation accuracy and provide feedback on articulation and phonological patterns. Despite these advancements, the majority of AI-based tools remain focused on written language skills or basic speech analysis. Complex oral performance tasks that involve expressive interpretation and communicative delivery, such as poetry recitation, have received comparatively limited attention in the existing literature. Recent developments in AI-assisted language assessment further indicate that artificial intelligence can support more effective evaluation of speaking and listening performance through automated analysis and adaptive feedback mechanisms (Goh & Aryadoust, 2025).

The assessment of poetry recitation requires a multidimensional framework that integrates linguistic competence with expressive performance. Fluency theory provides an important conceptual basis for understanding oral reading performance. Rasinski (2010) conceptualizes fluency as a multidimensional construct that includes accuracy, automaticity, and prosodic features such as rhythm, intonation, and stress. Prosody plays a crucial role in shaping meaning and conveying the emotional tone of a text, particularly in literary works such as poetry. In poetry recitation, readers must interpret the text and communicate its meaning through expressive voice modulation and effective phrasing. Performance-based learning theories further emphasize that oral expression involves both linguistic competence and artistic interpretation (Spolin, 1999). Educational frameworks proposed by the National Council of Teachers of English (Senokossoff, 2016) also advocate holistic assessment approaches that capture linguistic accuracy, comprehension, and expressive delivery. Based on these theoretical perspectives, poetry recitation tasks can be evaluated using multiple criteria, including pronunciation, intonation, stress, fluency, expression, and comprehension.

A number of studies have investigated methods to improve the reliability and effectiveness of oral language assessment in EFL contexts.A growing body of research on digital phonetics in ELT also highlights the increasing role of AI technologies in analyzing pronunciation, stress, and prosodic features in language learning environments (Shaheed et al., 2026). Rubric-based assessment is widely recognized as an effective approach for evaluating complex language performances because it provides clear criteria and structured performance descriptors. According to Andrade (Andrade, 2005), Rubrics enhance transparency and fairness by clarifying expectations and guiding both instructors and learners during the evaluation process. In parallel, recent research has begun to explore the potential of AI technologies to support language assessment. AI-driven speech analysis systems have demonstrated the ability to analyze pronunciation accuracy, detect speech patterns, and evaluate prosodic features such as pitch, rhythm, and articulation (Kuddus, 2022; Zou et al., 2020). More recent studies also suggest that generative AI tools can assist educators in analyzing spoken language data and generating detailed feedback on students' performance (Kohnke et al., 2023). Previous studies have shown that automatic speech recognition technology can positively influence EFL learners' pronunciation development and speaking performance by providing immediate corrective feedback (Sun, 2023). Similarly, AI-driven instructional interventions have been found to improve learners' pronunciation skills and increase confidence in oral communication tasks (Xodabande et al., 2025). These developments indicate that AI technologies have the potential to complement human evaluators by providing objective analysis while enabling instructors to focus on higher-level interpretive aspects of language performance.

https://doi.org/10.51878/cendekia.v6i3.10634

CENDEKIA: Jurnal Ilmu Pengetahuan
Vol. 6, No. 3, Mei-Juli 2026
e-ISSN : 2774-4183 | p-ISSN : 2774-8030
Online Journal System : https://jurnalp4i.com/index.php/cendekia

Despite the increasing use of AI in language education, several significant gaps remain in the literature. First, most AI-assisted assessment systems have been developed for written tasks or limited speech analysis, leaving performative oral activities relatively unexplored. Tasks that involve expressive interpretation, such as poetry recitation, require the integration of linguistic accuracy, prosody, and artistic delivery, making them more complex to evaluate using automated systems. Second, existing studies often focus primarily on the technical accuracy and algorithmic performance of AI tools, while their pedagogical implications for language learning and skill development receive less attention. Consequently, there is still limited empirical evidence on how Generative Artificial Intelligence can be effectively integrated into rubric-based assessment frameworks to support the evaluation of expressive oral tasks in EFL contexts.

To address these limitations, the present study proposes an innovative approach that integrates Generative Artificial Intelligence with rubric-based assessment to evaluate EFL students' poetry recitation performances. Unlike previous studies that concentrate mainly on written language evaluation or isolated pronunciation analysis, this research examines how GAI can support the assessment of both linguistic and expressive aspects of oral performance. By applying an AI-assisted rubric to analyze students' poetry recitation videos, the study introduces a novel model of assessment that combines technological efficiency with theoretically grounded evaluation criteria. This approach contributes to the emerging field of AI-enhanced language assessment by demonstrating how generative AI tools can be used to support more holistic evaluation practices in language classrooms.

Based on the issues discussed above, this study aims to investigate the use of Generative Artificial Intelligence in supporting the assessment of EFL students' poetry recitation videos in a university classroom context. Specifically, the study seeks to examine how GAI can assist in evaluating key aspects of oral performance, including pronunciation, intonation, fluency, and expressive delivery. In addition, the research explores the reliability and consistency of AI-assisted assessment compared to traditional instructor-based evaluation. Finally, the study investigates students' perceptions of receiving AI-supported feedback on their oral performance. By addressing these objectives, the study aims to contribute to the development of innovative, efficient, and pedagogically meaningful assessment practices in technology-enhanced EFL learning environments.

METHODS

This study employed a descriptive-exploratory research design using a mixed-methods approach to investigate the implementation of Generative Artificial Intelligence (GAI) in supporting the assessment of English as a Foreign Language (EFL) students' poetry recitation performances. The quantitative approach was used to analyze students' oral performance scores obtained from AI-assisted assessment and human evaluation, while the qualitative approach was used to explore students' perceptions of AI-generated feedback in language assessment. The integration of quantitative and qualitative data enabled the study to provide a more comprehensive understanding of the effectiveness, reliability, and pedagogical implications of AI-supported assessment in EFL learning contexts. The study focused specifically on the use of AI technology to evaluate both linguistic and interpretive dimensions of oral performance. Through this approach, the research aimed to examine how GAI could contribute to more efficient and meaningful oral assessment practices in higher education.

https://doi.org/10.51878/cendekia.v6i3.10634

CENDEKIA: Jurnal Ilmu Pengetahuan
Vol. 6, No. 3, Mei-Juli 2026
e-ISSN : 2774-4183 | p-ISSN : 2774-8030
Online Journal System : https://jurnalp4i.com/index.php/cendekia

The participants of this study were sixty-nine undergraduate students enrolled in an English as a Foreign Language course at a university. The participants were selected using purposive sampling because all students were required to complete an oral performance assignment as part of their course assessment. The students had intermediate to upper-intermediate levels of English proficiency, which were identified through placement tests and previous academic records. The participants represented varying speaking abilities, allowing the study to capture differences in pronunciation accuracy, fluency, prosody, and expressive delivery during poetry recitation performances. Participation in the study was voluntary, and all participants provided informed consent prior to the data collection process. Ethical considerations were carefully maintained by ensuring the confidentiality and anonymity of all collected data throughout the research process.

Several instruments were used to collect data in this study, including a poetry recitation performance task, an assessment rubric, and a student perception questionnaire. In the performance task, students were instructed to select an English poem, practice their recitation, and record a video of their oral performance. Poetry recitation was chosen because it integrates linguistic accuracy, fluency, pronunciation, prosody, and expressive interpretation, making it suitable for evaluating multiple aspects of oral proficiency in EFL learning. The assessment rubric was developed based on theories of oral language assessment, reading fluency, and rubric-based evaluation proposed by Brown (2010), Harris (1969), Rasinski (2010), and Andrade (2005). The rubric consisted of five assessment dimensions, namely pronunciation, intonation and stress, fluency, expression, and comprehension, which were evaluated using a five-point scale ranging from very poor to excellent. Pronunciation focused on the accuracy and clarity of sound production, while intonation and stress assessed the appropriate use of prosodic features during reading. Fluency evaluated speech continuity, rhythm, and pacing, whereas expression and comprehension focused on interpretive and emotional aspects of the poetry performance.

The data collection procedure was conducted in several stages. First, students were given approximately two weeks to prepare and practice their selected poems before recording their performances. After completing the preparation stage, students submitted their poetry recitation videos through the university learning management system. A total of sixty-nine videos were collected and used as the primary data source for analysis. The videos were then processed using a Generative Artificial Intelligence platform equipped with automatic speech recognition and prosodic analysis features. The AI system analyzed several measurable speech characteristics, including pronunciation accuracy, speech rate, pause frequency, and pitch variation, and generated preliminary scores and automated feedback based on the assessment rubric. In addition to AI-assisted analysis, the performances were independently evaluated by two human raters consisting of the researcher and an experienced EFL instructor to ensure scoring consistency and reliability.

The collected data were analyzed using both quantitative and qualitative techniques. Quantitative analysis was conducted on the rubric scores obtained from the AI system and the human evaluators using descriptive statistics, including mean scores and score distributions across the five assessment criteria. Correlation analysis and Cohen's kappa coefficient were also applied to examine the level of agreement between AI-generated scores and human ratings. Qualitative analysis was conducted on the responses obtained from the open-ended questionnaire items using thematic analysis procedures. The responses were coded and categorized to identify recurring themes related to students' perceptions of the usefulness,

https://doi.org/10.51878/cendekia.v6i3.10634

CENDEKIA: Jurnal Ilmu Pengetahuan
Vol. 6, No. 3, Mei-Juli 2026
e-ISSN : 2774-4183 | p-ISSN : 2774-8030
Online Journal System : https://jurnalp4i.com/index.php/cendekia

effectiveness, fairness, and limitations of AI-assisted assessment. Ethical approval for the study was obtained from the university research ethics committee prior to data collection, and all recordings and questionnaire responses were anonymized and used solely for academic research purposes.

RESULT AND DISCUSSION

Result

The results of this study are presented in two parts. The first section reports the quantitative findings derived from the rubric-based assessment of students' poetry recitation performances. The second section presents qualitative findings obtained from the student perception questionnaire regarding the use of Generative Artificial Intelligence (GAI) in the assessment process.

Quantitative Findings

A total of sixty-nine student poetry reading videos were collected and evaluated using the AI-assisted assessment rubric. The evaluation combined automated analysis generated by the AI system and human scoring for interpretive aspects of performance. The AI system primarily analyzed pronunciation, intonation, and stress, and fluency, while expression and comprehension were assessed by two human raters due to the interpretive nature of these dimensions. The distribution of scores across the five rubric dimensions is presented in Table 1.

Table 1. Distribution of Scores for 69 Student Poetry Reading Videos

Rubric Dimension	Score 4 (Excellent)	Score 3 (Good)	Score 2 (Fair)	Score 1 (Poor)	Mean Score
Pronunciation	25 (36%)	26 (38%)	12 (17%)	6 (9%)	3.1
Intonation & Stress	15 (22%)	18 (26%)	22 (32%)	14 (20%)	2.6
Fluency	20 (29%)	25 (36%)	15 (22%)	9 (13%)	2.9
Expression	12 (17%)	16 (23%)	25 (36%)	16 (23%)	2.5
Comprehension	18 (26%)	22 (32%)	18 (26%)	11 (16%)	2.7

The results indicate that students generally performed better in technical aspects of oral performance than in expressive aspects. Pronunciation obtained the highest mean score among the five categories, with a mean of 3.1 out of 4. Approximately seventy-four percent of the students achieved scores in the Good or Excellent categories, suggesting that most participants demonstrated relatively accurate articulation of English sounds. This finding indicates that the majority of the students were able to produce clear and understandable pronunciation when reading the selected poems.

Fluency also showed relatively strong performance, with a mean score of 2.9. Around sixty-five percent of the students were rated as Good or Excellent, indicating that many participants were able to maintain a relatively smooth reading pace with minimal hesitation. The AI analysis identified speech rate and pause patterns as indicators of fluency, while human raters observed that students with higher fluency scores tended to maintain better rhythm and continuity during their recitation.

In contrast, students experienced more difficulty with intonation and stress, which had a lower mean score of 2.6. Less than half of the participants achieved Good or Excellent scores

https://doi.org/10.51878/cendekia.v6i3.10634

CENDEKIA: Jurnal Ilmu Pengetahuan
Vol. 6, No. 3, Mei-Juli 2026
e-ISSN : 2774-4183 | p-ISSN : 2774-8030
Online Journal System : https://jurnalp4i.com/index.php/cendekia

in this category. The findings suggest that although students could pronounce words correctly, many struggled to apply appropriate prosodic features such as pitch variation, stress placement, and natural phrasing. These features are particularly important in poetry reading, where prosody contributes significantly to conveying meaning and emotional tone.

Expression was the lowest-scoring dimension, with a mean score of 2.5. Only forty percent of the students were rated as Good or Excellent in expressive delivery. Many students were observed reading in a relatively monotonous tone with limited variation in facial expression, voice modulation, or emotional engagement. These results indicate that expressive interpretation remains a challenging aspect of poetry recitation for many EFL learners, as it requires both linguistic competence and performance skills.

Comprehension scores were slightly higher than expression scores, with a mean of 2.7. Approximately fifty-eight percent of the students demonstrated adequate understanding of the poem's meaning. Human raters evaluated comprehension based on the appropriateness of phrasing, pauses, and interpretive delivery that reflected an understanding of the poem's content. Students who showed stronger comprehension often demonstrated more natural phrasing and more coherent expressive interpretation during their recitation.

To examine the reliability of the evaluation process, inter-rater agreement between the two human raters was calculated using Cohen's kappa coefficient. The analysis produced a coefficient of 0.82, indicating strong agreement between the evaluators. This result suggests that the rubric provided consistent guidelines for evaluating the expressive and interpretive aspects of students' performances. In addition, correlation analysis between AI-generated scores and instructor evaluations showed a high level of agreement ($r = 0.87$, $p < 0.01$), indicating that the AI-assisted assessment system provided reliable and consistent measurements for the technical aspects of oral performance.

Qualitative Findings

In addition to performance assessment data, qualitative data were collected through a student perception questionnaire administered after students received AI-supported feedback. Thematic analysis of the open-ended responses revealed three major themes related to students' experiences with AI-assisted assessment.

The first theme was perceived usefulness. Many students reported that the AI-generated feedback helped them identify pronunciation errors and areas for improvement that they had not previously noticed. Several participants mentioned that the automated analysis provided clear and specific feedback on their speech patterns, including pacing, pronunciation accuracy, and pause frequency. This type of feedback enabled students to reflect more critically on their oral performance and provided practical guidance for improving their reading skills.

The second theme was motivation and engagement. A number of participants indicated that receiving immediate feedback from the AI system encouraged them to practice their poetry recitation more frequently. Students noted that the availability of automated evaluation created a sense of challenge and curiosity, motivating them to improve their pronunciation and fluency. Some students also reported that the use of technology in assessment made the learning process more engaging and interactive compared to traditional teacher-centered evaluation.

The third theme concerned perceived limitations of AI-based assessment. Although many students appreciated the efficiency of the AI system, several participants expressed concerns that AI feedback could not fully capture the emotional and interpretive nuances of poetry performance. Students pointed out that elements such as emotional expression,

https://doi.org/10.51878/cendekia.v6i3.10634

CENDEKIA: Jurnal Ilmu Pengetahuan
Vol. 6, No. 3, Mei-Juli 2026
e-ISSN : 2774-4183 | p-ISSN : 2774-8030
Online Journal System : https://jurnalp4i.com/index.php/cendekia

creativity, and audience engagement were difficult for automated systems to evaluate accurately. These responses suggest that while AI can effectively analyze measurable speech features, human evaluation remains essential for assessing the artistic and interpretive dimensions of oral performance.

The findings demonstrate that AI-assisted assessment can provide reliable and efficient evaluation of technical aspects of poetry recitation, such as pronunciation and fluency, while human evaluators continue to play an important role in assessing expressive and interpretive performance. The integration of AI and human assessment, therefore, offers a complementary approach that can enhance both the efficiency and the pedagogical value of oral language evaluation in EFL contexts.

Discussion

AI-Supported Assessment for Technical Oral Performance

This study demonstrates that GAI can effectively support the evaluation of technical aspects of students' oral performance, particularly pronunciation and fluency. The relatively high scores obtained in these categories indicate that most students were able to produce understandable speech and maintain adequate reading continuity during poetry recitation. The AI-assisted system was capable of analyzing measurable speech features such as pronunciation accuracy, speech rate, pause frequency, and rhythm, allowing the assessment process to be conducted more efficiently and consistently. These findings suggest that AI technologies can reduce the workload of instructors in evaluating large numbers of oral performances while still maintaining reliable scoring procedures. In EFL classrooms where oral assessment often requires substantial time and effort, AI-supported evaluation offers practical advantages for improving assessment efficiency and feedback delivery.

The strong correlation between AI-generated scores and human evaluations further indicates that AI systems can provide reliable measurements for technical dimensions of oral performance. The correlation coefficient obtained in this study showed a high level of agreement between automated and human scoring, particularly in pronunciation and fluency assessment. Similar findings were reported by Deveci (2026), who found that AI-based systems produced evaluation results that closely aligned with human judgments of EFL learners' intelligibility and speaking performance. This finding reinforces the growing evidence that AI-assisted assessment can function as a valid supplementary tool in language evaluation contexts. The consistency between AI and human scoring also indicates that speech recognition and automated analysis technologies are becoming increasingly capable of identifying linguistic patterns associated with oral proficiency.

The findings are also consistent with previous studies emphasizing the effectiveness of AI-powered pronunciation technologies in language learning environments. Research conducted by Abdelhalim and Alsehibany (2025) demonstrated that AI-assisted pronunciation tools significantly improved learners' pronunciation accuracy and increased their learning motivation. Similarly, Sun (2023) reported that automatic speech recognition technologies positively influenced EFL learners' speaking and pronunciation development through immediate corrective feedback. In the present study, students also perceived AI-generated feedback as informative and useful for identifying weaknesses in pronunciation and speech delivery. This suggests that AI-supported assessment does not merely function as an evaluative tool but also contributes to formative learning processes by enabling students to reflect on their oral performance and practice more independently.

https://doi.org/10.51878/cendekia.v6i3.10634

CENDEKIA: Jurnal Ilmu Pengetahuan
Vol. 6, No. 3, Mei-Juli 2026
e-ISSN : 2774-4183 | p-ISSN : 2774-8030
Online Journal System : https://jurnalp4i.com/index.php/cendekia

Prosody and Expressive Challenges in Poetry Recitation

Despite the positive results in pronunciation and fluency, the findings reveal that students experienced greater difficulty in applying appropriate intonation, stress, and expressive delivery during poetry recitation. The lower mean scores obtained in these categories indicate that prosodic competence remains a significant challenge for EFL learners. Poetry reading requires learners not only to pronounce words accurately but also to communicate rhythm, emotional tone, and interpretive meaning through vocal expression. According to Rasinski and Reutzel (2020), prosody represents an essential component of reading fluency because it reflects a reader's ability to convey meaning through appropriate phrasing, stress, and intonation. Students who struggle with prosodic features often produce monotonous delivery, reducing the communicative and aesthetic quality of oral performance.

The findings also indicate that expressive oral interpretation involves more complex cognitive and emotional processes than technical pronunciation accuracy alone. Students may demonstrate acceptable articulation and speech clarity while still experiencing difficulty in conveying emotion, dramatic emphasis, and interpretive understanding. Similar findings were reported by Xue et al. (2024), who emphasized that poetry reading performance is strongly influenced by learners' sensitivity to rhythm, emotional engagement, and literary interpretation. In poetry recitation, effective delivery requires students to understand not only the literal meaning of the text but also its emotional and artistic dimensions. This explains why expressive performance obtained lower scores compared to pronunciation and fluency in the present study.

Human Judgment in Performative Oral Assessment

The findings further reinforce the inherently performative nature of poetry reading. Spolin et al. (1999) emphasized that performance-based language activities involve creativity, emotional interpretation, and spontaneous expression that extend beyond measurable linguistic features. In poetry recitation, students are expected to communicate meaning through voice modulation, facial expression, emotional tone, and interpretive engagement with the audience. These elements are difficult to evaluate solely through automated systems because they involve subjective and context-dependent interpretation. As a result, human evaluators continue to play a crucial role in assessing expressive authenticity and emotional effectiveness in oral performance tasks.

More recent studies also support the importance of human judgment in performance-based language assessment. Lee and Liu (2022) found that drama-based speaking activities require evaluators to consider audience interaction, emotional delivery, and communicative authenticity in addition to linguistic accuracy. Similarly, Burton (2026) argued that spoken-word poetry and performative oral tasks involve affective and interpretive dimensions that cannot be fully captured through automated analysis alone. The present study produced similar findings, as human raters were better able to evaluate expressive interpretation, emotional nuance, and artistic delivery compared to the AI system. These results suggest that although AI technologies can provide efficient technical analysis, human evaluation remains essential for assessing higher-order performative aspects of oral communication.

Pedagogical Implications of AI in EFL Oral Assessment

Another important finding of this study concerns the pedagogical value of AI-supported feedback in EFL learning environments. Students generally perceived the AI-generated feedback as useful, motivating, and beneficial for improving their oral performance. Immediate feedback allowed students to identify weaknesses in pronunciation, pacing, and fluency more

https://doi.org/10.51878/cendekia.v6i3.10634

CENDEKIA: Jurnal Ilmu Pengetahuan
Vol. 6, No. 3, Mei-Juli 2026
e-ISSN : 2774-4183 | p-ISSN : 2774-8030
Online Journal System : https://jurnalp4i.com/index.php/cendekia

efficiently than conventional delayed evaluation methods. This finding aligns with formative assessment theory proposed by Black and Wiliam (1998), which emphasizes the importance of timely feedback in supporting learning improvement. AI-assisted systems therefore have the potential to strengthen formative assessment practices by providing faster and more consistent responses to students' oral performances.

The findings also support recent studies highlighting the growing role of AI technologies in language teaching and assessment. Goh and Aryadoust (2025) argued that AI-assisted speaking assessment can improve feedback quality and assessment efficiency through automated speech analysis and adaptive evaluation systems. Similarly, Ramos-Saltos et al. (2026) emphasized that AI-enhanced speaking practices can support learners' oral development while reducing teachers' assessment burden in EFL classrooms. In the present study, the integration of AI and human evaluation created a complementary assessment model in which AI focused on measurable linguistic features while instructors concentrated on interpretive and expressive aspects of performance. This blended approach offers important pedagogical implications for higher education institutions seeking to integrate AI technologies into language learning and assessment practices.

Overall, the findings suggest that GAI-supported assessment should be viewed as a complementary pedagogical tool rather than a replacement for human evaluators. AI technologies demonstrate strong potential in evaluating technical aspects of oral performance efficiently and consistently, particularly in large classroom settings. However, expressive interpretation, emotional authenticity, and artistic communication still require human judgment and contextual understanding. The combination of AI-assisted analysis and instructor evaluation therefore provides a more balanced and comprehensive assessment framework for EFL oral performance tasks. This integrated approach can enhance assessment efficiency while preserving the pedagogical value of human-centered language evaluation.

CONCLUSION

This study examined the use of Generative Artificial Intelligence (GAI) to support the assessment of EFL students' poetry recitation videos. The findings demonstrate that AI-assisted assessment can effectively evaluate several technical aspects of oral performance, including pronunciation, fluency, intonation, and expressiveness. The results also reveal a strong alignment between AI-generated scores and instructor evaluations, suggesting that AI can provide consistent and reliable measurements. Moreover, students responded positively to the AI-generated feedback, perceiving it as informative, motivating, and helpful for improving their oral communication skills. These findings indicate that GAI has the potential to enhance assessment practices in EFL contexts, particularly in situations where instructors must evaluate large numbers of students.

However, the study also highlights important limitations of AI-based assessment. While AI tools can efficiently analyze technical speech features, they remain limited in capturing the deeper interpretive and emotional dimensions of poetry performance. Elements such as emotional authenticity, interpretive depth, and artistic expression still require human judgment and contextual understanding. Therefore, the findings support the adoption of a blended assessment approach in which AI functions as a supportive tool rather than a replacement for instructors. By combining AI-generated analysis with human evaluation, educators can achieve a more balanced, efficient, and pedagogically meaningful assessment process that supports both linguistic accuracy and expressive communication in EFL learning.

https://doi.org/10.51878/cendekia.v6i3.10634

CENDEKIA: Jurnal Ilmu Pengetahuan
Vol. 6, No. 3, Mei-Juli 2026
e-ISSN : 2774-4183 | p-ISSN : 2774-8030
Online Journal System : https://jurnalp4i.com/index.php/cendekia

REFERENCES

Abdelhalim, S. M., & Alsehibany, R. A. (2025). Integrating AI-powered tools in EFL pronunciation instruction: effects on accuracy and L2 motivation. *Computer Assisted Language Learning*, 1-25. https://doi.org/10.1080/09588221.2025.2534015

Andrade, H. G. (2005). Teaching With Rubrics: The Good, the Bad, and the Ugly. *College Teaching*, *53*(1), 27–31. https://doi.org/10.3200/CTCH.53.1.27-31

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Brown, H. D. (2010*). Language Assessment: Principles and Classroom Practices*. Pearson.

Burton, J. (2026). *Translingual spoken word poetry in an online English for academic purposes class*. The Language Learning Journal, 54(2), 302–316. https://doi.org/10.1080/09571736.2026.2628332

Deveci, C. (2026). Comparing Human and AI Judgments of Turkish EFL Learners' Intelligibility. *Journal of Uludag University Faculty of Education*, *39*(1), 33-49. https://doi.org/10.19171/uefad.1728485

Goh, C. C., & Aryadoust, V. (2025). Developing and assessing second language listening and speaking: Does AI make it better?. *Annual Review of Applied Linguistics*, 1-21. https://doi.org/10.1017/S0267190525100111

Harris, D. P. (1969). *Testing English as a Second Language*. McGraw-Hill Book

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education promises and implications for teaching and learning*. Center for Curriculum Redesign.

Holmes, Wayne., & Luckin, Rose. (2016). *Intelligence unleashed : an argument for AI in education*. Pearson : UCL Knowledge Lab.

Ji, D. (2026). The impact of transformer-based chatbots on second language speaking competence in blended learning. *Discover Computing*, *29*(1), 145. https://doi.org/10.1007/s10791-026-10035-x

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for Language Teaching and Learning. *RELC Journal*, *54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Kuddus, K. (2022). Artificial Intelligence in Language Learning: Practices and Prospects. In *Advanced Analytics and Deep Learning Models* (pp. 1–17). Wiley. https://doi.org/10.1002/9781119792437.ch1

Lee, Y. J. J., & Liu, Y. T. (2022). Promoting oral presentation skills through drama-based tasks with an authentic audience: A longitudinal study. *The Asia-Pacific Education Researcher*, *31*(3), 253-267. https://doi.org/10.1007/s40299-021-00557-x

Liu, Y., binti Ab Rahman, F., & binti Mohamad Zain, F. (2025). A systematic literature review of research on automatic speech recognition in EFL pronunciation. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2466288

Lun, C. W. W., Pau, L. C., & Yunus, M. M. (2026). The Impact of Popular Artificial Intelligence Tools on English Language Teaching and Learning: A Systematic

https://doi.org/10.51878/cendekia.v6i3.10634

Literature Review (2021-2025). *International Journal of Research and Innovation in Social Science (IJRISS)*, *9*(12). https://dx.doi.org/10.47772/IJRISS.2025.91200296

Ramos-Saltos, L. A., Vélez-Olvera, E. R., Piña-Roldán, V. A., & Pereira-Loor, J. M. (2026). AI-Enhanced Speaking Practice in Upper-Secondary EFL Classroom: A Systematic Review of Recent Evidence. *Revista Científica Ciencia y Método*, *4*(1), 516-532. https://doi.org/10.55813/gaea/rcym/v4/n1/168

Rasinski, T. V. (2010). *The fluent reader: Oral & silent reading strategies for building fluency, word recognition & comprehension*. Scholastic.

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, *30*(2), 175–194. https://doi.org/10.1080/0260293042000264262

Santoso, K. H., Al Faruq, M. S., Kurniawati, K., Hidayati, S. P., & Aung, Y. M. M. (2025). Leveraging Chatgpt's Advanced Voice Mode For Enhancing Oral-Aural Skills In Higher Education Language Learning: A Mixed-Methods Study. *Improvement: Jurnal Ilmiah untuk Peningkatan Mutu Manajemen Pendidikan*, *12*(2), 115-124. https://doi.org/10.21009/improvement.v12i2.62614

Shaheed, A., Alam, I., & Al-Hamzi, M. A. (2026). Phonetics In Digital Language Learning: A Systematic Review Of ELT Research (2020–2025). *Journal of Applied Linguistics and TESOL (JALT)*, *9*(1), 40-56. https://doi.org/10.63878/jalt1729

Spolin, V. (1999). *Improvisation for the theater: A handbook of teaching and directing techniques*. Northwestern University Press.

Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: a mixed methods investigation. *Frontiers in psychology*, *14*, 1210187. https://doi.org/10.3389/fpsyg.2023.1210187

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

Xodabande, I., Shiri, S., & Zohrabi, M. (2025). Exploring the impacts of an AI-driven instructional intervention on Iranian EFL learners' pronunciation skill development. *Discover Education*, *4*(1), 307. https://doi.org/10.1007/s44217-025-00782-2

Xue, S., Gao, X., Wu, Y., Sun, J., Yang, W., Li, X., ... & Chen, S. (2025). Effects of Theme, Form, and Language on Poetry Reading: Evidence From Chinese ESL Learners. *Empirical Studies of the Arts*, *43*(1), 672-706. https://doi.org/10.1177/02762374241265866

Zou, B., Liviero, S., Hao, M., & Wei, C. (2020). *Artificial Intelligence Technology for EAP Speaking Skills: Student Perceptions of Opportunities and Challenges* (pp. 433–463). https://doi.org/10.1007/978-3-030-34212-8_17

https://doi.org/10.51878/cendekia.v6i3.10634