

Automatic Vs Human Evaluation of Students' Spoken English Proficiency: A Study of Students' Perceptions

Nur Aeni¹, Asriati Asriati² & Baso Jabu^{3*}

¹English Education, Faculty of Languages and Literature, Universitas Negeri Makassar, Makassar, Indonesia
Email: nur_aeni@unm.ac.id

²English Education, Faculty of Languages and Literature, Universitas Negeri Makassar, Makassar, Indonesia
Email: asriati@gmail.com

³English Education, Faculty of Languages and Literature, Universitas Negeri Makassar, Makassar, Indonesia
Email: basojabu@unm.ac.id

ARTICLE HISTORY

Receive : 10 February 2026
Accepted: 16 May 2026
Published: 24 May 2026

KEYWORDS

Speaking assessment
Student perception
AI evaluation
Human feedback
Hybrid assessment

LICENSE

Copyright © 2025 authors



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

ABSTRACT

Assessing students' spoken English proficiency is a vital component of language education. With recent advances in artificial intelligence (AI) and natural language processing (NLP), automated speaking assessments have emerged as alternatives to traditional human evaluations. This study investigates undergraduate students' perceptions of human versus automated speaking assessments within an English Language Education program at the State University of Makassar. Grounded in Swain's (2006) Linguaging Theory, the research explores how different assessment modes influence learners' motivation, anxiety, engagement, and perceived feedback quality. Employing a descriptive qualitative design, data were collected through semi-structured interviews with nine students who had experienced both human and automated assessment types. Thematic analysis revealed a strong preference for human evaluations due to their ability to capture contextual, emotional, and expressive aspects of speech. Human feedback was perceived as more personalized, meaningful, and motivational. In contrast, automated systems were valued for their efficiency, objectivity, and usefulness in practicing technical elements such as pronunciation and fluency. However, students noted limitations in the depth and clarity of automated feedback. Most participants advocated for a hybrid model that combines the strengths of both approaches. These findings yield important pedagogical implications for the incorporation of artificial intelligence in language assessment, while emphasizing the continued relevance of human involvement in fostering communicative competence.

*Corresponding author: basojabu@unm.ac.id

Aeni, N., Asriati, A., & Jabu, B. (2026). Automatic Vs Human Evaluation of Students' Spoken English Proficiency: A Study of Students' Perceptions. *Journal of English Culture, Language, Literature and Education*, 14(1), 124-133. <https://doi.org/10.53682/z3pbmz48>.

INTRODUCTION

This may constrain their pedagogical effectiveness. The present study builds upon this insight by further examining the issue within English language education, where evaluating learners' oral proficiency plays a crucial role in measuring their overall communicative competence. Traditionally, human evaluators have played a key role in assessing spoken language through face-to-face or oral exams. However, the advent of technology has led to the rise of automatic evaluation systems, which promise more efficient, scalable, and consistent methods for evaluating spoken English. These systems, often powered by artificial intelligence (AI) and natural language processing (NLP), offer a more objective approach to assessment.

Despite their potential, concerns remain regarding their accuracy and their ability to capture the nuances of spoken communication, such as intonation, fluency, and psychological aspects. One relevant theoretical lens for examining this issue is Swain's (2006) *Language Theory*, which underscores the cognitive and metacognitive processes involved in language production, particularly as they unfold through interpersonal interaction and responsive feedback. Although automated assessment systems offer the advantage of delivering swift evaluations, they often fall short in providing the depth of human interaction and dialogic feedback that is essential for fostering meaningful learner development.

In light of these considerations, it is essential to examine learners' perspectives on both automated and human assessment methods. An important question arises: Do students perceive feedback from human evaluators as more meaningful, or do they favor the efficiency and impartiality offered by automated systems? This study seeks to explore these perceptions, along with the underlying factors that shape students' preferences in the context of assessing spoken English proficiency.

LITERATURE REVIEW

Oral Assessment in English Language Learning

Oral assessment is widely recognized as an essential component in evaluating learners' speaking proficiency in English language education. Anggraeni, Endriyati, and Pratolo (2019) emphasize that despite its pedagogical benefits, oral assessment often generates anxiety and undermines learner confidence. This indicates a need for more supportive and transparent assessment practices. Building on their insights, the current study highlights that students' emotional responses to assessment should be a core consideration when designing both human and automated evaluation frameworks, especially in high-stakes speaking tasks.

Furthermore, Widiastuti et al. (2019) highlight the pivotal importance of timely and substantive feedback in supporting student learning and development. Their research underscores the pedagogical view that feedback functions not only as an

evaluative tool but also as a formative mechanism that guides learning. Building upon this perspective, the current study examines whether learners regard human evaluators as more effective than automated systems in providing formative feedback.

Technological Advancements in Automated Evaluation

Technological innovations-particularly in the fields of Artificial Intelligence (AI) and Natural Language Processing (NLP)-have prompted a shift in the way language assessments are conducted. Gao et al. (2023), in their systematic review, note that large language models such as ChatGPT can facilitate scalable, consistent, and time-efficient assessment. While this development is promising, it raises questions about the pedagogical trade-offs between automation and human judgment.

Liang (2025) presents empirical findings indicating that evaluation tools utilizing Natural Language Processing (NLP) surpass conventional assessment methods in both processing speed and scoring accuracy. Nevertheless, this research aims to expand the discussion by examining whether such technical superiority translates into pedagogical effectiveness from the learners' perspective. Moreover, existing critiques-such as those presented by Excelsoft Corp. (2023) point to possible algorithmic bias and lack of contextual understanding in automated systems. The present study builds upon this critique by exploring whether students detect such limitations and how these perceptions influence their preferences for human versus machine evaluators.

Learners' Perceptions of Assessment Methods

The efficacy of any assessment method is inevitably tied to how learners perceive its fairness, relevance, and utility. Kristiyanti et al. (2021) report that students generally view online formative assessments favorably due to their ease of access and user-friendly design. However, the absence of detailed feedback whether automated feedback mechanisms are perceived as meaningful and actionable by students in the context of speaking assessments. In a related study, Babanoğlu and Ağçam (2016) found that students often feel excluded from the assessment design process, which can lead to disengagement and lower motivation. In response to this, the present study explores how involving learners either directly or indirectly in the evaluation discourse can enhance their sense of agency and trust in the assessment process.

Strengths and Limitations of Human and Automated Approaches

The existing body of literature delineates distinct differences between human-led and automated approaches to language assessment. Human assessors are capable of providing detailed, personalized feedback and are better equipped to evaluate pragmatic and sociolinguistic subtleties. Nonetheless, this form of assessment tends

to be time-consuming and may be affected by issues of reliability and inter-rater variability (Anggraeni et al., 2019; Babanoğlu & Ağçam, 2016). Conversely, automated evaluation systems, while efficient and objective, may lack the interpretative flexibility necessary for assessing more nuanced aspects of spoken language, such as intonation or discourse coherence (Liang, 2025; Excelsoft Corp., 2023). This study seeks to bridge these two paradigms by exploring how students weigh the trade-offs between consistency and contextual sensitivity.

Positioning the Current Study in the Existing Body of Research

Although previous studies have significantly advanced the understanding of the advantages and limitations inherent in both human and automated assessment systems, there remains a noticeable gap concerning learners' subjective experiences, especially in the context of spoken English proficiency. By centering on students' perspectives, this research seeks to fill that gap and offers a more learner-oriented insight into the effectiveness of assessment practices. Furthermore, by incorporating Linguaging Theory (Swain, 2006), this research offers a theoretical framework to examine how interactional dynamics in assessment settings (human vs. machine) influence learners' cognitive and linguistic development.

RESEARCH METHOD

This study will adopt a descriptive qualitative research design to explore students' perceptions of both human and automated speaking assessment methods. The qualitative design is chosen to allow for deeper analysis of students' opinions, experiences, and affective reactions.

Participants

The participants in this study will consist of undergraduate students enrolled in the English Language Education program at Universitas Negeri Makassar. Participants will be selected using purposive sampling, ensuring that they have prior experience with both human (e.g., teacher-led assessments) and automated speaking evaluations (e.g., AI-powered language learning platforms, speech recognition tools) in their speaking courses. The selection will aim to represent students with a variety of academic levels (from lower to upper semesters) to gain diverse insights into students' perceptions of each evaluation method.

Instruments

The interview will use semi structure interview guide. Semi-structured interviews will conduct with participants to gain deeper insight into their experiences and reasoning.

Data Collection Procedure

The research will be conducted over four weeks. interviews will be conducted either via Zoom or face-to-face depending on participant availability. Ethical approval will be obtained prior to data collection, and informed consent will be secured from all participants.

Data Analysis

Quantitative data from questionnaires will be analyzed using descriptive statistics. Qualitative data from interviews will be analyzed using thematic analysis, following the six-phase framework by Braun & Clarke (2006): familiarization, coding, theme generation, review, naming, and final report writing.

FINDINGS AND DISCUSSION

Findings

This research explores students' perspectives on English speaking assessment, comparing evaluations made by humans (instructors or examiners) and automatic evaluations through AI-based applications or software. A total of 9 respondents, all students majoring in English, participated in the study and provided valuable feedback on their experiences in undergoing speaking assessments. The results of the study are summarized as follows:

Themes and Categories in Speaking Evaluation

Based on the data collected, several key themes and categories emerged related to the two methods of speaking assessment: human-based evaluation and automatic evaluation via AI applications or software. The discussion below highlights the respondents' views and experiences related to each of these methods.

Preference for Human Evaluation

A majority of the respondents (7 out of 9) preferred human evaluation, particularly from instructors or examiners. The reasons for this preference were consistent across their responses:

1. Ability to Capture Context and Expression

A number of participants underscored the capacity of human evaluators to more effectively interpret the broader contextual dimensions of their spoken language performance. For example: Respondent 1 stated, *"The feedback I get from a human evaluator is more nuanced. They can understand the context of what I'm saying, my tone, and whether I'm making a point clearly, something an automatic system might miss."* Similarly, Respondent 6 noted, *"Humans can grasp the intention behind my words and*

offer feedback on things like my body language or how my ideas are presented, which is something AI can't evaluate effectively."

2. More Personal Feedback

Respondents expressed that feedback from humans is more personal and relevant to their learning needs.

Respondent 9 expressed that feedback provided by a human evaluator is perceived as more meaningful, particularly because it includes specific, targeted suggestions for improvement. The participant noted that such feedback is especially helpful in addressing personal challenges, such as issues related to intonation and clarity in speech.

Respondent 8 echoed this, saying, *"Human evaluators can help me fix specific mistakes and explain why something didn't sound right in a way that makes sense to me."*

3. Support for Self-Development

Many respondents reported feeling motivated after receiving feedback from humans.

Respondent 4 mentioned, *"After being evaluated by my professor, I felt more confident. The feedback was constructive, and it pushed me to improve."* In contrast, Respondent 7 shared, *"I always feel more motivated when I receive human feedback because I know it's tailored to me and my weaknesses."*

Advantages and Disadvantages of Automatic Evaluation

While many respondents preferred human evaluation, most acknowledged the usefulness of automatic evaluation systems, especially for practicing technical aspects of speaking. The benefits mentioned included:

1. Time Efficiency and Consistency

Respondent 5 stated, *"Using an AI app like Duolingo helps me practice anytime. It's convenient, and I get immediate feedback."*

Respondent 3 also highlighted the speed of automatic systems: *"AI systems are fast. They can give me a lot of feedback in a short time, which is great for practice."*

2. Objectivity in Assessment

Respondent 6 emphasized that automatic evaluations are objective: *"AI doesn't have any bias. It evaluates based on my pronunciation and fluency without being influenced by my mood or the evaluator's preferences."*

However, respondents also identified several shortcomings of automatic evaluation:

1. Limited in Capturing Context and Emotion

Most respondents expressed that AI systems are not good at understanding the emotional context of their speaking.

Respondent 8 explained, *“AI doesn’t understand when I’m joking or emphasizing a point with my tone. It just looks at the technical side, like pronunciation.”* Similarly, **Respondent 9** mentioned, *“Sometimes AI doesn’t get what I’m trying to say, especially if I use a word in a non-standard way. It’s not flexible enough to understand meaning.”*

2. Lack of In-Depth Explanations

Respondent 4 noted, *“AI gives feedback, but it’s too technical and lacks explanations. I get corrected on pronunciation, but I don’t always know how to fix it or why I made the mistake.”*

Comparison Between Human and Automatic Evaluation

A substantial majority of participants (eight out of nine) expressed the view that an optimal assessment approach would involve a combination of both human and automated evaluations. Participants recognized the value of automated systems in facilitating technical skill development such as pronunciation and fluency while emphasizing the indispensable role of human evaluators in delivering more holistic and contextually informed assessments.

Respondent 3 said that, *“I think it’s best to combine both. I can use AI to get quick feedback on pronunciation, but I still need a human to help me with how to communicate my ideas more effectively.”*

- a. **Humans in Evaluation:** **Respondent 7** pointed out that humans excel at understanding the context, emotion, and The participant emphasized that *“human evaluators are able to provide feedback not only on linguistic accuracy but also on the effectiveness of idea delivery. As noted, humans can assess the underlying meaning being conveyed, offering insights into how clearly and effectively the speaker communicates their intended message, beyond mere language mechanics.”*
- b. **Automatic Systems for Practice:** **Respondent 2** agreed that automatic systems are ideal for practice *“One respondent acknowledged the usefulness of AI-based systems in supporting technical aspects of spoken language, such as pronunciation and fluency. However, they also noted that human input remains essential for addressing more complex elements of communication, including the organization and coherence of argumentative structures.”*

Emotional Impact of Evaluation on Students

Respondents felt differently about the emotional impact of human and automatic evaluations. Many noted that evaluations by humans often caused anxiety, but also motivated them to improve:

Human Evaluations: **Respondent 6** said that, *"I get nervous when I am being evaluated by humans, but it pushes me to work harder. The feedback is real and immediate, and I know it will help me to get improved."*

Automatic Evaluations: In contrast, **Respondent 5** felt more comfortable and at ease using automatic systems: *"I feel more relaxed with AI feedback because it's not personal. I have more practice as much as I want without the fear of judgment."*

Discussion

The findings suggest that both human and automated evaluations offer distinct advantages. Human evaluation offers a more comprehensive assessment of speaking, taking into account factors such as context, expression, and the clarity of ideas. Feedback from human evaluators is perceived as more personalized, contextually relevant, and motivational, thus contributing more effectively to the development of students' speaking skills.

On the other hand, automatic evaluation offers benefits like efficiency, objectivity, and the ability to provide quick feedback on technical aspects such as pronunciation and grammar. However, it lacks the depth and flexibility of human feedback, particularly in capturing emotional context or providing nuanced suggestions. The ideal method for speaking assessment, according to the majority of respondents, is a combination of both human and automatic evaluation. Automatic systems can be used for regular practice, particularly for technical aspects, while human evaluators remain essential for assessing more complex communication elements.

Suggestions for Improving Speaking Assessment for Students

Based on the findings, several recommendations for improving speaking assessments for students are as follows:

1. **Use of Clear and Standardized Rubrics:** **Respondent 4** suggested, *"Using clear rubrics would make the assessment process more objective and transparent, so students know exactly what they need to improve."*
2. **Practice Based on Real-Life Situations:** **Respondent 9** emphasized the importance of real-life tasks: *"Assessments should reflect real communication, like presentations or discussions, to better prepare us for real-world speaking situations."*
3. **Improvement of Automatic Evaluation Technology:** Several respondents, like **Respondent 2**, recommended improving AI systems to capture a broader range of speaking contexts: *"AI needs to improve so that it can evaluate more than*

just pronunciation. It should understand different accents and be able to provide more detailed feedback."

4. **Providing More Personal Feedback:** Respondent 7 advised that human evaluators should ensure their feedback is specific and actionable: *"When giving feedback, evaluators should be specific and clear. We need suggestions we can apply right away."*

CONCLUSION

This study aimed to explore students' perspectives on English speaking assessment by comparing human evaluations with automatic, AI-based evaluation systems. The findings, based on interviews with nine respondents, revealed that most students preferred human evaluators, citing their ability to understand emotional nuance, provide contextual feedback, and deliver more personalized and motivating input. However, students also recognized the benefits of automatic evaluation, particularly its speed, objectivity, and usefulness for practicing pronunciation and fluency at their own pace. Interestingly, the majority of respondents agreed that a combination of both methods would be the most effective approach.

While AI applications are helpful for technical practice, human evaluation remains essential for providing holistic assessments that account for expression, clarity of ideas, and communicative effectiveness. Students also noted the emotional dimensions of assessment: human evaluations could be anxiety-inducing but also motivational, whereas AI-based assessments were less stressful but sometimes lacked clarity.

ACKNOWLEDGMENTS

First and foremost, the researchers would like to express my deepest gratitude to the students who participated as respondents in this research. Their willingness to share their time, thoughts, and experiences made this study possible. The insights they provided were invaluable in understanding students' perspectives on English speaking assessments, and their honest responses greatly enriched the quality of this research. The authors would like to express their gratitude to Universitas Negeri Makassar for funding this research through the *PNBP (Non-Tax State Revenue)* scheme under Contract Number [168/DST/UN36.12/TU/2026]. This support was instrumental in the successful completion of the study.

REFERENCES

- Aeni, N., Nur, S., Limbong, S., & Yunus, M. (2024). Promoting EFL students' engagement by using Bamboozle: Digital game-based learning in Indonesian higher education. *Journal of English Education and Teaching*, 8(3), 508-527.

- Aeni, N., Muhalim, M., Ganteng, H., Tahir, M., & Talib, A. (2025). English Education Students' Perceptions of Automated vs Human Assessment in Spoken English Proficiency. *AL-ISHLAH: Jurnal Pendidikan*, 17(3), 5615-5625. <https://doi.org/10.35445/alishlah.v17i3.7655>
- Anggraeni, M., Endriyati, C., & Pratolo, B. W. (2019). Students' perception on the implementation of oral assessment in speaking class. *Journal of Language Teaching and Research*, 10(5), 1020-1027.
- Babanoğlu, M. P., & Ağçam, R. (2016). The evaluation of speaking skills in EFL classrooms. *Journal of Language and Linguistic Studies*, 12(2), 42-53.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Excelsoft Corp. (2023). *Bias and limitations in AI-powered assessment systems: A technical white paper*.
- Gao, L., Zhang, M., & Liu, X. (2023). Artificial intelligence in language assessment: A systematic review. *Computer Assisted Language Learning*, 36(1), 1-24.
- Kristiyanti, R., Anugerahwati, M., & Widiati, U. (2021). Students' perception toward online formative assessment in speaking class. *Journal of English Education and Linguistics Studies*, 8(2), 232-247.
- Liang, Y. (2025). Exploring the pedagogical validity of NLP-based speaking assessment tools in higher education. *Journal of Educational Technology and Language Learning*, 12(1). (In press)
- Swain, M. (2006). Languaging, agency and collaboration in advanced second language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95-108). Continuum.
- Widiastuti, I. A. M. S., Artini, L. P., & Jayantini, I. G. A. S. R. (2019). Students' perception toward the use of self-assessment and peer assessment in speaking class. *Jurnal Pendidikan Bahasa Inggris Undiksha*, 7(1), 10-19. (Example DOI only - adjust if needed)