p-ISSN 2252-7427, e-ISSN 2541-3635
https://ieflj.uniku.ac.id/pub

p-ISSN 2252-7427 e-ISSN 2541-3635
Indonesian EFL Journal
AISEE
The Association of Indonesia Scholars of English Education

AI SPEECH RECOGNITION IN HIGH-SCHOOL EFL SPEAKING ASSESSMENT: EFFECTIVENESS, MEANINGFULNESS, ENJOYMENT

Namira Amaliah Putri[1], Siti Zulaiha[*2], Herri Mulyono[3], Jihan Fakhira[4]
[1,2,3,4]*English Language Education, Graduate School*
Universitas Muhammadiyah Prof. DR. HAMKA, Jakarta, Indonesia

Corresponding author's email: siti.zulaiha@uhamka.ac.id

APA Citation: Putri, N. A., Zulaiha, S., Mulyono, H., & Fakhira, J. (2026). AI speech recognition in high-school EFL speaking assessment: Effectiveness, meaningfulness, enjoyment. *Indonesian EFL Journal*, 12(2), 387-396. https://doi.org/10.25134/ieflj.v12i2.67

Received: 02-01-2026 Accepted: 12-03-2026 Published: 14-05-2026

Abstract: This mixed-methods study examined the use of an AI-powered automatic speech recognition (ASR) tool (SpeechAce) for high-school EFL speaking assessment, with a focus on perceived effectiveness, meaningfulness, and enjoyment. Thirty students completed pre- and post-surveys, and ten students participated in semi-structured interviews; classroom observations were documented using a structured checklist. The intervention lasted two weeks (four 40-minute meetings) in which students accessed SpeechAce on their smartphones via the web and followed an iterative speak–check–revise routine. Quantitative results (paired-samples t-tests) showed significant gains in perceived speaking skill and meaningfulness, whereas enjoyment remained stable. Qualitative findings indicated that automated, non-judgmental feedback supported lower anxiety, increased willingness to speak, and greater learner autonomy, while teacher facilitation (warm-ups, modelling, and brief mini-lessons) helped students interpret feedback and set revision goals. These findings suggest that ASR-supported assessment can function as a formative learning episode in school contexts when tasks, feedback interpretation, and revision opportunities are explicitly structured.
Keywords: *AI-powered speech recognition; joyful learning; meaningful learning; speaking assessment; speaking proficiency.*

INTRODUCTION

Spoken English is a core competency in an interconnected world (Nageen, Sarwar, Alam, Jabeen, & Tayyab, 2025), yet classroom opportunities to practise speaking and receive actionable feedback remain limited in many EFL settings (Mingyan, Noordin, & Razali, 2025; Zheng, Chen, Zhang, & Chai, 2024). In school contexts, speaking assessment is often conducted through teacher-led performance tests that provide global scores but offer limited diagnostic information for revision, which may constrain students' engagement and perceived progress (Hattie & Timperley, 2007; Pekrun & Linnenbrink-Garcia, 2022)

How speaking is assessed therefore remains a pedagogical concern (Aedo Cancino, 2025). Conventional assessment formats can be experienced as high-stakes and judgmental, particularly when students have uneven pronunciation and fluency, which may intensify anxiety and reduce willingness to take risks in oral production (Aedo Cancino, 2025; Derwing & Munro, 2005; Özdemir & Seçkin, 2025; Yousif, 2025). Recent discussions on AI and automated scoring also highlight potential fairness issues related to accent and variation, underscoring the need for transparent classroom use and supportive teacher mediation (Prinos, Patwari, & Power, 2024; Veliche et al., 2024).

Advances in AI-powered automatic speech recognition (ASR) offer a complementary approach by providing immediate, standardized feedback that can be used as formative input for iterative practice (Kim, Park, Kim, Na, & Kim, 2024; Song, Rojas, & Pycha, 2025). Recent syntheses and classroom studies report that ASR-supported practice can contribute to pronunciation and speaking development and can strengthen learners' perceptions of control over

improvement when feedback is interpretable and revision opportunities are built into tasks (Liu, Rahman, & Zain, 2025; Ngo, Chen, & Lai, 2024; Sun, 2023)

However, empirical work in secondary-school classrooms remains comparatively limited, particularly regarding how students experience ASR-supported speaking assessment as a learning episode—that is, whether the process is perceived as effective, meaningful, and enjoyable in authentic classroom conditions. Tool-specific implementation features (e.g., access mode, task prompts, and feedback displays) may shape these perceptions and therefore warrant explicit reporting (SpeechAce, 2025; Zou et al., 2024).

This study responds to these gaps by implementing an AI-powered speech recognition tool in a high-school EFL speaking assessment and examining (1) perceived effectiveness based on pre- and post-surveys and (2) students' perceptions of meaningfulness and enjoyment based on surveys, interviews, and observations. (1) To what extent does the use of AI-powered speech recognition technology in speaking assessment improve students' perceived speaking skills, based on pre- and post-survey results? (2) How do students perceive the meaningfulness and enjoyment of learning when participating in an AI-powered speech recognition-based speaking assessment?

METHOD

This study used an explanatory sequential mixed-methods design in which quantitative pre–post survey data were collected first and then complemented by qualitative interviews and classroom observations to explain the quantitative patterns. This design was selected to capture both measurable changes in perceptions and the classroom processes through which students interpreted and used ASR feedback (Creswell & Creswell, 2017; Tashakkori & Teddlie, 2010). Procedures also followed recent recommendations for transparent reporting and ethical conduct in mixed-methods classroom research (Klefbeck, 2024; Schlunegger, Zumstein-Shaha, & Palm, 2024)

Participants were 30 high-school EFL students from one intact class in a public/private school context in Jakarta, Indonesia. The class followed a regular English curriculum and had prior experience using smartphones for learning activities. The teacher integrated SpeechAce as part of classroom speaking assessment activities during the intervention period.

Ethical approval and school permission were obtained prior to data collection. Students and their guardians received information sheets explaining the study aims, procedures, and data use. All participation was voluntary, and students could decline or withdraw without penalty. Survey and interview data were anonymized, and reporting avoids personally identifiable information.

Perceptions were measured with two parallel questionnaires administered before and after the intervention. The instrument contained 30 Likert-type items grouped into three subscales—perceived speaking skill, perceived meaningfulness, and enjoyment (10 items each). Items were adapted from recent technology-supported speaking and feedback studies and then contextualized to the ASR-assisted assessment setting (Grassini, 2023; Keumalasari, Iqbal, Aulia, & Pranata, 2024; Toro-Troconis, Inzolia, & Ahmad, 2023). To establish content validity, the draft items were reviewed by two EFL assessment experts for relevance, clarity, and alignment with the target constructs; minor wording revisions were made. A brief pilot with non-participating students was conducted to confirm comprehensibility and to refine instructions. Responses were recorded on a 5-point scale (1 = strongly disagree to 5 = strongly agree).

The AI speech recognition (ASR) tool used in this study was SpeechAce, a web-based speaking assessment and practice system that scores spoken responses automatically and provides immediate diagnostic feedback (Nguyen, Vo, & Tran, 2025). After each submission, SpeechAce generates an overall score and component indicators, including pronunciation and fluency, and also reports language-use dimensions such as grammar and vocabulary in its detailed report (SpeechAce, 2025; SpeechAce API, 2025). In this study, SpeechAce was used to deliver standardized prompts and to provide instant feedback that students could interpret and use to revise subsequent responses within the same lesson, thereby supporting an iterative speak–check–revise routine.

After the post-intervention survey, semi-structured interviews were conducted with a purposive subsample of ten students representing different genders and grade levels. Each interview lasted about 15 to 20 minutes and explored usefulness, meaningfulness, enjoyment, confidence changes, challenges, and suggestions for improvement. Open prompts and flexible probes were used to elicit concrete examples while

p-ISSN 2252-7427, e-ISSN 2541-3635
https://ieflj.uniku.ac.id/pub

keeping sessions comparable across participants (Staddon, 2023; Trundle, Gardner, & Phillips, 2024). Interviews were conducted in Bahasa Indonesia to help students express their views fully and to reduce language barriers. Using the participants' primary language supported detailed, context-rich accounts and clearer responses, which aligns with recommendations to interview in the language most comfortable for respondents (Xu, 2023).

Observation Checklist

Classroom observations were conducted in each meeting using a structured checklist to document implementation features (e.g., task flow, students' engagement, frequency of feedback checking, revision attempts, and technical constraints). Short field notes were written during activities and expanded immediately after each meeting to ensure procedural detail and to support triangulation with survey and interview findings.

Because the classroom teacher also served as the primary researcher, steps were taken to reduce researcher bias. Observation notes were divided into descriptive accounts and interpretive comments, and a brief reflexive memo was written after each meeting to record assumptions and emerging interpretations. In addition, interview questions were phrased to invite both positive and critical experiences, and quantitative and qualitative findings were compared systematically during integration.

The treatment was implemented for four meetings over a two-week period (40 minutes per meeting). Students accessed SpeechAce using their smartphones. Each meeting began with a short warm-up in which students engaged in paired conversation about a teacher-provided topic. Following this interaction, students worked individually with SpeechAce: the teacher instructed students to access SpeechAce via the web, select a relevant SpeechAce topic/form, and respond to the related prompts/questions. Students then reviewed the automated feedback (scores and descriptive indicators) and used it to adjust subsequent responses within the same lesson, consistent with an iterative speak–check–revise routine.

Survey data from the pre- and post-questionnaires were screened for completeness and coded consistently (including reverse-keyed items). Internal consistency was examined using Cronbach's alpha and McDonald's omega for each administration. For inferential analysis, item scores were aggregated into three subscale scores (speaking skill, meaningfulness, enjoyment), and paired-samples t-tests were conducted to compare pre- and post-intervention means (df = 29). Effect sizes were estimated using Cohen's dz ($t/\sqrt{n}$), and statistical significance was evaluated at $\alpha = .05$. Qualitative interview data were analysed thematically following Braun & Clarke (2021) procedures (familiarisation, initial coding, theme development, review, and definition), and observation notes were used to triangulate and contextualise participants' accounts. All quantitative analyses were conducted using Jamovi.

RESULTS AND DISCUSSION

Quantitative results

This section presents the classroom results for the AI-powered speech recognition tool.

The quantitative analysis proceeded in two steps. First, internal consistency of the three subscales was examined for the pre- and post-surveys. Second, paired-samples t-tests were conducted on the subscale scores to examine pre–post differences in perceived speaking skill, perceived meaningfulness, and enjoyment.

Reliability of measures

The scale demonstrated strong internal consistency at both administrations (Tables 1–2), indicating that the items cohered within each construct and were suitable for tracking change over time.

Table 1. *Pre*-survey's reliability test

Scale Reliability Statistics

	Cronbach's α	McDonald's ω
scale	0.883	0.891

Table 2. Post-survey's Reliability Test

Scale Reliability Statistics

	Cronbach's α	McDonald's ω
scale	0.862	0.857

Both administrations showed high internal consistency, supporting the use of the three subscales for subsequent pre–post comparisons.

Table 3. *Pre–post differences by subscale (paired-samples t-tests, n = 30)*
Note. Pre and Post values are subscale means and standard deviations. Mean Difference is computed as Pre – Post. df = 29.

Subscale	Pre-test (M ± SD)	Post-test (M ± SD)	Mean Difference (Pre–Post)	t(df)	p
Speaking skill	26.20 ± 5.28	29.93 ± 2.27	-3.73	-3.073 (29)	.005
Meaningfulness	24.53 ± 3.57	28.27 ± 3.66	-3.74	-5.901 (29)	< .001
Enjoyment	28.20 ± 4.25	27.00 ± 3.60	1.20	1.392 (29)	.174

Paired-samples t-tests indicated significant pre–post gains in perceived speaking skill and perceived meaningfulness (Table 3). Perceived speaking skill increased from M = 26.20 (SD = 5.28) to M = 29.93 (SD = 2.27), t(29) = -3.07, p = .005, dz = 0.56. Perceived meaningfulness increased from M = 24.53 (SD = 3.57) to M = 28.27 (SD = 3.66), t(29) = -5.90, p < .001, dz = 1.08. Enjoyment did not change significantly, suggesting that students' positive affect toward the assessment was maintained across the intervention (t(29) = 1.39, p = .174).

Enjoyment scores were relatively high at baseline and remained high after the intervention, which may indicate a ceiling effect or that students already viewed speaking assessment sessions positively before using the ASR tool. Accordingly, qualitative findings are used to clarify how students experienced enjoyment and what classroom conditions supported sustained positive engagement.

Construct-focused interpretation

Overall, the quantitative pattern suggests that students perceived the ASR-supported assessment as both more effective (in terms of speaking skill) and more meaningful after the intervention, while enjoyment remained consistently positive. This pattern is compatible with a formative interpretation of the activity: students reported clearer information about performance gaps and next-step actions, which aligns with feedback models that emphasise actionable, revision-oriented information (Hattie & Timperley, 2007)

Taken together, the reliability evidence and subscale comparisons provide a quantitative basis for interpreting students' qualitative accounts of reduced anxiety, clearer guidance for improvement, and sustained enjoyment during the sessions.

Qualitative results

The qualitative analysis used ten semi-structured interviews and classroom observations to explain the survey trends. Transcripts were coded inductively and grouped into themes that describe how students experienced ASR-based speaking assessment. Excerpts are combined to present a clear story for each theme, and observation notes are used to confirm or nuance what students said.

Non-judgmental feedback lowers anxiety and invites risk-taking

Students compared the AI's feedback with teacher-led tests. They called the AI "neutral," "not judging," and "just showing what to fix." This tone made speaking feel safer. One student said, "*I don't get too anxious with AI because it doesn't judge; I can speak more relaxed*." Another added, "*If I make a mistake, it just points to it, not like a person reacting.*" Several linked this to longer turns and trying new words: "*Before, I stopped quickly because I worried about being wrong. Now I continue and try to finish my idea*." Observation notes showed steady participation, including from usually quiet students, which supports the low-pressure feel of the sessions.

Immediate, clear feedback supports on-the-spot correction

Students stressed that feedback came "right away," was "specific," and "told exactly what to change." One explained, "*I really like how the AI gives instant and clear feedback, so I'm not confused about where to start correcting*." Another said, "*It shows the problem in my pronunciation and gives the exact word to practice. I try again immediately and hear the difference*." Learners described a speak–check–revise rhythm within the same task: "*The quick response makes me fix it on the spot instead of waiting and forgetting*." Observations showed students glancing at feedback panels, re-recording within seconds, and the teacher confirming corrections—evidence that self-repair became routine.

Perceived gains in accuracy and confidence

p-ISSN 2252-7427, e-ISSN 2541-3635
https://ieflj.uniku.ac.id/pub

Across interviews, students linked improvement to repeated, feedback-rich practice. One noted, "*It's much better now—the biggest change is confidence*." Another added, "*My pronunciation is clearer and I can keep the sentence going*." Many tied accuracy to confidence: "*Because I know what to change, I'm braver to speak longer*." Students also reported better pacing and fewer hesitations. Although these are perceptions, observations of longer utterances and fewer aborted attempts later in the unit support the pattern.

Flexibility increases practice and motivation

Students valued being able to practice "anytime," without waiting for a partner or a grade. Comments included, "*I can practice after school when it's quiet; I don't have to ask someone to help*," and "*If I have five minutes, I can try two or three times*." This access turned spare moments into extra rehearsal. Some set simple goals such as "*record until I reduce the mistakes.*" Observation notes showed students using the first or last minutes of class for additional attempts.

Enjoyment and a game-like feel

Many described the sessions as enjoyable and, for some, "like a game." One student said, "*At first it felt strange talking to AI, but later it became fun—like an exciting game—so I'm not afraid to speak; I even want to keep trying*." Others connected enjoyment to progress: "*Seeing fewer errors makes me want to continue*." The class atmosphere matched this blend of focus and energy, with active but task-oriented talk and frequent voluntary retries.

Persistent challenge: finding the right word

Even with overall positive experiences, students sometimes struggled with lexical retrieval. Typical comments were, "*Sometimes I know what I want to say but not the exact word*," and "*Pronunciation is clearer now, but vocabulary still blocks me*." Students framed this as a personal gap rather than a tool problem and suggested prompts or word banks before speaking. Observations captured brief pauses and self-talk during word searches, indicating that vocabulary remains a constraint while pronunciation and fluency improve.

Observation corroboration

Observation records showed high on-task behavior, appropriate tool use, and smooth movement between attempts. Students followed instructions, consulted feedback, and revised with little prompting. The teacher mainly clarified interface questions and highlighted recurring errors. Overall, students experienced ASR-supported assessment as low-threat, informative, and engaging. Neutral feedback reduced anxiety, instant guidance enabled quick correction, and flexible access increased practice. Enjoyment seemed to come from visible progress, not novelty alone. The continuing vocabulary challenge suggests pairing ASR-based assessment with explicit lexical support. These themes help explain the positive shifts seen in the quantitative data and highlight design features that matter for classroom use.

This study examined an AI-powered speech recognition (ASR) tool used in everyday speaking assessment for high-school EFL learners. The discussion links the findings to the research questions and places them within current work on assessment and motivation, while staying close to what happened in the classroom.

For the first research question, the quantitative results show that ASR worked as an assessment tool, not only as a practice aid. Reliability was high at both time points, and paired tests showed clear pre–post change. Students endorsed statements that the assessment helped them monitor pronunciation, fluency, and lexico-grammatical control and that expectations were clearer. Interview accounts pointed to real-time, criterion-aligned feedback that helped students locate problems and correct them immediately; observations recorded steady, on-task use in normal lessons. Together, these results suggest that ASR produced interpretable evidence of speaking performance within routine class activity, in line with reports on AI-supported speaking assessment (Bachiri, Mouncif, Bouikhalene, & Hamzaoui, 2024; Orosoo et al., 2025; Sutomo, 2024; Zou, Du, Wang, Chen, & Zhang, 2023).

For the second research question, perceptions of meaningfulness and enjoyment improved. Students viewed the assessment as more meaningful because feedback was specific, immediate, and reusable in the next attempt. This linked tasks to communicative goals and made criteria visible. Enjoyment rose as sessions shifted from high-pressure, one-shot testing to a low-threat, rehearse-and-resubmit routine with visible progress. Qualitative data connect these shifts to an impartial feedback stance that reduced evaluative threat and to prompt guidance that showed what to improve and how. These patterns

fit work showing that AI-mediated environments can build confidence, reduce anxiety, and sustain engagement (Liu, Darvin, & Ma, 2024; Qiao & Zhao, 2023; Wei, 2023; Zou et al., 2023)

Three mechanisms help explain the overall pattern. First, students described the feedback as factual and non-judgemental. This reframed assessment as information rather than interpersonal evaluation and encouraged risk-taking in speech, which helps explain lower anxiety and higher confidence (Banihashem, Kerman, Noroozi, Moon, & Drachsler, 2024; Escalante, Pack, & Barrett, 2023). Second, the speed and clarity of feedback supported noticing and learner-initiated correction. Learners identified mispronunciations or grammatical slips and changed their output within the same session, consistent with the value of timely, specific guidance (Alshammari & Alomaim, 2025; Liu, Hwang, Yu, Tu, & Wang, 2025; Sun, 2023). Third, flexible access increased practice frequency; students could rehearse without waiting for a partner or a grading cycle, extending opportunities beyond fixed lesson times (Ng, Tan, & Leung, 2024). These mechanisms align with assessment-for-learning principles and offer a coherent explanation for the gains in perceived effectiveness, meaningfulness, and enjoyment.

The qualitative evidence also clarifies the limits of these effects. A recurring difficulty was lexical retrieval during spontaneous production: several students still struggled to find precise words even as pronunciation and fluency improved. This supports the view that vocabulary needs targeted instruction and repeated, contextualized exposure to build fast access, and that pronunciation-focused feedback alone will not close lexical gaps (Nation & Hunston, 2013; Schmitt, 2008; Webb & Nation, 2017). It also matches how many ASR systems prioritize segmental and suprasegmental accuracy and basic form features rather than lexico-semantic growth, which reinforces the need to pair ASR with explicit vocabulary work (Bashori, van Hout, Strik, & Cucchiarini, 2024; Derwing & Munro, 2005; Ngo et al., 2024). A small number of students mentioned time pressure and connectivity interruptions. These issues show that pacing and infrastructure can influence perceived control if they are not managed carefully (Hew & Brush, 2007; Khoirunnisa & Trisanti, 2025; Ng et al., 2024). Although a novelty effect cannot be fully ruled out, engagement stayed stable across sessions and students gave specific examples of corrections prompted by feedback. This points to assessment design features—not short-term interest—as the main driver of the results (Kratochvil, Vaculik, & Macak, 2023).

CONCLUSION

In sum, the integrated survey, interview, and observation evidence suggests that ASR-supported speaking assessment can function as a formative learning episode in high-school EFL classrooms. Students reported higher perceived speaking skill and meaningfulness after the two-week implementation, and their enjoyment remained consistently positive. Qualitative accounts further indicate that immediate, non-judgmental feedback and repeated opportunities to revise helped students manage anxiety, monitor progress, and develop greater confidence in speaking.

For classroom practice, the findings underscore the importance of explicit task design and teacher facilitation. Short warm-ups, clear instructions for interpreting automated feedback, and structured opportunities for resubmission can help students translate ASR indicators into concrete revision goals. When used in this way, ASR tools such as SpeechAce may complement teacher assessment by providing timely diagnostic information while maintaining a supportive classroom climate.

This study has limitations. The sample was drawn from one intact class, which constrains generalisability. The intervention was short (two weeks), and the outcomes reflect students' perceptions rather than externally rated speaking proficiency. Because SpeechAce is an ASR system, feedback quality may vary across learners and accents; therefore, the tool should be used as formative guidance rather than as a high-stakes scoring device. Finally, the teacher–researcher role may have influenced classroom dynamics despite the mitigation steps described; future studies could involve independent observers and longer interventions to examine stability of effects.

ACKNOWLEDGEMENT

This research is funded by the Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology of the Republic of Indonesia based on the Master's Thesis research scheme contract No: 300/F.03.07/2025, 05 Juni 2025.

REFERENCES

Aedo Cancino, P. (2025). Prospective feedback to improve speaking and affective dimensions in higher education students. *HOW*, *32*(1), 29–48. https://doi.org/10.19183/how.32.1.639

Alshammari, E. M. M., & Alomaim, T. I. (2025). Students' preferences of oral corrective feedback: Traditional vs. online learning. *World Journal of English Language*, *15*(6), 77–86. https://doi.org/10.5430/wjel.v15n6p77

Bachiri, Y.-A., Mouncif, H., Bouikhalene, B., & Hamzaoui, R. (2024). Integrating AI-based speech recognition technology to enhance reading assessments within Morocco's TaRL program. *Turkish Online Journal of Distance Education*, *25*(4). https://doi.org/https://doi.org/10.17718/tojde.1335062

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00455-4

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024). I can speak: Improving English pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching*, *18*(5), 443–461. https://doi.org/10.1080/17501229.2024.2315101

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide* (A. Maher, Ed.). London: SAGE Publications Ltd.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (Fifth Edition; H. Salmon, Ed.). USA: SAGE Publications, Inc.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), 379. https://doi.org/10.2307/3588486

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00425-2

Grassini, S. (2023). Development and validation of the AI attitude scale (AIAS-4): A brief measure of general attitude toward artificial intelligence. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1191628

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hew, K. F., & Brush, T. (2007). Integrating technology into K-12 teaching and learning: Current knowledge gaps and recommendations for future research. *Educational Technology Research and Development*, *55*(3), 223–252. https://doi.org/10.1007/s11423-006-9022-5

Keumalasari, B., Iqbal, M., Aulia, F., & Pranata, A. F. (2024). Students' perception of Artificial Intelligence (AI) as English learning tools at MTsN 4 North Aceh. *International Journal of Educational Evaluation and Policy Analysis*, *1*(4), 54–66. https://doi.org/10.62951/ijeepa.v1i4.54

Khoirunnisa, A. W., & Trisanti, N. (2025). Improving pronunciation and speaking confidence with Speechache in 8th grade English learning. *Language Circle: Journal of Language and Literature*, *19*(2s), 281–293. https://doi.org/10.15294/lc.v19i2s.31420

Kim, D. J., Park, H. M., Kim, H., Na, S. H., & Kim, G. J. (2024). Special issue on speech and language AI technologies. *ETRI Journal*, *46*(1), 7–10. https://doi.org/10.4218/etr2.12666

Klefbeck, K. (2024). Implementing assessment protocols in compulsory school for students with intellectual disabilities: Challenges and opportunities for teachers educating students with intellectual disabilities. *International Journal for Lesson & Learning Studies*, *13*(5), 76–91. https://doi.org/10.1108/IJLLS-04-2024-0075

Kratochvil, T., Vaculik, M., & Macak, M. (2023). Gamification tailored for novelty effect in distance learning during COVID-19. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1051227

Liu, C. C., Hwang, G.-J., Yu, P., Tu, Y.-F., & Wang, Y. (2025). Effects of an automated corrective feedback-based peer assessment approach on students' learning achievement, motivation, and self-regulated learning conceptions in foreign language pronunciation. *Educational Technology Research and Development*, *73*(4), 2403–2424. https://doi.org/10.1007/s11423-025-10484-z

Liu, G. L., Darvin, R., & Ma, C. (2024). Exploring AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2024.2310288

Liu, Y., Rahman, F. A., & Zain, F. M. (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

Mingyan, M., Noordin, N., & Razali, A. B. (2025). Improving EFL speaking performance among

undergraduate students with an AI-powered mobile app in after-class assignments: An empirical investigation. *Humanities and Social Sciences Communications*, *12*(1), 370. https://doi.org/10.1057/s41599-025-04688-0

Nageen, S., Sarwar, M., Alam, M., Jabeen, M., & Tayyab, J. (2025). Assessing English-speaking proficiency among secondary school students in Pakistan: A quantitative cross-sectional study. *Language Testing in Asia*, *15*(1), 49. https://doi.org/10.1186/s40468-025-00384-7

Nation, I. S. P., & Hunston, S. (2013). *Learning Vocabulary in Another Language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139858656

Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology*, *55*(4), 1328–1353. https://doi.org/10.1111/bjet.13454

Ngo, T. T. N., Chen, H. H. J., & Lai, K. K. W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, *36*(1), 4–21. https://doi.org/10.1017/S0958344023000113

Nguyen, N. V., Vo, T. T., & Tran, V. D. T. (2025). AI-Driven pronunciation assessment: The impact of SpeechAce on EFL learners' pronunciation competency. *Computer-Assisted Language Learning Electronic Journal*, *26*(3), 84–106. https://doi.org/10.54855/callej.252635

Orosoo, M., Raash, N., Treve, M., Hassan, H. F., Alshammry, N., Ramesh, J. V. N., & Rengarajan, M. (2025). Transforming English language learning: Advanced speech recognition with MLP-LSTM for personalized education. *Alexandria Engineering Journal*, *111*, 21–32. https://doi.org/10.1016/j.aej.2024.10.065

Özdemir, O., & Seçkin, H. (2025). Exploring foreign language anxiety in higher education: Multifaceted insights into causes, impacts, and coping strategies. *Social Sciences & Humanities Open*, *11*, 101364. https://doi.org/10.1016/j.ssaho.2025.101364

Pekrun, R., & Linnenbrink-Garcia, L. (2022). Academic emotions and student engagement. In *Handbook of Research on Student Engagement* (pp. 109–132). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-07853-8_6

Prinos, K., Patwari, N., & Power, C. A. (2024). Speaking of accent: A content analysis of accent misconceptions in ASR research. *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1245–1254. New York, NY, USA: ACM. https://doi.org/10.1145/3630106.3658969

Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1255594

Schlunegger, M. C., Zumstein-Shaha, M., & Palm, R. (2024). Methodologic and data-analysis triangulation in case studies: A scoping review. *Western Journal of Nursing Research*, *46*(8), 611–622. https://doi.org/10.1177/01939459241263011

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Song, J. Y., Rojas, C., & Pycha, A. (2025). Factors modulating perception and production of speech by AI tools: A test case of Amazon Alexa and Polly. *Frontiers in Psychology*, *16*. https://doi.org/10.3389/fpsyg.2025.1520111

SpeechAce. (2025). SpeechAce: speech recognition and scoring platform. Retrieved January 6, 2026, from https://www.speechace.com

SpeechAce API. (2025). SpeechAce API documentation.

Staddon, R. V. (2023). Exploring higher education students' perspectives on factors affecting use, attitudes and confidence with learning technologies. *International Journal of Instruction*, *16*(2), 31–52. https://doi.org/10.29333/iji.2023.1623a

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*. https://doi.org/10.3389/fpsyg.2023.1210187

Sutomo, N. (2024). The implementation of Automated Speech Recognition (ASR) in ELT classroom: A systematic literature review from 2012-2023. *VELES (Voices of English Language Education Society)*, *7*(3), 816–828. https://doi.org/10.29408/veles.v7i3.23978

Tashakkori, A., & Teddlie, C. (2010). *SAGE handbook of mixed methods in social & behavioral research*. California, United States: SAGE Publications, Inc. https://doi.org/10.4135/9781506335193

Toro-Troconis, M., Inzolia, Y., & Ahmad, N. (2023). Exploring attitudes towards embedding education for sustainable development in curriculum design. *International Journal of Higher Education*, *12*(4), 42. https://doi.org/10.5430/ijhe.v12n4p42

Trundle, C., Gardner, J., & Phillips, T. (2024). The ethnographic interview: An interdisciplinary guide for developing an ethnographic disposition in health research. *Qualitative*

Health Research. https://doi.org/10.1177/10497323241241225

Veliche, I.-E., Huang, Z., Ayyat Kochaniyan, V., Peng, F., Kalinli, O., & Seltzer, M. L. (2024). Towards measuring fairness in speech recognition: Fair-Speech dataset. *Interspeech 2024*, 1385–1389. ISCA: ISCA. https://doi.org/10.21437/Interspeech.2024-2273

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1261955

Xu, Y. (2023). Beyond language and culture: A qualitative exploration of mental health barriers for Chinese international students seeking therapy in the U.S. *Graduate Student Journal of Psychology*, *21*. https://doi.org/10.52214/gsjp.v21i.11845

Yousif, A. S. A. (2025). Pronunciation anxiety among EFL learners: Causes, consequences, and classroom interventions. *Journal of Language Teaching and Research*, *16*(6), 1955–1963. https://doi.org/10.17507/jltr.1606.17

Zheng, C., Chen, X., Zhang, H., & Chai, C. S. (2024). Automated versus peer assessment: Effects of learners' English public speaking. *Language Learning & Technology*, *28*(2), 210–228. https://doi.org/10.64152/10125/73578

Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An investigation into Artificial Intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *SAGE Open*, *13*(3). https://doi.org/10.1177/21582440231193818

Zou, B., Liviero, S., Ma, Q., Zhang, W., Du, Y., & Xing, P. (2024). Exploring EFL learners' perceived promise and limitations of using an artificial intelligence speech evaluation system for speaking practice. *System*, *126*, 103497. https://doi.org/10.1016/j.system.2024.103497