THE EFFECT OF USING ELEVENLABS PLATFORM BASED ON AI AUDIO TOWARDS STUDENTS' LISTENING SKILL

Amelia Junisa Putri[1], Yulia Nur Ekawati[2], Endang Sulistianingsih[3]
[123]*English Education Program, Teacher Training and Education Faculty, Universitas Pancasakti Tegal (INDONESIA)*
**Corresponding author: ameliajunisaputri85@gmail.com*

Abstract

This study investigated whether ElevenLabs AI-generated audio significantly affected students' listening skills and described its implementation in listening instruction. A quantitative pre-experimental one-group pre-test–post-test design was employed. The population comprised 35 tenth-grade students at SMK NU AL-YAMAN in the 2025/2026 academic year. Using purposive sampling, 20 students of Class X TKJ 1 were selected as the experimental group, while 15 students of Class X TKR 1 participated in the instrument try-out. Data were collected through a 25-item listening test and an observation checklist completed by the English teacher. The participants completed a pre-test, received six treatment sessions using recount-text audio generated through ElevenLabs, and completed a post-test. The test scores were analyzed with SPSS using the Shapiro–Wilk normality test and a paired-samples t-test, while the observation data were summarized descriptively and supported by thematic coding. The mean score increased from 12.75 on the pre-test to 21.65 on the post-test. The paired-samples t-test yielded $t(19) = -14.175$, $p < .001$, indicating a statistically significant improvement after treatment. Ten of twelve observation indicators were achieved, showing that the audio was implemented through pre-listening, while-listening, and post-listening activities and that students generally demonstrated attention, participation, and interest. ElevenLabs AI-generated audio therefore supported a significant within-group improvement in listening performance. Nevertheless, because the design did not include a control group, the result should not be interpreted as conclusive evidence that ElevenLabs alone caused the improvement.

Keywords: ElevenLabs; AI-generated audio; listening skills; text-to-speech;

1 INTRODUCTION

Listening is an essential language skill because it enables learners to receive, process, and interpret spoken communication. It also supports the development of speaking, reading, and writing by exposing learners to pronunciation, vocabulary, and language use. Listening is not a passive activity; learners actively construct meaning from the sounds they hear (Brown, 2004). Harmer (2007) similarly emphasizes that listening is fundamental to language acquisition because it provides learners with meaningful spoken input.

In English as a Foreign Language (EFL) classrooms, listening remains difficult for many students. Their difficulties may result from limited exposure to spoken English, unfamiliar accents, natural speech rates, unclear pronunciation, and learning media that do not match their proficiency. When listening instruction consists only of playing a recording and answering questions, learners can lose interest and receive insufficient guidance for understanding the text.

Preliminary observation at SMK NU AL-YAMAN showed that listening activities were conducted infrequently because instruction mainly emphasized written exercises and theoretical explanations. Tenth-grade students, generally at the A1–A2 proficiency level, had limited opportunities to practise listening and often required repeated exposure to recognize vocabulary, identify main ideas, and locate specific information. The limited use of varied audio also reduced their familiarity with spoken English.

Artificial intelligence provides teachers with new ways to prepare adaptable listening materials. Text-to-speech technology converts written text into spoken language, while recent AI systems can produce clearer and more natural-sounding speech. ElevenLabs is an AI-based text-to-speech platform that allows users to select voices and adjust settings related to stability, similarity, style, and speaking speed (ElevenLabs, n.d.). Teachers can therefore prepare audio that corresponds to learners' proficiency and lesson objectives.

AI-generated audio may be pedagogically relevant to Krashen's Input Hypothesis. Krashen (1985) argues that language acquisition occurs when learners receive comprehensible input that is slightly beyond their current competence but remains understandable. Clear pronunciation, controlled speed, repeated listening, vocabulary preparation, and teacher guidance can make spoken input more accessible without removing the cognitive challenge of listening.

Previous studies indicate that AI and synthetic speech can support language learning. Andriani et al. (2020) reported an effect of synthetic-speech materials on listening comprehension. Akmal et al. (2024) found that text-to-speech resources supported EFL listening through human-like audio input. Mulyanto et al. (2024) likewise reported improvement associated with AI-based text-to-speech technology. Studies of ElevenLabs have also found positive responses to its natural audio, although some focused on speaking or on Arabic rather than English listening (Budianto et al., 2024; Sunarko et al., 2026).

The pedagogical value of AI-generated audio lies partly in its adaptability. Unlike a fixed textbook recording, a teacher-generated text-to-speech recording can be revised when vocabulary, sentence length, pronunciation, or speaking rate is unsuitable. Baxramova (2025) describes AI-generated listening materials as dynamic and customizable, although the study also warns that synthetic voices may not fully reproduce the emotional depth and human nuance of natural speech. This limitation means that AI audio should be treated as a carefully selected instructional resource rather than an automatic replacement for human voices.

Students' responses are also relevant because listening performance can be affected by motivation, anxiety, and willingness to participate. Huda et al. (2025) found that AI-enhanced listening tools helped learners understand spoken English and increased their motivation. Loebis (2025) similarly emphasized that AI-based applications can provide personalized and interactive learning experiences. These findings suggest that the classroom value of AI audio is not limited to sound quality: it may also support repeated practice and a learning environment in which students feel more willing to attempt difficult listening tasks.

However, the effectiveness of technology depends on how it is integrated into instruction. Listening development requires more than exposure to an audio file. Learners need preparation before listening, a clear purpose while listening, and feedback after listening. Studies of project-based and interactive listening instruction have likewise shown that active, meaningful classroom procedures contribute to improvement (Ekawati, 2017; Triana et al., 2023). Accordingly, the present study combined ElevenLabs-generated audio with staged activities rather than evaluating the platform as a stand-alone application.

ElevenLabs produces speech through text-to-speech technology. In the platform's Text-to-Speech feature, users enter a prepared script, choose a voice and model, and adjust available controls such as speed and stability. For listening instruction, these controls allow the teacher to select a voice with clear articulation and generate the same text again after correcting problematic pronunciation, punctuation, or pauses. The resulting file can be replayed consistently, which is useful when learners need repeated exposure to the same spoken input.

AI audio in this study refers specifically to spoken output generated from written English text. It does not refer to automatic assessment, speech recognition, voice cloning, or conversational tutoring. This boundary is important because different AI functions involve different pedagogical and ethical questions. The study examined ElevenLabs only as a means of producing listening material, while the researcher and English teacher remained responsible for selecting the text, designing the tasks, managing the classroom, and evaluating students' answers.

Two previous studies are especially close to the present research. Budianto et al. (2024) examined students' responses to ElevenLabs in English-speaking activities and reported support for pronunciation, fluency, and participation, but the focus was speaking rather than listening. Sunarko et al. (2026) investigated ElevenLabs-generated audio for Arabic listening in a pesantren and found stronger improvement for the experimental group as well as positive responses to audio clarity. Their context, target language, participants, and design differed from those of the present research.

Research specifically examining ElevenLabs-generated English audio for vocational high-school listening instruction remains limited. This study therefore addressed two questions: (1) Is there a significant difference between students' listening scores before and after instruction using ElevenLabs AI-generated audio? and (2) How is ElevenLabs AI-generated audio implemented in teaching listening? The study aimed to measure the within-group change in listening performance and describe the instructional process.

2 METHODOLOGY

2.1 Research Approach and Design

This study used a quantitative approach and a pre-experimental one-group pre-test–post-test design. One intact class completed a pre-test, received the treatment, and completed a post-test. The design was selected to measure the change in the same participants' listening scores after the use of ElevenLabs-generated audio. Because no comparison group was included, the statistical result represents within-group improvement and must be interpreted cautiously when discussing causality.

The study was conducted at SMK NU AL-YAMAN from May 11 to May 26, 2026. It comprised eight meetings: one pre-test meeting, six treatment sessions, and one post-test meeting.

The one-group design can identify whether the participating students' scores changed after treatment, but it cannot control all alternative explanations. Repeated testing, ordinary classroom development, teacher assistance, or other events occurring during the research period may also contribute to change. For this reason, the analysis used careful language: a significant paired-samples result demonstrates a difference within the group, whereas stronger causal attribution would require a control or comparison group.

2.2 Population, Sample, and Sampling

The population comprised 35 tenth-grade students in the 2025/2026 academic year. Purposive sampling was used. Class X TKJ 1, consisting of 20 students, was selected as the experimental group because it met the study's practical and instructional requirements, including limited previous listening practice, availability, and permission from the school and English teacher. Fifteen students of Class X TKR 1 participated only in the instrument try-out.

2.3 Variables and Instruments

The independent variable was the use of ElevenLabs platform-based AI-generated audio in listening instruction. The dependent variable was students' listening skill, operationalized through scores on a multiple-choice listening test. The assessed areas included vocabulary recognition, general information, main ideas, specific information, sequence of events, and overall comprehension of spoken recount texts.

The initial try-out contained 48 multiple-choice items and was administered to Class X TKR 1 on May 9, 2026. Item analysis showed that 25 items were valid and 23 were invalid. The reliability analysis yielded Cronbach's alpha = .876, indicating high internal consistency; consequently, the 25 valid items were retained for the pre-test and post-test.

An observation checklist completed by the Class X TKJ 1 English teacher was used as supporting descriptive evidence. It contained 12 Yes/No indicators covering preparation, while-listening activities, task completion, interaction, comprehension, feedback, discussion, and student interest. Additional notes documented classroom conditions and responses.

The listening test was administered under comparable conditions before and after the treatment. Each correct response contributed one point, so the maximum raw score was 25. The observation checklist did not function as a second achievement test. Instead, it documented whether the planned stages were implemented and recorded visible classroom responses. Thus, test scores answered the question about change in listening performance, while observation evidence answered the question about implementation.

2.4 Treatment and Data Collection

Before each treatment, the researcher prepared a recount text suited to the students' proficiency and lesson focus. The text was checked for spelling, punctuation, sentence structure, names, and abbreviations before being entered into ElevenLabs. A suitable voice and settings were selected, and the generated audio was reviewed for pronunciation, intonation, pauses, clarity, and speaking rate. Unclear sections were revised and generated again.

Table 1. Treatment focus and classroom activities

Meeting	Listening focus	Main student activity
1	Vocabulary recognition	Identify words heard in a spoken recount text.
2	General information	Identify the topic and select an appropriate title.
3	Main idea	Determine the central idea through listening questions.
4	Specific information	Identify people, places, times, and events.
5	Sequence of events	Arrange events in the correct chronological order.
6	Overall comprehension	Summarize the recount text based on the audio.

Every treatment followed three stages. In pre-listening, the researcher introduced the topic, activated background knowledge, and explained essential vocabulary. In while-listening, students listened to ElevenLabs-generated audio and completed tasks aligned with the focus of the meeting; the audio was replayed when necessary. In post-listening, answers were discussed, difficult sections were clarified, and feedback was provided.

The pre-test was administered on May 11, 2026, the six treatments took place from May 12 to May 25, and the post-test was administered on May 26. The same 25-item format was used to compare listening performance before and after treatment.

The six lesson focuses were arranged from relatively local processing to broader comprehension. Vocabulary recognition introduced students to identifying words in connected speech. General information and main-idea activities then required them to construct an overall understanding. Specific-information and sequencing tasks directed attention to details and relationships among events. The final summarizing activity asked students to combine information from the recording into a concise representation of the recount text. This progression aligned the use of technology with explicit listening objectives.

2.5 Data Analysis

The quantitative data were analyzed using SPSS. Descriptive statistics summarized the pre-test and post-test means, standard deviations, and standard errors. Because the same students were measured twice, the assumption of normality was assessed on the difference scores using the Shapiro–Wilk test. A significance value above .05 indicated that the normality assumption was satisfied.

A paired-samples t-test compared the pre-test and post-test means at $\alpha = .05$. When $p < .05$, the null hypothesis of no difference was rejected. Observation checklist responses were summarized descriptively, while the observer's notes were coded and grouped to describe implementation. These qualitative observations supported the quantitative findings but did not change the study into a mixed-methods design.

3 RESULTS

3.1 Instrument Quality and Normality

Of the 48 try-out items, 25 met the validity criterion and were used in the final instrument. Cronbach's alpha was .876, exceeding the study's .60 threshold and indicating that the retained listening test was reliable.

Table 2. Shapiro–Wilk normality test for difference scores

Variable	Statistic	df	Sig.
Pre-test–post-test difference	.948	20	.366

The Shapiro–Wilk significance value was .366. Because .366 > .05, the pre-test–post-test difference scores were normally distributed, satisfying the normality assumption for the paired-samples t-test.

3.2 Descriptive Statistics and Hypothesis Test

Table 3. Paired-samples descriptive statistics

Test	Mean	N	Std. deviation	Std. error mean
Pre-test	12.75	20	4.102	.917
Post-test	21.65	20	2.254	.504

The mean increased by 8.90 points, from 12.75 before treatment to 21.65 after treatment. The post-test scores also had a smaller standard deviation, suggesting that students' performance became more concentrated after the six treatment sessions.

Table 4. Paired-samples t-test

Mean difference	SD	SE	95% CI lower	95% CI upper	t	df	p
−8.900	2.808	.628	−10.214	−7.586	−14.175	19	< .001

The test produced $t(19) = -14.175$, $p < .001$. The 95% confidence interval for the pre-test minus post-test difference ranged from −10.214 to −7.586 and did not include zero. The null hypothesis was therefore rejected. Students' listening scores were significantly higher after the ElevenLabs-supported instruction. This result establishes a statistically significant within-group change; the one-group design does not establish that ElevenLabs was the sole cause.

3.3 Implementation Findings

Table 5. Summary of observation checklist results

Stage/aspect	Achieved indicators	Main finding
Pre-listening	2 of 2	Topic introduction and vocabulary explanation were completed.
While-listening	3 of 3	Audio was played clearly; students generally listened attentively; replay was provided when needed.
Task and interaction	2 of 2	Students completed questions and followed instructions.
Comprehension	0 of 2 fully achieved	Some students still had difficulty identifying main ideas and specific information.
Post-listening	2 of 2	The teacher provided feedback and students joined the discussion.
Motivation	1 of 1	Students showed interest and enthusiasm.

Ten of the twelve indicators were achieved. The teacher introduced each topic, explained relevant vocabulary, played and replayed the ElevenLabs audio, guided task completion, and provided feedback. Students generally listened attentively, responded to instructions, completed the tasks, participated in discussion, and showed interest in the audio-based activities.

Two comprehension indicators were not fully achieved: not all students consistently identified the main idea and specific information. Thus, the technology supported engagement and structured input, but it did not remove every listening difficulty. Learners continued to need repetition, vocabulary support, strategy instruction, and teacher clarification.

The observer's notes distinguished interest from measurable achievement. Students' interest was visible in their attention, responses to instructions, task completion, and willingness to join discussion. These behaviours indicate positive classroom engagement, but they were not measured with a separate pre-test and post-test scale. The study therefore does not claim a statistically significant increase in motivation or engagement; it reports only the positive patterns observed during implementation.

The thematic organization used seven categories derived directly from the checklist: pre-listening, while-listening, task, interaction, comprehension, post-listening, and motivation. The while-listening category contained the largest number of indicators because it covered playing the audio, attentive listening, and replay. Its visual prominence in the NVivo hierarchy chart represented the amount of coded material, not a higher level of success. Likewise, the colours in the chart were used only to distinguish categories and did not indicate achievement or positive and negative results.

3.4 Implementation Across the Six Treatments

During the first treatment, students focused on recognizing vocabulary in a spoken recount text. The teacher introduced the context and key words, then played the recording while students identified vocabulary from the audio. This meeting familiarized students with the AI-generated voice and established the routine of listening, completing a task, replaying difficult sections, and discussing answers.

The second and third treatments moved from general understanding to main ideas. In the second meeting, students listened for the overall topic and selected a suitable title. In the third, they answered questions designed to identify the main idea. The observer noted active participation, although the main-idea indicator was not fully achieved by every student. This showed that general familiarity with the topic did not always lead automatically to accurate identification of the central message.

The fourth treatment focused on specific information such as people, places, time, and events. Some students still missed details even after replay, which explains why the related observation indicator was not fully achieved. The fifth treatment required students to arrange events chronologically, connecting listening comprehension with the structure of recount text. In the sixth treatment, students summarized the text, drawing together vocabulary, general meaning, main ideas, details, and event sequence.

Across the six meetings, the researcher checked the generated recordings before classroom use. Punctuation and sentence structure were reviewed because they influenced pauses and intonation, while voice and speed settings were adjusted to maintain clarity for A1–A2 learners. Replay was provided when students missed information, but the teacher continued to require active responses rather than explaining the complete text in advance.

4 DISCUSSION

4.1 Implementation of ElevenLabs in Listening Instruction

The implementation followed a systematic progression from material preparation to pre-listening, while-listening, and post-listening. ElevenLabs was not used as an isolated technological novelty; it functioned as the principal audio source within planned classroom tasks. The researcher first aligned recount texts with the students' level, generated and checked the recordings, and then connected each recording to a specific listening objective.

The pre-listening activities helped learners activate background knowledge and recognize important vocabulary before processing the recording. During listening, repeated playback gave students additional opportunities to segment speech, identify words, and verify information. Post-listening discussion and feedback allowed the teacher to correct misunderstandings and connect the students' answers with the content of the audio. This organization corresponds with the view that text-to-speech technology is most useful when embedded in guided instruction rather than merely played to students (Akmal et al., 2024; Hazaymeh et al., 2025).

The observation findings indicated positive participation and interest. Natural-sounding and consistent speech appeared to attract attention, while replay and teacher support reduced pressure during difficult sections. Similar positive responses to AI-supported language learning have been reported by Budianto et al. (2024) and Huda et al. (2025). Nevertheless, the difficulties with main ideas and specific information demonstrate that audio quality alone cannot guarantee comprehension. Students at A1–A2 level still require explicit strategies, repeated exposure, and graduated task difficulty.

The findings also clarify the division of responsibility between technology and teacher. ElevenLabs generated the spoken output, but pedagogical quality depended on the researcher's script selection, editing, audio checking, sequencing of tasks, and feedback. When a generated word or pause sounded unsuitable, the text or settings had to be revised. This supports the position of Guzmán Alvarado and Naranjo Andrade (2025) that AI should remain an instructional aid rather than replacing the teacher. In this study, the platform increased flexibility in preparing audio, while the teacher preserved control over content and interaction.

4.2 Effect on Students' Listening Skills

The significant increase from 12.75 to 21.65 indicates that the students performed better after six sessions of listening instruction using ElevenLabs-generated audio. The statistical result, $t(19) = -14.175$, $p < .001$, is consistent with earlier research reporting that synthetic speech and AI-based text-to-speech can support listening comprehension (Andriani et al., 2020; Akmal et al., 2024; Mulyanto et al., 2024).

Several instructional features may explain this change. ElevenLabs provided clear pronunciation, natural intonation, and consistent delivery. The materials were adjusted to the students' proficiency, and replay was permitted when clarification was needed. In Krashen's (1985) terms, these conditions helped make the listening input comprehensible while retaining an appropriate level of challenge. Brown's (2004) account of listening as active meaning construction is also reflected in the tasks: students identified vocabulary, inferred general and main ideas, located details, ordered events, and summarized texts rather than listening passively.

The quantitative and observation findings should be read together. Improved scores occurred alongside attention, task completion, discussion, and interest, suggesting that both the accessibility of the input and the organization of the classroom activities supported learning. However, the study used only one class and did not compare ElevenLabs with another audio source or instructional method. Maturation, repeated testing, teacher support, and other classroom factors cannot be separated from the treatment. Accordingly, the evidence supports a significant positive change after the intervention, not an exclusive causal claim.

The size of the score change is educationally meaningful within the scale used. The post-test mean of 21.65 out of 25 corresponds to substantially more correct responses than the pre-test mean of 12.75. At the same time, the smaller post-test standard deviation (2.254 versus 4.102) indicates less dispersion among students after treatment. These descriptive patterns are consistent with the inferential result, although they should not be interpreted independently of the sample size and research design.

The negative t-value does not indicate a negative effect. SPSS calculated the paired difference as pre-test minus post-test; because post-test scores were higher, the mean difference was −8.900 and the t-value was negative. The direction would become positive if the subtraction order were reversed, while the magnitude and significance of the test would remain the same. The substantive conclusion is therefore improvement from pre-test to post-test.

The remaining comprehension difficulties are also important. Identifying main ideas requires learners to integrate information across a text, whereas locating specific information requires accurate perception of details in continuous speech. Clear AI-generated pronunciation can make the signal more accessible, but students must still apply vocabulary knowledge, attention, memory, and listening strategies. The two unachieved indicators therefore do not contradict the significant test result; they show that improvement did not mean complete mastery for every learner.

Taken together, the results support the use of ElevenLabs as one component of structured listening instruction for vocational high-school students. The most defensible interpretation is that the participating class improved significantly during an intervention in which ElevenLabs-generated audio, repeated listening, staged tasks, discussion, and teacher feedback operated together. Further research is needed to isolate the contribution of the audio platform from the contribution of the broader instructional procedure.

4.3 Pedagogical Implications and Study Limitations

The findings have practical implications for teachers who wish to prepare listening materials that are more closely aligned with learners' needs. A teacher can revise a script, generate the recording, check its pronunciation and pacing, and create another version when the output is unsuitable. This process provides more flexibility than relying only on a fixed textbook recording. Nevertheless, customization should serve a defined learning objective. Voice selection or speed adjustment is useful only when it helps learners practise a specific aspect of listening rather than making every recording artificially easy.

The treatment sequence also suggests that teachers should move gradually from recognition to comprehension. Beginning learners may first need to notice vocabulary and the general topic before they can determine a main idea, locate details, reconstruct the sequence of events, and summarize the text. Replay can support this progression, but repeated listening should have a clear purpose. For example, the first playback may focus on general meaning, the second on key information, and the final playback on checking answers. Feedback should then explain both the correct answer and the listening cues that support it.

The study also illustrates that positive classroom participation does not necessarily mean that all comprehension objectives have been mastered. Students appeared interested and completed the assigned activities, yet several still experienced difficulty with main ideas and specific information. Teachers should therefore avoid evaluating an AI tool only through enthusiasm or task completion. Achievement evidence and observation evidence answer different questions and should be interpreted together.

Several limitations restrict the generalization of the findings. The sample consisted of only 20 students from one vocational high-school class selected purposively. The treatment period was limited to six sessions, and the same group completed both measurements without a control group. Consequently, the result cannot determine whether ElevenLabs is more effective than teacher-recorded audio, textbook audio, or another instructional medium. The observation checklist was completed by the English teacher and provided useful classroom evidence, but motivation and engagement were not measured through separate validated scales.

Future research should use a quasi-experimental or true experimental design with comparable groups, a larger and more diverse sample, and a longer intervention. Researchers could compare different audio sources while keeping texts, tasks, duration, and teacher support constant. They could also examine whether the effects differ across proficiency levels, listening-text genres, voice settings, or repeated-listening procedures. Such studies would provide stronger evidence about which benefits arise from ElevenLabs itself and which arise from the surrounding instructional design.

5 CONCLUSIONS

This study found a statistically significant improvement in the listening scores of 20 students of Class X TKJ 1 after six treatment sessions using ElevenLabs AI-generated audio. The mean increased from 12.75 to 21.65, and the paired-samples t-test yielded $t(19) = -14.175$, $p < .001$. Therefore, the null hypothesis of no pre-test–post-test difference was rejected.

ElevenLabs was implemented through preparation of level-appropriate recount texts, generation and review of the audio, pre-listening topic and vocabulary activities, guided while-listening tasks with replay, and post-listening discussion and feedback. Ten of twelve observation indicators were achieved, and students generally demonstrated attention, participation, interaction, and interest. Some students nevertheless continued to struggle with main ideas and specific information.

ElevenLabs can be considered a useful instructional medium for providing clear, adjustable, and repeatable spoken input. Teachers should combine it with vocabulary preparation, structured tasks,

listening strategies, repetition, and feedback. Future studies should involve larger samples, a control or comparison group, longer treatment periods, and, where appropriate, comparisons between AI-generated and human-recorded audio to strengthen causal conclusions.

ACKNOWLEDGEMENTS

The authors thank the principal, English teacher, and students of SMK NU AL-YAMAN for their permission, participation, and support during the study.

REFERENCES

Akmal, A., Perkasa, A. B., & Aziz, R. A. (2024). Artificial intelligence for EFL students' listening skills. Celt: A Journal of Culture, English Language Teaching & Literature, 24(2), 229–250. https://doi.org/10.24167/celt.v24i2

Andriani, T. N., Herawati, Y. W., & Sulistyo, T. (2020). Text-to-speech application for foreign language learners' listening comprehension in Indonesia. In Proceedings of ICONQUHAS 2018. https://doi.org/10.4108/eai.2-10-2018.2295544

Brown, H. D. (2004). Language assessment: Principles and classroom practices. Pearson Education.

Budianto, L., Zulaiha, S., Ar-Rahma, A. P., & Addiyaullami, N. D. (2024). Students' responses on ElevenLabs AI for English-speaking learning. In Proceeding International Seminar on Islamic Education and Peace (Vol. 4).

Baxramova, M. M. (2025). Effectiveness of AI-generated listening materials vs. textbook audio in EFL learning. Ta'lim Innovatsiyasi va Integratsiyasi, 49(1), 53–55.

Creswell, J. W., & Creswell, J. D. (2017). Research design: Qualitative, quantitative, and mixed methods approaches. Sage.

Ekawati, Y. N. (2017). Improving English students' listening skill through project-based learning model. Cakrawala: Jurnal Pendidikan, 11(2), 135–143.

ElevenLabs. (n.d.). Text to speech: Product guide. Retrieved July 23, 2026, from https://elevenlabs.io/docs/eleven-creative/playground/text-to-speech

Guzmán Alvarado, M. V., & Naranjo Andrade, S. S. (2025). The impact of artificial intelligence on English language learning: A systematic review of tools, methods, and outcomes in language skills. RUNAS Journal of Education and Culture, 6(12), e250287. https://doi.org/10.46652/runas.v6i12.287

Harmer, J. (2007). How to teach English. Pearson Longman.

Hazaymeh, W. A., Alshaikhi, T., Abdul Wahab, M. O., & Khasawneh, M. A. (2025). Evaluating AI applications for enhancing listening comprehension among EFL students in real-world scenarios. World Journal of English Language, 15(6), 209–220. https://doi.org/10.5430/wjel.v15n6p209

Huda, M. C., Roistika, N., & Triani, C. N. (2025). Exploring the use of AI-enhanced listening comprehension tools by EFL students. Teaching English and Language Learning English Journal, 5(3), 290–300.

Krashen, S. D. (1985). The input hypothesis: Issues and implications. Longman.

Loebis, I. A. (2025). Effectiveness of AI-based English language learning apps in improving listening skills. Lingeduca: Journal of Language and Education Studies, 4(1), 9–16. https://doi.org/10.70177/lingeduca.v4i1.2123

Mulyanto, D., Wahyudi, M., Ridho, A. M. A., & Zaki, M. (2024). Utilization of artificial intelligence with text-to-speech technology based on natural language processing to enhance Arabic listening skills. Alsinatuna: Journal of Arabic Linguistics and Education, 10(1), 44–58.

Sunarko, A., Solehudin, M., & Sakinah, N. (2026). ElevenLabs text-to-speech and Arabic listening achievement in pesantren. An Nabighoh, 28(1), 1–20. https://doi.org/10.32332/an-nabighoh.v28i1.1-20

Triana, D. A., Ekawati, Y. N., & Santoso, M. E. (2023). Implementation of the project-based learning model in teaching macro listening skills. In Seminar Nasional Literasi Pedagogi (SRADA) III (pp. 153–160). Universitas Pancasakti Tegal.