**Issues in Applied Linguistics & Language Teaching**
Volume 07, Issue 02, 2025, pp. 374-383
E-ISSN: 2597-9825
Open Access: https://doi.org/10.37253/iallteach.v7i2.11675

# Which Voice Leads Learning? A Qualitative Comparison of AI-Generated Input and Teacher Talk in Listening Instruction

**Nurhayati Ganinda[1*], Rini Lindawati[2], Asih Andriyati Mardliyah[3]**
[1,2] English Language Teaching Education Department, University Islam Majapahit, Mojokerto, Indonesia
E-mail addresses: nurhayatiganinda@unim.ac id

**ARTICLE INFO**

**Article history:**
Received November 30, 2025
Revised December 09, 2025
Accepted December 11, 2025
Available online December 17, 2025

***Kata Kunci :***
*Pembelajaran Menyimak, Tuturan Guru, Kecerdasan Buatan*

**Keywords:**
AI-generated Input, Teacher Talk, Listening Instruction, Artificial Intelligence

**ABSTRAK**

*Penelitian ini bertujuan untuk mengkaji persepsi pemelajar Bahasa Inggris sebagai Bahasa Asing (EFL) terhadap suara hasil generasi kecerdasan buatan (AI-generated voices) dan tuturan guru sebagai sumber masukan menyimak, serta menganalisis pengaruh keduanya terhadap pemahaman dan keterlibatan belajar di kelas. Penelitian ini menggunakan pendekatan kualitatif dengan pengumpulan data melalui observasi kelas, wawancara semi-terstruktur, dan jurnal reflektif mahasiswa EFL tahun pertama. Data dianalisis secara tematik. Hasil penelitian menunjukkan bahwa suara hasil generasi AI dipersepsikan memiliki kejelasan dan konsistensi yang tinggi sehingga mendukung latihan menyimak secara mandiri, membantu pengenalan kosakata, dan mengurangi kecemasan belajar. Namun, masukan berbasis AI dinilai kurang memberikan interaksi dan keterlibatan emosional. Sebaliknya, tuturan guru dipersepsikan lebih autentik dan kontekstual karena memungkinkan klarifikasi langsung, scaffolding yang adaptif, serta dukungan afektif, meskipun variasi tuturan alami menuntut upaya kognitif yang lebih besar dari pemelajar. Temuan ini menegaskan bahwa kedua sumber masukan bersifat saling melengkapi dan perlu diintegrasikan secara seimbang dalam pembelajaran menyimak EFL.*

**ABSTRACT**

This study explores the comparative influence of artificial intelligence (AI)-generated input and teacher talk as listening materials for first-year university students learning English as a foreign language (EFL). As AI technologies become more prevalent in education, tools such as AI-generated speech offer new modes of listening input, while teacher talk continues to serve as an authentic and interactive source of language exposure. Employing a qualitative research design, this study involves classroom observations, semi-structured interviews, and reflective journals to examine students' comprehension strategies, engagement, perceptions, benefits, and challenges with both types of listening input. The participants consist of university EFL students. The data will be analysed thematically to uncover patterns in learner responses of perceived benefits, challenges, comprehension strategies, and engagement associated with each input type. Findings indicate that learners perceived AI-generated voices as clear, consistent, and supportive of self-paced listening practice, facilitating vocabulary recognition and reducing listening anxiety. However, AI input was also described as limited in interaction and emotional presence. Teacher talk, in contrast, was experienced as more authentic and socially meaningful, providing real-time clarification, contextual scaffolding, and affective support, although natural speech variation sometimes increased cognitive demand. In terms of comprehension and engagement, AI-generated input promoted focused individual effort and autonomous learning, while teacher talk encouraged active participation, interaction, and sustained motivation. Overall, the findings suggest that AI-generated voices and teacher talk function as complementary listening sources, highlighting the importance of balanced integration to support effective EFL listening instruction.

## 1. INTRODUCTION

Listening has long been recognized as a central component in second language acquisition, providing learners with comprehensible input that facilitates vocabulary development, grammatical awareness, and pragmatic competence (Vandergrift, L., & Goh, 2012). In English as a Foreign Language (EFL) contexts, listening instruction becomes even more crucial, as learners often have limited exposure to authentic

language use outside the classroom, relying heavily on instructional materials and classroom discourse as their primary source of input (Field, 2019). Without sufficient listening practice that mirrors real-world communication, learners may struggle to develop the ability to process spoken English in authentic contexts, making classroom listening activities a vital bridge between controlled learning and communicative use.

Traditionally, teacher talk has been one of the most significant sources of input in EFL classrooms. Beyond delivering content, teacher talk functions as a dynamic pedagogical tool that adapts to learners' needs, offering linguistic models that reflect real-time language use while also incorporating scaffolding, interactional adjustments, and contextual cues to support comprehension (Walsh, 2011). Unlike pre-recorded or scripted materials, teacher talk is responsive and situated, allowing learners to negotiate meaning, seek clarification, and experience language within a meaningful social interaction—elements that are central to effective listening development. Furthermore, teacher talk carries an affective dimension: through tone, emphasis, and interpersonal engagement, teachers can foster a supportive classroom atmosphere that reduces learner anxiety and promotes active participation (Gibbons, 2015). In this way, teacher talk does not merely transmit language; it embodies the immediacy, authenticity, and dialogic nature of communication that helps bridge the gap between classroom instruction and real-world language use. Nevertheless, the scope of teacher talk is often limited by factors such as the teacher's own linguistic repertoire, accent, and time constraints in the classroom, which may restrict learners' exposure to the diversity of speech patterns they are likely to encounter beyond the academic setting (Richards, 2017).

In response to these limitations, recent technological advances have introduced alternative sources of listening input, most notably through artificial intelligence (AI). AI-generated voices, driven by natural language processing and text-to-speech technologies, have become increasingly sophisticated, producing speech with near-human fluency, varied accents, and natural prosody (Tan et al., 2021). These tools allow learners to access a broader range of listening materials beyond what teacher talk alone can provide, offering repeated exposure, flexible pacing, and opportunities for individualized practice (Ní Chiaráin & Chasaide, 2020). Unlike traditional audio materials, which are often fixed and inflexible, AI-generated input can be tailored to learners' proficiency levels, learning goals, and preferred modes of practice, thereby enhancing both accessibility and personalization (Jian, 2023). In addition, the scalability of AI-generated resources makes them particularly valuable in large or resource-limited classrooms, where teachers may not always have the time or capacity to provide varied listening input for every learner. Moreover, AI-generated input can be designed to simulate authentic communicative contexts, giving learners exposure to diverse voices and registers that more closely resemble real-world listening situations while simultaneously allowing controlled conditions for practice and reflection. Nevertheless, despite these advantages, AI-generated voices often lack the spontaneity, affective warmth, and adaptive responsiveness that characterize human teacher talk, raising questions about how effectively such input can replicate the richness of interactive classroom communication (Zhu & Wang, 2024).

**Teacher Talk as Listening Source**

Teacher talk remains a principal source of comprehensible input in EFL classrooms, not only because it delivers linguistic models but also because it creates interactional space where meaning is jointly negotiated. Recent work on Classroom Interactional Competence (CIC) highlights how teachers manage turn-taking, repair, and questioning strategies to provide learners with opportunities for listening development (Sert, 2015; Walsh, 2011). Such adaptive talk supports learner engagement by shaping input in ways that align with student needs, thus fostering deeper comprehension in real time.

From pedagogic perspective, teacher talk functions as scaffolding, shaping learners' access to academic language and tasks. Through modelling, recasting, and contextualizing content, teachers help learners reduce cognitive load and process new language more effectively (Gibbons, 2015). Teacher talk is therefore not just delivery but an instructional design feature that integrates linguistic and conceptual support.

A growing body of evidence links the quality of teacher language in terms of linguistic richness, clarity, and interactional supports with student achievement. For instance, (Tao & Chen, 2024) found that teacher language quality significantly predicts student learning outcomes, underscoring that it is not the sheer amount of talk but it's interaction that matters most. Discourse-analytic studies further demonstrate the role of teacher discourse moves, such as probing, recontextualizing, and using metadiscourse to signal emphasis or transitions. These moves guide learners' attention, making oral input more navigable and supporting listening comprehension in cognitively demanding settings (Vasheghani Farahani, 2021).

Finally, research cautions that intelligibility variables such as accent, speech rate, and prosody affect the accessibility of teacher talk. (Saito et al., 2025) show that learners' comprehension varies with accent familiarity and speech delivery, highlighting the need for calibrated input and exposure to accent diversity.

These findings reinforce the dual role of teacher talk as both a facilitator and potential barrier to listening development depending on how it is enacted.

In sum, teacher talk is a dynamic and adaptive listening source that provides real-time scaffolding and responsive feedback, distinguishing it from scripted or pre-recorded materials. Its effectiveness depends not on quantity but on the quality and accessibility of the input it provides, making it a crucial benchmark for evaluating alternative listening sources such as AI-generated voices.

**AI-Generated Voices as Listening Source**

With rapid advances in artificial intelligence (AI), listening instruction has begun to incorporate AI-generated voices as an alternative to traditional input sources. Text-to-speech (TTS) systems and natural language processing technologies have improved dramatically in fluency, naturalness, and prosodic features, making synthetic speech increasingly indistinguishable from human voices (Shen et al., 2018; Tan et al., 2021). These developments position AI-generated input as a valuable tool for expanding the range and accessibility of listening materials in EFL contexts.

One of the most significant affordances of AI-generated voices is flexibility. Learners can control playback, repeat input, and adjust speed, which supports individualized pacing and repeated exposure to target language (Ní Chiaráin & Chasaide, 2020). This feature aligns with research on listening comprehension, which emphasizes the importance of multiple exposures and controlled practice for deep processing of language (Vandergrift, L., & Goh, 2012).

Another strength lies in the diversity of linguistic exposure. AI-generated speech can simulate different accents, registers, and contexts that might not be readily available in teacher talk or textbook (Jian, 2023) note that AI-generated voices can provide learners with varied and authentic-sounding input, helping them prepare for real-world listening situations that involve multiple English varieties. In this way, AI-generated voices may reduce the risk of over-reliance on a single model of speech, broadening learners' listening repertoires.

Additionally, AI voices offer opportunities for autonomous learning. Learners can engage with AI-generated materials outside classroom settings, thus increasing overall listening time and supporting learner agency (Zhu & Wang, 2024). This self-directed practice is particularly valuable in EFL contexts, where authentic input is often scarce.

However, while AI-generated input provides consistency and accessibility, it lacks the interactive dimension of teacher talk. Unlike teachers, AI voices cannot negotiate meaning, adjust discourse in response to learners, or provide socio-pragmatic cues during interaction (Porubay et al., 2024). Therefore, while AI-generated voices expand exposure and accessibility, they cannot fully replace the responsive and relational aspects of classroom interaction.

In sum, AI-generated voices function as a complementary listening source. They extend learners' access to varied and repeatable input, support individualized pacing, and foster autonomy. Yet, their limitations in interactional adaptability underscore the importance of balancing AI input with teacher talk to maximize listening development.

Although AI technologies are increasingly incorporated into language education, empirical evidence comparing AI-generated voices and teacher talk as sources of listening input in authentic classroom settings remains limited. Existing studies have primarily focused on digital listening resources such as podcasts, audiobooks, and text-to-speech applications (Ní Chiaráin & Chasaide, 2020), leaving learners' comparative responses to AI-generated input and teacher talk largely unexplored. As a result, important questions remain regarding how these different input sources shape listening comprehension, engagement, and affective experience. Addressing this gap is crucial, as pedagogical decisions about AI integration should be informed not only by technological capabilities but also by learners' perceptions and classroom experiences (Zhu & Wang, 2024). Examining learners' responses to both input types can therefore offer valuable insights for designing balanced listening instruction that integrates innovation without diminishing the human dimensions of language teaching.

Despite the increasing integration of AI technologies in education, the question of how AI-generated speech compares to teacher talk as a source of listening input remains underexplored. Teacher talk is widely regarded as an authentic, interactive, and pedagogically adaptive form of input that provides learners with scaffolding and contextual support (Gibbons, 2015; Walsh, 2011). In contrast, AI-generated voices, while offering consistency, accessibility, and diversity of listening materials, may lack the immediacy, affective presence, and interactive feedback that characterize teacher talk.

Existing studies have examined digital listening tools such as podcasts, audiobooks, and text-to-speech applications (Jian, 2023), but few have directly compared students' experiences and perceptions of AI-generated input with those of teacher talk in authentic classroom settings. As a result, little is known

about how different input sources influence learners' comprehension, engagement, and attitudes in EFL listening instruction.

This research addresses the gap by investigating the following problem: to what extent do AI-generated input and teacher talk differ in shaping students' listening comprehension and engagement, and how do learners perceive the benefits and challenges of each input type.

By qualitatively examining classroom interactions, student reflections, and interviews, this study aims to provide insights into the pedagogical implications of integrating AI-generated voices alongside traditional teacher talk. Such an investigation is critical to ensuring that the adoption of AI in EFL classrooms enhances, rather than diminishes, the quality of listening instruction and learner experience. To achieve these objectives, the study is guided by the following research questions:

a. How do EFL learners perceive and experience AI-generated voices compared to teacher talk in listening instruction?
b. What benefits and challenges do learners associate with AI-generated input and teacher talk as sources of listening practice?
c. In what ways do AI-generated voices and teacher talk influence learners' comprehension strategies and engagement in the classroom?

Practically, the findings provide valuable insights for teachers, curriculum designers, and policymakers who are seeking to balance the integration of AI technologies with traditional pedagogical practices. For teachers, the results may highlight effective ways to combine AI-generated input with classroom discourse to maximize learner exposure and engagement. For curriculum designers, the study offers evidence-based considerations for incorporating AI-driven tools into listening instruction without diminishing the interpersonal and adaptive benefits of teacher talk. Finally, for policymakers and institutions, the study underscores the importance of adopting technology in a measured and pedagogically informed manner, ensuring that innovations in AI support rather than replace the human dimension of language teaching.

## 2. METHODS

This study employs a qualitative research design to explore the comparative influence of teacher talk and AI-generated voices as sources of listening input in an EFL context. A qualitative approach was chosen because it enables an in-depth understanding of learners' experiences, perceptions, comprehension, and engagement with different types of listening input, which cannot be fully captured through quantitative measures alone (Creswell, J.W. & Poth, 2018). Data were collected through classroom observations, semi-structured interviews, and reflective journals, allowing for triangulation and a holistic exploration of the phenomenon. Participants consisted of EFL students in university, selected purposively to represent learners with equal exposure to teacher talk and AI-generated talks. The data were analyzed thematically using (Braun, V., & Clarke's, 2006) framework to identify patterns and themes related to learner comprehension, affective responses, and perceived benefits or challenges of each listening source.

### Research Design

This study adopts a qualitative case study design to investigate how teacher talk and AI-generated voices function as listening sources in an EFL classroom. A case study approach was chosen because it allows for a detailed and context-sensitive exploration of a bounded system, namely a group of university students engaged in listening instruction. Unlike experimental or survey-based designs, the case study provides an opportunity to capture the richness of learners' experiences and the interplay between classroom dynamics and technological input. The study emphasizes interpretive inquiry, aiming not to generalize findings to all EFL contexts but to generate deep insights into how learners perceive, engage with, and respond to different listening inputs. By combining classroom observations, semi-structured interviews, and reflective journals, the design ensures multiple perspectives on the phenomenon, thus enhancing validity through triangulation (Merriam, S. B., & Tisdell, 2016).

### Participant

The participants of this study were 20 undergraduate students enrolled in an English as a Foreign Language (EFL) program at a private university in Indonesia. They were taking an intensive and extensive listening course, which is designed to provide systematic exposure to spoken English through both classroom instruction and independent practice. The course was held twice a week and was supplemented by multiple hours of self-directed listening practice outside the classroom. A purposive sampling strategy was employed to ensure that the participants represented learners with relatively similar proficiency levels. The 20 participants consisted of both male and female students, aged between 18 and 20 years. This

group was selected because they had nearly equal exposure to teacher talk and AI-generated spoken input, making them an appropriate sample for investigating the role of different listening sources. In this course, AI-generated input was delivered through Gliglish, an AI-powered website that allows learners to practice English through spoken conversations with an artificial intelligence tutor, while teacher talk was provided in the usual instructional manner, particularly during explanations and classroom instructions. The researcher also served as the course instructor, ensuring consistent implementation of the instructional procedures, while other members of the research team were responsible for assisting with data analysis to maintain objectivity. Participation in the study was voluntary, with informed consent obtained prior to data collection, and ethical considerations such as anonymity and confidentiality were strictly maintained in accordance with standard research ethics in applied linguistics.

**Instruments**

To capture a comprehensive understanding of learners' experiences with teacher talk and AI-generated voices as listening sources, three primary instruments were employed: classroom observations, semi-structured interviews, and reflective journals. Classroom observations were conducted to document the dynamics of listening activities, focusing on learners' engagement, and comprehension strategies. Field notes were taken systematically, guided by an observation protocol to ensure consistency. Semi-structured interviews were used to elicit learners' perceptions and attitudes toward the two types of listening input. This format allowed flexibility in probing students' responses while maintaining comparability across interviews (Kvale & Brinkmann, 2015). Additionally, reflective journals were collected from students over the course of the study. These journals provided ongoing insights into learners' subjective experiences, challenges, and perceived benefits of engaging with teacher talk and AI-generated input. The combination of these instruments ensured data triangulation, enhancing the trustworthiness and richness of the findings.

**Data Collection Procedures**

Data collection was carried out over a six-week period during the students' regular listening classes. In the first stage, classroom observations were conducted on a weekly basis to capture learners' interactions with both teacher talk and AI-generated listening input. A structured observation guide was used to take notes and focusing on several predefined categories, including: (1) learners' behavioral engagement (e.g., attention, participation, and responsiveness); (2) observable listening and comprehension strategies (e.g., note-taking, asking for clarification, use of contextual cues); and (3) learners' reactions to different sources of spoken input. In the second stage, semi-structured individual interviews were conducted with selected participants, each lasting approximately 20–30 minutes. These interviews explored students' perceptions of the clarity, authenticity, and pedagogical usefulness of teacher talk and AI-generated speech as listening sources. In the third stage, participants maintained reflective journals throughout the study by writing weekly entries guided by prompts such as: (1) Which listening input (teacher talk or AI-generated speech) was easier or more difficult to understand this week? (2) What strategies did you use to help your comprehension? (3) What challenges or benefits did you experience when listening to each type of input? All data were collected in compliance with ethical procedures, including informed consent, voluntary participation, and the confidentiality of participants' responses (Cohen, Manion, & Morrison, 2018).

**Data Analysis**

The collected data were analyzed using thematic analysis, following (Braun & Clarke's, 2006) six-phase framework. The analysis adopted data-driven coding approach, allowing codes and themes to emerge from the data rather than being predetermined by existing theory. Observation notes, interview transcripts, and reflective journals were first transcribed and read repeatedly to achieve data familiarization. Initial coding was conducted manually without the use of qualitative analysis software, with codes generated to capture recurring ideas related to learners' comprehension strategies, engagement, experiences, perceptions, as well as the perceived benefits and challenges associated with the two listening sources. These codes were then iteratively reviewed, refined, and grouped into broader themes that represented patterns across the combined data sets. To enhance the trustworthiness of the analysis, data triangulation was employed by comparing findings across the three data sources. In addition, coding consistency was ensured through discussion and review of the codes among the research team, and member checking was conducted by sharing preliminary findings with selected participants to confirm their accuracy.

## 3. RESULT AND DISCUSSION

### Results

### Students' Perception and Experience of AI-generated Voices and Teacher Talk in Listening Instruction

Analysis of semi-structured interviews and reflective journals revealed three interrelated themes: (a) clarity and comprehensibility, (b) authenticity and natural interaction, and (c) flexibility and personalization of practice. Across these themes, learners emphasized that each source of input offered unique strengths and limitations.

a. Clarity and Comprehensibility

AI-generated voices were frequently described as "clear," "standardized," and "easy to understand." Students noted that the consistent pronunciation and lack of background noise reduced listening anxiety, especially for beginners: "The AI voice is very clear; I don't get confused by fast speech or accents." However, others commented that the uniformity of AI delivery made it feel less natural and limited their preparedness for real-life communication.

In contrast, teacher talk was perceived as more challenging due to variations in pace, intonation, and occasional digressions. Yet, learners also valued this challenge, recognizing it as closer to authentic communication. As one student reflected, "Sometimes the teacher speaks faster or uses different expressions, but it helps me learn how English really sounds." Thus, while AI voices supported initial comprehension, teacher talk was seen as crucial for developing listening resilience in authentic contexts.

b. Authenticity and Natural Interaction

Students consistently associated teacher talk with authenticity and social engagement. They highlighted the spontaneity of teacher speech, including hesitations, emphasis, and real-time responses. Importantly, teacher talk enabled opportunities for clarification and negotiation: "If I don't understand, I can ask the teacher to explain again or use examples." This interactive nature gave learners a stronger sense of connection and contextual understanding.

By contrast, AI-generated input was perceived as "artificial" or "robotic" despite improvements in prosody. While students acknowledged that AI could mimic natural speech, they still found it lacking in emotional tone and spontaneity: "AI sounds human, but it cannot respond to me or show feeling like my teacher does." Reflective journals emphasized that the absence of interaction made AI input feel "isolated" compared to the richer communicative environment fostered by teacher talk.

c. Flexibility and Personalization of Practice

AI-generated voices were praised for their adaptability in self-directed learning. Students appreciated features such as playback control, adjustable speed, and exposure to multiple accents. One journal entry noted, "I like that I can repeat the AI voice many times without feeling shy." This flexibility allowed learners to practice independently and at their own pace, making AI a useful supplement outside class. Teacher talk, however, was perceived as more responsive in real time. Learners observed that teachers adjusted their speech according to students' reactions, simplified explanations when necessary, and added contextual cues like gestures. As one interviewee remarked, "The teacher knows when we look confused and slows down or explains in another way. AI cannot do that." Thus, while AI input supported autonomous listening practice, teacher talk was viewed as indispensable for adaptive, socially meaningful listening instruction.

Overall, learners perceived AI-generated voices and teacher talk not as competing but as complementary listening sources. AI was valued for its clarity and flexibility in individualized practice, whereas teacher talk was appreciated for its authenticity, interactional richness, and adaptive scaffolding. Together, these findings suggest that balanced integration of both input types may provide learners with optimal listening development opportunities.

### What are the Perceived Benefits and Challenges of Controlled Extensive Listening Activities Using AI-generated Voices and Teacher Talk

Analysis of the data revealed four key themes: (a) motivational impact, (b) listening skill development, (c) emotional and affective responses, and (d) limitations and challenges.

a. Motivational Impact

AI-generated input was perceived as motivating for self-directed practice because learners could engage without fear of judgment. Students expressed that AI reduced performance anxiety: "I

can listen to the AI voice many times without worrying the teacher will think I am slow." The variety of AI accents also maintained learner curiosity and novelty.

Teacher talk, however, was seen as motivating due to the social presence and real-time encouragement teachers provided. Learners emphasized the teacher's ability to inspire effort and persistence: "When the teacher praises us or jokes while speaking, it makes me more interested in listening." Thus, AI was linked to reduced anxiety and independent motivation, while teacher talk was linked to relational and emotional motivation.

b. Listening Skill Development

AI voices were praised for improving comprehension strategies especially vocabulary recognition and pronunciation awareness because of their consistent clarity. Reflective journals noted that repeated exposure to AI speech enhanced word recognition: "I started to notice how some words are pronounced clearly in the AI recording."

Teacher talk, in contrast, was associated with broader communicative competence. Learners highlighted how exposure to spontaneous speech, natural hesitations, and context-driven language improved their real-world listening skills. As one student stated: "The teacher's talk is not always perfect, but that makes it closer to how people actually speak in daily life."

c. Emotional and Affective Responses

Learners frequently described AI input as "neutral" or "emotionless," which sometimes led to disengagement: "It is clear but feels boring after a while because there is no feeling in the voice." Although useful for practice, students felt AI could not replicate the human warmth of a teacher.

By contrast, teacher talk elicited stronger emotional connections. The tone, gestures, and humor embedded in teacher talk helped learners feel supported and valued. "I feel more confident when the teacher explains because I know she cares about us," wrote one participant. This emotional resonance was described as a crucial factor in sustaining long-term engagement with listening tasks.

d. Limitations and Challenges

Despite its advantages, AI input was seen as limited in interaction and contextual adaptability. Learners noted that AI could not clarify meaning, adjust to student confusion, or respond spontaneously: "If I don't understand, AI just repeats the same thing. It cannot explain like my teacher."

Teacher talk, while valued, was also perceived as challenging. Some students struggled with fast pace, varied accents, and the unpredictability of real-time explanations. As one learner admitted: "Sometimes the teacher speaks too quickly, and I miss the meaning. I cannot pause or replay like AI." Thus, while AI lacked adaptability, teacher talk sometimes demanded higher processing effort that overwhelmed less confident learners.

Findings suggest that learners view both AI-generated voices and teacher talk as beneficial but in different ways. AI supports clarity, repeated practice, and independent learning, but lacks emotional warmth and interactivity. Teacher talk fosters authentic communication, social motivation, and emotional connection, though it can be cognitively demanding. Together, these complementary benefits highlight the need for integrated use of AI and teacher talk to maximize listening instruction effectiveness.

**In What Ways Do AI-generated Voices and Teacher Talk Influence Learners' Comprehension and Engagement in the Classroom**

Findings from classroom observations and interviews revealed three main themes: (a) comprehension strategies supported by input type, (b) patterns of engagement in classroom activities, and (c) the complementary role of AI and teacher talk in sustaining participation.

a. Comprehension Strategies Supported by Input Type

Observation data showed that AI-generated voices promoted comprehension by providing clear articulation and predictable pacing. Students were able to identify key vocabulary and phrases more accurately during listening tasks when AI voices were used, often requiring fewer repetitions compared to teacher talk. Interviews confirmed this, with one student stating: "I can catch more words from AI because the pronunciation is very standard."

In contrast, comprehension during teacher talk was supported through interactional adjustments and contextual scaffolding. Observers noted that when students appeared confused, the teacher often repeated, slowed down, or rephrased explanations, which immediately improved comprehension. One learner explained: "Sometimes I don't understand at first, but when the teacher explains again with examples, it becomes clearer." This indicates that while AI facilitated comprehension through clarity and consistency, teacher talk enhanced comprehension through adaptability and context.

b. Patterns of Engagement in Classroom Activities

AI-generated voices were associated with focused individual engagement. During AI-based tasks, observers noted that students were more concentrated on the audio itself, with less side-talk and more frequent note-taking. Learners described AI activities as "serious" moments where they felt responsible for understanding on their own. However, some reported that engagement felt "less lively" because there was limited interaction with peers or the teacher.

Teacher talk, on the other hand, fostered interactive engagement. Observation revealed that learners were more likely to raise their hands, ask questions, or respond spontaneously during teacher-led sessions. Students described teacher talk as "more motivating" because they could immediately interact with the teacher and peers. One student noted: "I pay more attention when the teacher is speaking, because I know she might ask me something."

c. Complementary Role in Sustaining Participation

Both sources influenced classroom participation in distinct but complementary ways. AI-generated input encouraged learners to take responsibility for their own comprehension and to practice independently, which boosted confidence in subsequent activities. Teacher talk, however, sustained participation by creating a supportive classroom atmosphere, where learners felt encouraged to try, even when uncertain. As one student reflected: "AI helps me prepare, but the teacher makes me brave to use what I hear."

Classroom observations and interviews indicate that AI-generated voices influence comprehension by offering clarity and consistency, and engagement through focused, independent effort. In contrast, teacher talk supports comprehension by adapting to learners' needs and fosters engagement through interaction and immediacy. Together, these sources play complementary roles in enhancing both comprehension and engagement in EFL listening instruction.

**Triangulation**

To ensure the trustworthiness and validity of the findings, this study employed triangulation by integrating multiple data sources and methods. Specifically, classroom observations, semi-structured interviews, and reflective journals were used to examine learners' perceptions, experiences, and engagement with AI-generated voices and teacher talk in listening instruction.

Data source triangulation was achieved by comparing perspectives from different instruments. For example, learners' reports in interviews about the clarity of AI-generated voices were cross-checked with observational notes that showed fewer requests for repetition during AI-based tasks. Similarly, students' reflections in journals about feeling more motivated during teacher talk sessions were confirmed by observations of increased hand-raising and classroom interaction.

Methodological triangulation was established by combining qualitative instruments that captured both observable behaviors and personal reflections. While classroom observations provided evidence of how learners engaged with listening tasks in real time, interviews offered deeper insights into their reasoning and affective responses, and reflective journals revealed ongoing patterns of individual experience across lessons.

Through this triangulated approach, the study reduced the risk of bias inherent in relying on a single data source and strengthened the credibility of interpretations. The convergence of evidence across instruments provides a more comprehensive understanding of how different listening sources, AI-generated input and teacher talk shape EFL learners' comprehension and engagement.

**Discussion**

The findings of this study provide deeper insight into how EFL learners experience and engage with AI-generated voices and teacher talk as listening input, revealing that each source affords different cognitive, affective, and pedagogical benefits. Drawing on triangulated data from classroom observations, semi-structured interviews, and reflective journals, the discussion illustrates not only what learners perceived, but also why these perceptions emerged and how they shape listening comprehension and engagement in the classroom.

To begin with, learners consistently perceived AI-generated voices as clear, stable, and easy to understand, particularly at earlier stages of listening practice. This finding can be interpreted through the lens of cognitive processing: the acoustic consistency of AI speech appears to reduce listeners' processing load, enabling learners to focus on decoding linguistic forms rather than adapting to variation in accent or delivery. This interpretation aligns with prior research indicating that text-to-speech technology can support comprehension by minimizing prosodic and phonological variability (Ní Chiaráin & Chasaide, 2020; Tan et al., 2021). However, classroom observations add an important layer of nuance. While AI-generated input supported accurate comprehension, learners were often observed engaging in individual

and relatively passive listening behaviors, such as silent note-taking and limited peer or teacher interaction. This suggests that although AI input facilitates understanding, it may not inherently prompt interactive engagement unless deliberately mediated by instructional design.

In contrast, teacher talk emerged as a more interactionally rich and pedagogically responsive input source. Despite being less acoustically uniform than AI-generated voices, teacher talk provided contextual explanations, spontaneous repetition, clarification, and real-time adjustment to learners' needs. Observational data showed that during teacher-led listening activities, learners more frequently asked clarification questions, responded verbally, and negotiated meaning. These findings support (Walsh's, 2011) argument that teacher talk functions as an interactional resource that actively shapes classroom discourse and supports learning through scaffolding. The effectiveness of teacher talk, therefore, lies not only in linguistic input, but in its adaptability and responsiveness to learner cues, which AI-generated speech currently cannot fully replicate.

Beyond comprehension, the findings also highlight clear differences in affective and motivational responses to the two listening sources. Learners reported feeling more comfortable and motivated when listening to their teacher, often attributing this to the interpersonal connection and immediacy of feedback. This supports the view that listening is not solely a cognitive activity, but a socially situated process shaped by relationships and classroom dynamics (Vandergrift, L., & Goh, 2012). In contrast, although AI-generated input was initially perceived as novel and engaging, its motivational impact tended to diminish over time. Observations reinforce this interpretation, as students demonstrated sustained attention but fewer interactional behaviors during AI-based tasks, suggesting a form of engagement that was cognitively focused yet socially limited.

Taken together, these findings indicate that AI-generated input and teacher talk serve complementary rather than competing roles in EFL listening instruction. AI-generated voices offer clarity, consistency, and opportunities for self-paced practice, while teacher talk provides interactional authenticity, emotional support, and pedagogical scaffolding. This complementarity helps explain why learners benefited from exposure to both sources, yet consistently valued teacher talk for maintaining engagement and motivation. In this sense, the findings extend (Jian's, 2023) argument by showing not only that AI should be balanced with human instruction, but why such balance is pedagogically necessary.

In conclusion, this study underscores the importance of thoughtful pedagogical integration of AI-generated listening input in EFL classrooms. Rather than replacing teacher talk, AI tools should be positioned as supplementary resources that expand learners' exposure to spoken language and support autonomous practice. Teacher talk, meanwhile, remains central to fostering interactive, socially meaningful, and motivating listening experiences. By leveraging the distinct strengths of both input sources, educators can create a more balanced and effective listening environment that addresses learners' cognitive, affective, and interactional needs.

## 4. CONCLUSION

This study explored how EFL learners perceive and experience AI-generated voices compared to teacher talk, as well as how both input sources influence comprehension and engagement in the classroom. By providing empirical, classroom-based evidence on the pedagogical roles of AI-generated listening input in comparison with teacher talk, this study contributes to the growing body of EFL research on technology-enhanced listening by clarifying how cognitive, interactional, and affective dimensions of listening are differentially supported by each input source. Findings from interviews, reflective journals, and classroom observations demonstrate that while AI-generated voices offer clarity, consistency, and flexible access to diverse listening input, teacher talk provides interactional authenticity, scaffolding, and affective support that are central to meaningful listening development. Learners valued the accessibility of AI input for practice and comprehension, yet expressed stronger engagement and motivation during teacher-led listening activities.

The results suggest that AI-generated voices and teacher talk function not as competing sources but as complementary modalities that, when integrated thoughtfully, can enhance both comprehension and engagement. AI can extend exposure and support individualized learning, while teacher talk remains vital for fostering interaction, negotiation of meaning, and sustained motivation. However, the study is limited by its relatively small sample size and short implementation period, which may constrain the generalizability of the findings.

Pedagogically, the study highlights the need for balanced listening instruction that leverages the strengths of both input types. Rather than viewing AI as a substitute for teacher talk, educators should employ AI-based tools as supplementary resources alongside teacher-led listening activities. This integrated approach can provide learners with richer, more varied listening experiences that prepare them

for authentic language use beyond the classroom. Future research is recommended to involve larger and more diverse learner populations, incorporate longitudinal designs, and examine the impact of AI–teacher input integration on measurable listening outcomes to further strengthen the evidence base.

## 5. REFERENCES

Creswell, J.W. and Poth, C. N. (2018). *Qualitative Inquiry and Research Design Choosing among Five Approaches* (4th ed.). Sage Publications.

Field, J. (2019). *Listening in the Language Classroom*. Cambridge University Press.

Gibbons, P. (2015). *Scaffolding Language, Scaffolding Learning: Teaching English Language Learners in the Mainstream Classroom*. Heinemann.

Jian, M. (2023). Personalized learning through AI. *Advances in Engineering Innovation*, *5*. https://doi.org/10.54254/2977-3903/5/2023039

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative Research: A Guide to Design and Implementation*. Jossey-Bass.

Ní Chiaráin, N., & Chasaide, A. (2020). *The Potential of Text-to-Speech Synthesis in Computer-Assisted Language Learning* (pp. 149–169). https://doi.org/10.4018/978-1-7998-1097-1.ch007

Porubay, I., Tashmatova, G., Akhmadjonova, S., Uktamova, M., Kuziev, S., & Mirzaakhmedov, M. (2024). Artificial Intelligence in Language Classrooms: Improving Engagement and Outcomes. *2024 International Conference on IoT, Communication and Automation Technology (ICICAT)*, 335–340. https://doi.org/10.1109/ICICAT62666.2024.10923388

Richards, J. C. (2017). *Teaching Listening and Speaking: From Theory to Practice*. SEAMEO Regional Language Centre.

Sert, O. (2015). *Social Interaction and L2 Classroom Discourse*. https://doi.org/10.1515/9780748692651

Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., Zhang, Y., Wang, Y., Skerrv-Ryan, R., Saurous, R. A., Agiomvrgiannakis, Y., & Wu, Y. (2018). Natural TTS Synthesis by Conditioning Wavenet on MEL Spectrogram Predictions. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4779–4783. https://doi.org/10.1109/ICASSP.2018.8461368

Tan, X., Qin, T., Soong, F., & Liu, T.-Y. (2021). *A Survey on Neural Speech Synthesis*. http://arxiv.org/abs/2106.15561

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and Learning Second Language Listening: Metacognition in Action*. Routledge.

Vasheghani Farahani, M. (2021). METADISCOURSE: EXPLORING INTERACTION IN WRITING, Second Edition By Ken Hyland. *TESOL Journal*.

Walsh, S. (2011). *Exploring Classroom Discourse: Language in Action*. Routledge. https://doi.org/10.4324/9780203827826

Zhu, M., & Wang, C. (2024). *A Systematic Review of Artificial Intelligence in Language Education from 2013 to 2023: Current Status and Future Implications*. https://doi.org/10.2139/ssrn.4684304