Didaktika: Jurnal Kependidikan
P-issn 1978-0214; E-ISSN 2657-196X

AI-Generated Transcripts in EFL Listening: Supporting, Monitoring, Regulation, and Strategy Shifts

Cendikia Flory Aristia[1], Dellis Pratika[2], Niwayan Sukraini[3]
[1] Universitas Palangka Raya, Indonesia; cendikiafa@fkip.upr.ac.id
[2] Universitas Palangka Raya, Indonesia; dellispratika@fkip.upr.ac.id
[3] Institut Agama Hindu Negeri Tampung Penyang Palangka Raya, Indonesia;niwayansukraini@iahntp.ac.id

Keywords:
AI-generated transcripts; EFL listening; gist-oriented understanding; metacognitive monitoring

Abstract

This study explores how 31 intermediate EFL learners used AI-generated transcripts to monitor and regulate their listening comprehension over seven weeks of structured listening activities. Data were collected through open-ended questionnaires and focus group discussions and analyzed thematically. Findings indicate that participants used transcripts to compare spoken input with written text, identify misheard or unfamiliar words, and verify their interpretations. These activities appeared to assist learners in adjusting their listening strategies and noticing comprehension gaps. However, the results are perceptual and context-specific, reflecting participants' subjective experiences within a limited sample and cannot be generalized broadly. The study offers preliminary insights into the potential of AI transcripts as reflective tools in EFL listening instruction.

Kata kunci:
Transkrip hasil AI; *EFL Listening*; pemahaman berorientasi gagasan utama; pemahaman menyimak; pemantauan metakognitif

Abstrak

Studi ini menyelidiki bagaimana 31 pembelajar EFL tingkat menengah menggunakan transkrip yang dihasilkan AI untuk memantau dan mengatur pemahaman menyimak mereka selama tujuh minggu kegiatan menyimak terstruktur. Data dikumpulkan melalui kuesioner terbuka dan diskusi kelompok terarah, kemudian dianalisis secara tematik. Temuan menunjukkan bahwa peserta menggunakan transkrip untuk membandingkan input lisan dengan teks tertulis, mengenali kata yang salah didengar atau kosakata yang belum dikenal, serta memverifikasi interpretasi mereka. Aktivitas ini tampak membantu peserta dalam menyesuaikan strategi menyimak mereka dan memperhatikan kesenjangan pemahaman. Namun, temuan bersifat persepsional dan kontekstual, mencerminkan pengalaman subjektif peserta dalam sampel terbatas, dan tidak dapat digeneralisasi secara luas. Studi ini memberikan wawasan awal mengenai potensi transkrip AI sebagai alat bantu refleksi dalam pembelajaran menyimak EFL.

Article history:
Received: 01-05-2026
Revised 10-06-2026
Accepted 00-06-2026

Corresponding Author:
Cendikia Flory Aristia
Universitas Palangka Raya, Indonesia; cendikiafa@fkip.upr.ac.id

INTRODUCTION

Second language (L2) listening is widely recognized as one of the most difficult challenges for English as a Foreign Language (EFL) learners (Vandergrift & Goh, 2011). Unlike reading, listening occurs in real time, requiring learners to process fleeting acoustic input while

simultaneously decoding phonemes, recognizing vocabulary, and interpreting syntax within the limits of working memory (Field, 2010; Goh, 2000). These difficulties are intensified by learners' limited ability to identify words in continuous speech and use prior or contextual knowledge to build a coherent interpretation (Goh, 2023). EFL students often struggle most with decoding naturalistic, rapid, or accented speech, where phonological barriers can disrupt word-boundary recognition and prevent the shift from bottom-up acoustic processing to top-down meaning construction (Richards, 1983; Field, 2010; Goh, 2023).

Effective listening comprehension is generally understood as the dynamic interaction of bottom-up and top-down processes (Richards, 1983; Vandergrift & Goh, 2012). As Jones (2025) notes, L2 listeners must coordinate lower-level decoding of the speech signal with higher-level meaning construction in context. Bottom-up processing involves identifying phonemes, word boundaries, and grammatical structures from the acoustic input, while top-down processing draws on prior knowledge, situational context, and expectations to infer meaning and grasp the overall message (Field, 2010; Jones, 2025; Vandergrift & Goh, 2012). For many EFL learners, comprehension breaks down when they over-rely on bottom-up decoding and become "stuck" at the word level, expending cognitive resources on isolated lexical items rather than activating top-down strategies (Goh, 2000). As a result, they may recognize scattered words without integrating them into a coherent mental model (Goh, 2000; Vandergrift & Goh, 2012).

The successful coordination of bottom-up and top-down processes depends heavily on metacognition, particularly learners' ability to monitor and evaluate their comprehension during listening tasks (Vandergrift & Goh, 2012). Metacognitive monitoring is an active, cyclical process of self-diagnosis, planning, and regulation that helps learners detect gaps between the incoming acoustic signal and their mental representation of the text (Liu, 2020). When learners can identify where comprehension breaks down, they can use repair strategies to recover meaning (Vandergrift & Goh, 2012). However, the rapid and transient nature of listening often makes real-time monitoring difficult for EFL learners (Field, 2010; Goh, 2000). Although metacognitive awareness is essential, strategy training appears most effective when paired with cognitive supports that enable immediate reflection (Liu, 2020). Without a stable reference point for checking their interpretations, learners may remain unaware of misperceptions or become overwhelmed by cognitive load (Goh, 2000; Vandergrift & Goh, 2012).

With the rapid advancement of educational technology, AI-generated transcripts, including those from YouTube and automated speech recognition (ASR) platforms, have become accessible support tools for EFL listening. By visualizing spoken input, they transform transient acoustic signals into stable textual representations that learners can review at their own pace (Boltiziar & Munkova, 2024; Hsieh, 2020; Perez, 2022). This visual anchor helps learners connect spoken and written forms, identify unfamiliar or misheard words, and verify comprehension when facing fast speech, unfamiliar accents, or unknown vocabulary (Boltiziar & Munkova, 2024). Rather than replacing real-time decoding, AI-generated transcripts may reduce the cognitive burden of word recognition, allowing learners to devote more attention to higher-level comprehension, meaning construction, and gist-oriented understanding (Nguyen & Nguyen, 2025; Xiao, 2025). Thus, AI transcripts serve not only as textual support but also as a metacognitive anchor that helps learners check, notice, and strategically regulate their listening comprehension.

While the potential of automated speech recognition and AI-assisted listening tools is increasingly recognized, existing studies have not yet made transcript-mediated listening behavior their central focus. Recent research has mainly examined AI-supported listening through cognitive load, learner perceptions, or broad skill development. For example, Nguyen and Nguyen (2025) reported reported that AI-generated audio transcription reduced cognitive load and helped English majors manage task complexity. Luu and Doan (2025) examined Vietnamese EFL university learners' perceptions of Chat GPT for L2 listening through an explanatory sequential mixed-method design. Their findings showed that students recognized ChatGPT's benefits for vocabulary development, grammar knowledge, and working memory. However, the study focused mainly on perceived usefulness across listening-related domains rather than on the detailed processes through which learners monitor comprehension while listening. Similarly, Sakuta et al. (2026) used a qualitative descriptive approach to examine how YouTube and AI tools, including real-time transcription, automatic subtitles, summarization, and adaptive feedback, supported extensive listening. Although their study showed that learners used AI tools to plan, monitor, and evaluate listening performance, its focus remained broad, covering extensive listening, motivation, pronunciation, autonomy, and perceived skill development. It did not specifically examine the fine-grained cognitive and metacognitive behaviors that occur when learners interact with AI-generated transcripts during listening.

These studies confirm the pedagogical value of AI-supported listening tools, but they also reveal a methodological gap. Existing research has not yet explained in sufficient detail how EFL learners use AI-generated transcripts as a real-time or near-real-time reference point to diagnose mishearing, repair comprehension breakdowns, connect acoustic input with written forms, and shift from fragmented word-level decoding toward gist-oriented understanding. In other words, the field still lacks a process-oriented qualitative account of how learners interact with AI-generated transcripts to regulate listening behavior and strategy use. Although AI-generated audio transcriptions appear to help manage mental effort, little is known about the monitoring behaviors through which EFL learners reshape their cognitive strategies and develop a more integrated metacognitive approach to listening.

This study examines how EFL learners use AI-generated transcripts to monitor listening comprehension and how this use reshapes their listening processes. Addressing the identified gaps, it focuses on qualitative shifts in learner behavior when a visual linguistic anchor is available, particularly the transition from labor-intensive decoding to more efficient meaning-making (Field, 2010; Vandergrift & Goh, 2012). Moving beyond conventional asynchronous listening tasks, which are often criticized as monotonous or inconsistent in audio quality, the study aligns with innovative AI-supported approaches that aim to enhance engagement and listening fluency (Chua et al., 2025). Ultimately, it investigates whether technology-mediated support can help learners overcome the transience of spoken language and develop a more autonomous, meaning-oriented listening style.

To address these gaps, this study is guided by two research questions. RQ1: How do EFL learners use AI-generated transcripts to monitor and regulate their listening comprehension? This question examines how learners engage with transcripts to bridge the gap between perception and understanding. RQ2: How does this reshape their listening processes toward gist-oriented understanding? This question explores whether visual transcripts can

reduce the bottleneck of word-level decoding and support a shift toward more global, top-down processing.

METHOD

This study employed a qualitative case study design to explore how EFL learners used AI-generated transcripts to monitor and reshape their listening processes within a bounded instructional context over time. The case consisted of a single cohort of 31 undergraduate EFL learners enrolled in a university English course. Participants were purposively selected because they had prior experience with English listening activities, access to digital tools, and intermediate-level English proficiency, determined through institutional placement equivalent to the B1–B2 levels of the Common European Framework of Reference (CEFR). This design enabled an in-depth examination of learners' metacognitive monitoring and strategy development within an authentic classroom setting rather than focusing on measurable outcomes (Creswell, 2013; Merriam & Tisdell, 2016; Yin, 2018; Creswell & Creswell, 2023). All participants provided informed consent.

The study was conducted over seven weeks during regular classroom instruction. Learners engaged in weekly listening activities using transcripts obtained from YouTube's auto-generated caption/ASR feature for selected BBC Learning English videos. The listening materials consisted of authentic and semi-authentic audio recordings, including interviews, informational talks, and academic monologues appropriate for intermediate-level learners. Each session followed a structured sequence: listening without transcripts to encourage prediction and global comprehension, identifying unclear segments, consulting the AI-generated transcript to compare perceived and actual input, and reflecting on sources of misunderstanding and strategy adjustments. This repeated cycle encouraged learners to externalize their listening processes, identify comprehension gaps, and develop greater metacognitive awareness.

Data were collected through open-ended questionnaires and focus group discussions (FGDs) to capture learners' perceptions, monitoring practices, challenges, and strategy use while enabling methodological triangulation (Creswell, 2013; Hennink, 2014). The questionnaire contained open-ended items examining learners' experiences with transcript use, perceived benefits and limitations, and changes in listening behavior. Following the instructional period, FGDs were conducted using a semi-structured protocol that explored learners' reflections on transcript accuracy, listening difficulties, monitoring processes, and strategy development. Example questions included, "*Were there any parts of the AI transcript that you disagreed with or thought were errors? How did you decide if the AI was right or wrong?*" or "*How did the AI-generated transcript help you understand the listening text?* All discussions were audio-recorded and transcribed verbatim for analysis.

Data were analyzed using reflexive thematic analysis (Braun & Clarke, 2006; Maguire & Delahunt, 2017). The analysis involved repeated reading of questionnaire responses and FGD transcripts, inductive generation of initial codes, and the organization of related codes into broader themes representing patterns across the dataset. Examples of codes included mishearing words, checking assumptions, confirming interpretations, noticing pronunciation differences, and adjusting listening strategies. Themes were reviewed, refined, and defined

through iterative discussions among the researchers to ensure coherence and representativeness. To enhance credibility, three researchers independently coded a subset of the data and resolved discrepancies through discussion, while triangulation across questionnaires and FGDs and the maintenance of analytic memos and an audit trail supported the trustworthiness of the findings.

FINDINGS AND DISCUSSION

3.1 Monitoring Listening Through AI Transcripts

3.1.1 Metacognitive Monitoring via Comparison

Across the datasets, learners consistently engaged in systematic comparison between audio input and AI-generated transcripts as a central mechanism for monitoring their listening comprehension. This process involved deliberate alignment between what learners heard and what they read. As one participant explained, "we try to understand what we hear and what we read", indicating an active effort to reconcile the two sources of input.

This comparative behavior was also evident in the survey data. Many learners reported that, upon accessing the transcript, they immediately looked for specific words they had missed or compared the transcript with their notes. Common responses included actions such as "looked for specific words you missed" and "compared it to your notes," suggesting that the transcript was used as a tool to verify their initial understanding.

Learners' reliance on this strategy appeared to be related to their initial listening experience. Self-reported comprehension during the first listening attempt varied, but many responses clustered around mid-range levels (approximately 50–80%), indicating partial understanding. This suggests that learners experienced some degree of uncertainty, which led them to consult the transcript to confirm or refine their interpretations. The comparison supported ongoing monitoring of understanding. Learners described how their initial interpretations were checked and sometimes corrected after consulting the transcript. One participant stated, "At the first time I misheard it… then I look at the transcript, so I know it." Similarly, survey responses included statements such as "check my understanding," "see what I missed," and "learn from mistakes," indicating repeated cycles of listening, checking, and revising.

The effectiveness of this process is reflected in learners' evaluations of the AI transcript. Most participants rated the transcript as highly helpful in making them aware of their listening mistakes, with responses largely concentrated in the upper range of the scale. Learners' responses indicate an affective dimension in their engagement with the comparison process. While some learners initially expressed discomfort when noticing discrepancies, many also reported that the process helped them become more aware of their listening difficulties. Learners noted that the transcript "helped me see exactly which words I missed," "made me realize my listening weaknesses," and "helped me understand the words I missed."

Although most learners viewed transcript comparison as beneficial, the data also revealed variation in their experiences. Some participants provided only general comments such as "I don't know," "nothing," or "very helpful," suggesting limited reflection on the causes of their listening difficulties. Others reported continued challenges with unfamiliar accents or rapid speech even after consulting the transcript. A small number also expressed concerns about

dependence on transcript support, with one learner stating, "Listen less because I know there will be transcript," while another noted that transcript use was "making me depend on transcript more than my understanding." These responses indicate that the benefits of AI-generated transcripts were not experienced uniformly across all learners and that transcript-supported monitoring did not always lead to deeper reflection or independent listening.

3.1.2 Noticing Comprehension Gaps

Beyond supporting monitoring, transcript comparison enabled learners to identify specific comprehension difficulties. Analysis of the focus group and survey data revealed two main types of gaps i.e., perceptual gaps (misheard words) and lexical gaps (unknown vocabulary).

First, learners demonstrated clear awareness of misheard words, particularly those caused by rapid speech, accent variation, and phonological ambiguity. In both the focus group discussion and survey responses, learners explicitly identified causes of misunderstanding, stating that "the speaker talked too fast," "the accent was strange," or "I didn't know the vocabulary." This suggests that learners were not only recognizing errors but also becoming aware of the factors contributing to those errors. Survey responses provided numerous examples of mishearing. Learners reported hearing "affect" as "effect," "traits" as "trains", "anger" as "Angga", "poll" as "Paul," "heat" as "hate," and "derive" as "drive." In other cases, learners indicated that they could not recognize certain words at all until they saw the transcript, such as "affluence" or "mod cons." Confusion also occurred with proper names, such as hearing "Neil" as "Mill". These examples indicate that learners often relied on incomplete or inaccurate sound perception during initial listening.

Second, learners identified lexical gaps that were distinct from mishearing. In these cases, learners either did not know the word or were unable to recognize it in spoken form. Survey responses frequently included statements such as "I didn't know the vocabulary," with examples including "affluence," "mod cons," and "sophisticated." These responses indicate that some comprehension problems were related to lack of lexical knowledge rather than perceptual errors. The data suggest that learners were able to distinguish between these two types of gaps. Mishearing involved incorrectly interpreting sounds, while vocabulary gaps involved not recognizing or understanding the word at all. This distinction was reflected in learners' explanations of their listening difficulties.

Noticing these gaps also influenced learners' overall understanding of the listening material. Several learners reported that identifying a single misheard or unknown word changed their interpretation of the entire message. For example, recognizing "overconsumption" instead of "overpopulation" led to a clearer understanding of the topic. In addition to cognitive awareness, the noticing process was accompanied by affective responses. Many learners described the process as helpful and engaging rather than frustrating. They reported that the transcript "helped me see exactly which words I missed" and "made me realize my listening weaknesses." Some learners also noted that this awareness made them more careful and focused in subsequent listening tasks.

Furthermore, to these positive outcomes, the data revealed some variation in learners' responses to the noticing process. While many participants reported becoming more attentive

to key words, pronunciation, and overall meaning, others indicated little change in their listening approach, providing responses such as "nothing," "it's still the same," or "I don't know." A small number of learners also appeared to rely heavily on transcript support, with one participant stating that transcript use was "making me depend on transcript more than my understanding." These responses suggest that noticing comprehension gaps did not always lead to meaningful strategy development and, in some cases, may have encouraged dependence on textual support.

3.2 Reshaping Listening Processes Toward Gist-Oriented Understanding

The second research question examines how AI-generated transcripts reshaped learners' listening processes toward gist-oriented understanding. The findings show that transcript use helped learners move from word-level decoding to broader meaning construction. At the beginning of the listening process, many learners focused heavily on catching individual words. They often became confused when the speaker talked fast, used unfamiliar vocabulary, or pronounced words in ways they did not expect. After consulting the transcript, learners could connect unclear sounds with written forms and understand how specific words contributed to the overall message. This process helped them move beyond isolated word recognition and focus more on the main idea of the listening text.

The data indicate that AI-generated transcripts made unclear acoustic input more visible and manageable. Several learners reported that they initially heard the audio as unclear or confusing, especially when they encountered unfamiliar words. For example, one participant explained that the word "amygdala" was difficult to recognize at first because it was unfamiliar. After checking the transcript, the learner could identify the word and understand its meaning in context. Another learner described a similar experience with the word "enroll," stating that the transcript helped clarify the actual word and made the meaning easier to understand. These examples show that the transcript supported the transition from raw sound perception to meaningful word recognition.

Transcript use also helped learners revise their understanding of the whole listening text. In some cases, correcting one misheard word changed how learners understood the topic. For example, one learner initially confused "overconsumption" with "overpopulation." After reading the transcript, the learner realized that the speaker was discussing the difference between the two concepts. This correction did not only fix a single word-level error. It also helped the learner understand the main point of the listening material more accurately. This suggests that transcripts supported gist-oriented understanding by helping learners connect key words to the broader meaning of the text.

The findings further show that some learners began to use more global listening strategies after repeated transcript-supported activities. Instead of trying to understand every word, learners reported focusing first on the main topic, predicting the direction of the talk, and using the transcript later to confirm unclear parts. One participant stated that they tried to identify the main topic first and then predict where the discussion was going. Another participant explained that the best strategy was to listen first, check the transcript, and compare carefully to understand mistakes. These responses indicate that transcript use encouraged learners to combine bottom-up recognition with top-down prediction.

However, the shift toward gist-oriented listening was not uniform across all participants. While many learners became more confident and strategic, some still relied strongly on the transcript to understand the audio. This suggests that AI-generated transcripts can support more meaningful listening, but their benefits depend on how learners use them. When learners listen first, identify unclear parts, and use the transcript for confirmation, the transcript functions as a reflective tool. When learners depend on it too early, it may reduce independent listening effort. Overall, the findings suggest that guided transcript use can reshape listening from a word-catching activity into a more reflective, strategic, and meaning-oriented process.

DISCUSSION

4.1 AI as a Tool for Monitoring Listening

The findings indicate that AI-generated transcripts function as a tool for monitoring listening, particularly through learners' systematic comparison between auditory input and textual representation. According to Vandergrift & Goh (2012), such processes reflect core components of metacognitive theory, in which learners actively monitor, evaluate, and regulate their understanding during task performance. More recent work by Bozorgian et al. (2025) further confirms that metacognitive intervention plays a significant role in improving listening performance and reducing cognitive load.

In this study, the transcript served as an external reference that supported noticing, allowing learners to become aware of discrepancies between what they heard and what was actually said. By comparing audio with text, learners were able to identify misheard words and missing information, making otherwise unnoticed comprehension gaps visible. This aligns with the role of noticing in language learning, where awareness of mismatches between input and perception is a critical step in developing comprehension, as emphasized in Schmidt's (1990, 2001) Noticing Hypothesis.

At the same time, the transcript supported evaluation of comprehension, as learners used it to verify their understanding and assess the accuracy of their interpretations. Actions such as "checking what I missed" and "comparing with notes" indicate that learners were not only identifying errors but also judging the quality of their listening performance. This reflects metacognitive evaluation, a process that Vandergrift (2003) associates with skilled listeners, and which Graham (2017) identifies as central to effective listening strategy use in classroom contexts.

Importantly, the transcript functioned as a scaffold for these processes. Because spoken input is transient and difficult to revisit, learners often have limited opportunities to reflect on their comprehension. The availability of an AI-generated transcript transformed listening into a more stable and analyzable activity, enabling learners to repeatedly compare, check, and revise their understanding. As Vandergrift (2007) explains, metacognitive monitoring is often triggered when learners become aware of comprehension difficulties. This is further supported by Noushad et al. (2024), who show that explicit monitoring cues enhance learners' ability to evaluate their understanding. This interpretation is reinforced by research highlighting the role of AI-supported environments in fostering learner reflection, autonomy, and self-regulation (Mohebbi, 2025; Xu & Ouyang, 2022).

At the same time, the findings suggest that transcript-based scaffolding may involve potential trade-offs. While most learners reported increased attention to pronunciation, vocabulary, and listening accuracy, a small number perceived transcripts as reducing the need to attend carefully to the audio. This variation indicates that AI-generated transcripts do not automatically promote metacognitive engagement; rather, their effectiveness depends on how learners use them. Interestingly, many participants recommended listening to the audio independently before consulting the transcript and cautioned against relying on AI uncritically, describing transcripts as tools for verification rather than substitutes for listening. Such recommendations suggest that learners developed an awareness of both the benefits and limitations of AI support, reflecting an emerging capacity for self-regulated and strategic use of AI-assisted learning resources. Consistent with perspectives on self-regulated learning, transcript support appears most beneficial when learners actively employ it to evaluate and refine their listening rather than as a substitute for listening effort.

4.2 Reshaping Listening Processes Through AI Interaction

The findings suggest that AI-generated transcripts contributed to a restructuring of learners' listening processes by shifting their attention from word-level decoding toward meaning construction. Initially, many listening difficulties were associated with bottom-up processing challenges, including difficulties caused by rapid speech, accent variation, and phonological ambiguity. Such challenges are consistent with Field's (2008) argument that L2 listeners often struggle with segmenting continuous speech and with Cutler's (2012) observation that unfamiliar phonological patterns frequently result in misperception. As learners became aware of these comprehension difficulties through transcript-supported reflection, listening appeared to evolve from a process centered primarily on identifying words to one focused on understanding the message as a whole, reflecting the development of metacognitive awareness in listening (Goh, 2018).

This transformation can be understood through the interaction of bottom-up and top-down processing. Accurate recognition of lexical items provides the foundation for comprehension, yet successful listening also depends on the ability to draw upon contextual knowledge, prior experience, and inferential reasoning. By helping learners resolve decoding difficulties, transcripts appeared to facilitate a more effective coordination of these processing levels. This interpretation aligns with contemporary models of listening, which view comprehension as the outcome of multiple interacting cognitive processes rather than the result of word recognition alone (Rost, 2024).

The role of AI transcripts in this process can also be explained in terms of cognitive support. By reducing the attentional demands associated with decoding spoken input, transcripts enabled learners to allocate greater cognitive resources to higher-level comprehension processes. Rather than concentrating exclusively on identifying individual sounds or words, learners were able to engage more deeply with relationships between ideas, speaker intentions, and overall message coherence. Similar effects have been reported in multimodal learning research, where textual support facilitates segmentation and promotes deeper processing of audiovisual input (Lee et al., 2021; Perez, 2022). In this sense, the transcript

functioned not merely as corrective feedback but as a scaffold that supported meaning-oriented listening.

The findings also suggest broader implications for learner engagement. As learners gained greater awareness of how comprehension difficulties emerged and how they could be addressed, they reported becoming more attentive and focused during subsequent listening activities. This pattern is consistent with Efklides' (2011) model, which emphasizes the interaction of metacognition, motivation, and affect in self-regulated learning. Together with evidence from AI-assisted language learning research, the findings indicate that AI-generated transcripts may support listening development not only by improving comprehension but also by encouraging more reflective and self-regulated engagement with listening tasks.

Despite these positive outcomes, the findings suggest that the shift toward more strategic and meaning-oriented listening was not experienced uniformly across learners. While many participants reported greater attention to key information and overall meaning, others described little change in their listening approach. A small number of learners also reported increased reliance on transcript support, indicating that textual scaffolding may sometimes substitute for, rather than strengthen, independent listening processes. These variations suggest that the benefits of AI-generated transcripts depend not only on the availability of textual support but also on learners' ability to engage with it strategically. Therefore, transcript-based scaffolding may be most effective when combined with activities that encourage active listening and independent meaning construction.

CONCLUSION

The most important finding of this study is that most learners did not use AI-generated transcripts passively. Instead, many learners used them as reflective tools to compare, verify, and revise their listening comprehension. This showed that learners used transcripts to become more active, reflective, and strategic. They listened first, identified unclear parts, checked the transcript, and reflected on why comprehension had broken down. This process helped them recognize misheard words, notice unfamiliar vocabulary, revise inaccurate interpretations, and move from word-level decoding toward gist-oriented understanding. Thus, AI transcripts functioned not merely as textual support but as a metacognitive scaffold that made learners' listening problems visible and manageable.

The study contributes to EFL listening scholarship by confirming previous research on the importance of metacognitive monitoring and scaffolded input while also offering a more specific perspective on AI-mediated listening. Rather than treating AI transcripts only as tools for reducing cognitive load, this study shows how learners use them to monitor, regulate, and reshape listening processes through a listen–compare–reflect cycle. In this sense, the study contributes a process-oriented perspective that explains how AI support can facilitate the transition from fragmented decoding to more meaning-oriented listening.

This study is limited by its small and context-specific sample, involving 31 intermediate EFL learners in one university classroom over seven weeks. It is also limited by its qualitative case study design and reliance on self-reported data from questionnaires and focus group discussions. Future research should involve larger and more diverse samples, different proficiency levels, varied learning contexts, and additional methods such as classroom

observation, listening tests, or longitudinal tracking. Broader and deeper evidence would allow a more comprehensive understanding of AI transcript use and help formulate more effective pedagogical policies for EFL listening instruction.

REFERENCES

Adhikari, D. P., & Pandey, G. P. (2025). Integrating AI into higher education: transforming teachers' roles to boost student agency. *Educational Technology Quarterly*, *2025*(2), 151–168. https://doi.org/10.55056/etq.943

Amoozadeh, M., Daniels, D., Nam, D., Kumar, A., Chen, S., Hilton, M., Ragavan, S. S., & Alipour, M. A. (2024). Trust in Generative AI among students: An exploratory study. *SIGCSE 2024 - Proceedings of the 55th ACM Technical Symposium on Computer Science Education*, *1*, 67–73. https://doi.org/10.1145/3626252.3630842

Boltiziar, J., & Munkova, D. (2024). Emergency remote teaching of listening comprehension using YouTube videos with captions. *Education and Information Technologies*, *29*, 11367–11383. https://doi.org/10.1007/s10639-023-12282-7

Bozorgian, H., Muhammadpour, M., & Alehchali, A. B. (2025). The effects of metacognitive intervention and metacognitive intervention through dialogic interaction on L2 learners' listening comprehension and cognitive load. *Language Awareness, 1*–27. https://doi.org/10.1080/09658416.2025.2489937

Braun, V., & Clarke, V. (2006). Using thematic analysis in Psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chen, Z., Wei, W., & Zou, D. (2026). Generative AI technology and language learning: global language learners' responses to ChatGPT videos in social media. *Interactive Learning Environments*, *34*(2), 907–920. https://doi.org/10.1080/10494820.2025.2511248

Chua, W. C., Law, C. P., Ooi, Y. H., & Yunus, M. M. (2025). Leveraging AI-generated audio in the metaverse for enhanced English listening skills: an innovative approach. *Malaysia Journal of Invention and Innovation*, *4*(4), 71–90. https://doi.org/10.64382/mjii.v4i4.116

Creswell, J. W. (2013). *Qualitative inquiry and research design* (3rd ed.). SAGE Publications, Inc.

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications, Inc.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, *46*(1), 6–25. https://doi.org/10.1080/00461520.2011.538645

Field, J. (2008). Revising segmentation hypotheses in first and second language listening. *System*, *36*(1), 35–51. https://doi.org/10.1016/j.system.2007.10.003

Field, J. (2010). Listening in the language classroom. *ELT Journal*, *64*(3), 331–333. https://doi.org/10.1093/elt/ccq026

Field, J. (2013). Cognitive validity of the lecture-listening component of the IELTS. In L. Taylor & C. J. Weir (Eds.), *IELTS Collected Papers 2: Research in speaking and writing assessment* (pp. 77–121). Cambridge University Press. https://books.google.co.id/books?id=1dB2HxqLz_EC

Gass, S., Van Gorp, K., & Winke, P. (2019). Using different carrots: How incentivization affects proficiency testing outcomes. *Foreign Language Annals*, *52*(2), 216–236. https://doi.org/https://doi.org/10.1111/flan.12389

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, *28*(1), 55–75. https://doi.org/10.1016/S0346-251X(99)00060-3

Goh, C. C. M. (2018). Metacognition in second language listening. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1–7). https://doi.org/10.1002/9781118784235.eelt0572

Goh, C. C. M. (2023). Learners' cognitive processing problems during comprehension as a basis for L2 listening research. *System*, *119*, 103164. https://doi.org/10.1016/j.system.2023.103164

Graham, S. (2017). Research into Practice: Listening strategies in an instructed classroom setting. *Language Teaching*, *50*(1), 107–119. https://doi.org/10.1017/S0261444816000306

Hennink, M. M. (2014). *Focus group discussion: Understanding qualitative research*. Oxford University Press.

Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning*, *33*(5–6), 567–589. https://doi.org/10.1080/09588221.2019.1577898

Huang, J., Saleh, S., & Liu, Y. (2021). A review on Artificial Intelligence in education. *Academic Journal of Interdisciplinary Studies*, *10*(3), 206–217. https://doi.org/10.36941/AJIS-2021-0077

Hui, B., & Godfroid, A. (2021). Testing the role of processing speed and automaticity in second language listening. *Applied Psycholinguistics*, *42*(5), 1089–1115. https://doi.org/10.1017/S0142716420000193

James, E. E., Sampson, E. A., Usani, N. E., & Inyang, I. B. (2025). A principal component analysis of the factors influencing university students' trust in AI-based educational technologies. *African Journal of Advances in Science and Technology Research*, *18*(1), 111–141. https://doi.org/10.62154/ajastr.2025.018.010691

Jones, T. (2025). Bottom-Up and Top-Down listening. In M. Reed & J. M. Levis (Eds.), *The Handbook of Second Language Listening* (pp. 246–259). Wiley. https://doi.org/10.1002/9781394312375.ch18

Lee, P.-J., Liu, Y.-T., & Tseng, W.-T. (2021). One size fits all? In Search of the desirable caption display for second language learners with different caption reliance in listening comprehension. *Language Teaching Research*, *25*(3), 400–430. https://doi.org/10.1177/13621688198564

Liu, Y. (2020). Effects of metacognitive strategy training on Chinese listening comprehension. *Languages*, *5*(2), 1–22. https://doi.org/10.3390/languages5020021

Liu, Y., & Li, Y. (2025). AI-driven listening systems in language acquisition: Redefining auditory cognition in the intelligent era. *Discover Artificial Intelligence*, *6*(60), 1–32. https://doi.org/10.1007/s44163-025-00748-1

Luu, T. M. V., & Doan, Q. V. (2025). ChatGPT's impact on listening comprehension: Perspectives from Vietnamese EFL university learners. *Computer-Assisted Language Learning Electronic Journal, 26*(3), 43–63. https://doi.org/10.54855/callej.252633

Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *The All Ireland Journal of Teaching and Learning in Higher Education (AISHE-J)*, *8*(3), 3351–3364. http://ojs.aishe.org/index.php/aishe-j/article/view/335

Martín-Moncunill, D., & Martínez, D. A. (2025). Students' trust in AI and their verification strategies: A case study at Camilo José Cela University. *Education Sciences*, *15*(10), 1–16. https://doi.org/10.3390/educsci15101307

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

Milliner, B., & Dimoski, B. (2024). The effects of a metacognitive intervention on lower-proficiency EFL learners' listening comprehension and listening self-efficacy. *Language Teaching Research*, *28*(2), 679–713. https://doi.org/10.1177/13621688211004646

Mohebbi, A. (2025). Enabling learner independence and self-regulation in language education using AI tools: a systematic review. Cogent Education, 12(1), 2433814. https://doi.org/10.1080/2331186X.2024.2433814

Nguyen, N. L., & Nguyen, T. P. L. (2025). The impact of implementing AI-Generated audio transcriptions on English majors' cognitive load. *AsiaCALL Online Journal*, *16*(1), 140–158. https://doi.org/10.54855/acoj.251617

Noushad, B., Van Gerven, P. W. M., & de Bruin, A. B. H. (2024). Exploring the use of metacognitive monitoring cues following a diagram completion intervention. *Advances in Health Sciences Education, 29*(4), 1323–1351. https://doi.org/10.1007/s10459-023-10309-9

Perez, M. M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Language Teaching*, *55*(2), 163–192. https://doi.org/10.1017/S0261444821000501

Pitts, G., Rani, N., Mildort, W., & Cook, E.-M. (2025, June 16). *Students' reliance on AI in higher education: identifying contributing factors*. arXiv. https://doi.org/10.48550/arXiv.2506.13845

Richards, J. C. (1983). Listening comprehension: approach, design, procedure. *TESOL Quarterly*, *17*(2), 219–240. https://doi.org/10.2307/3586651

Rost, M. (2024). *Teaching and researching listening*. Routledge.

Safa, M. A., & Motaghi, F. (2024). Cognitive vs. metacognitive scaffolding strategies and EFL learners' listening comprehension development. *Language Teaching Research*, *28*(3), 987–1010. https://doi.org/10.1177/13621688211021821

Saito, K., Kachlicka, M., Suzukida, Y., Mora-Plaza, I., Ruan, Y., & Tierney, A. (2024). Auditory processing as perceptual, cognitive, and motoric abilities underlying successful second language acquisition: Interaction model. *Journal of Experimental Psychology: Human Perception and Performance*, *50*(1), 119–138. https://doi.org/10.1037/xhp0001166

Sakuta, D. A., Jusriati, J. & Sari, P. (2026). From scroll to skill 2.0: Integrating YouTube and AI tools to enhance students' extensive listening skills. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, *11*(1), 850–872. https://doi.org/10.23969/jp.v11i01.38933

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3–32). Cambridge University Press.

Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Susanto, S. & Nanda, D. S. (2024). Harmonizing sound and script: integrating phonetic symbols within the lexical approach to improve EFL listening skill. *English Review: Journal of English Education*, *12*(1), 177–192. https://doi.org/10.25134/erjee.v12i1.9224

Teng, M. F., & Teng, L. S. (2025). Metacognition, self-regulation, and self-efficacy in second and foreign language learning. *International Journal of TESOL Studies*, *7*(1), 1–3. https://doi.org/10.58304/ijts.250624

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457–479. https://doi.org/10.1093/applin/ams074

Vandergrift, L. (2003). Orchestrating strategy use: Toward a model of the skilled second language listener. *Language Learning*, *53*(3), 463–496. https://doi.org/10.1111/1467-9922.00232

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, *40*(3), 191–210. https://doi.org/10.1017/S0261444807004338

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.

Xia, L., Shen, K., Sun, H., An, X., & Dong, Y. (2025). Developing and validating the student learning agency scale in generative artificial intelligence (AI)-supported contexts. *Education and Information Technologies*, *30*, 13999–14021. https://doi.org/10.1007/s10639-024-13137-5

Xiao, Y. (2025). The impact of AI-driven speech recognition on EFL listening comprehension, flow experience, and anxiety: a randomized controlled trial. *Humanities and Social Sciences Communications*, *12*, Article 425, 1–14. https://doi.org/10.1057/s41599-025-04672-8

Xu, W., & Ouyang, F. (2022). The Application of AI technologies in STEM education: A systematic review from 2011 to 2021. *International Journal of STEM Education*, *9*, Article 59, 1-20. https://doi.org/10.1186/s40594-022-00377-5

Yin, R. K. (2018). *Case study research and applications: Designs and methods* (6th ed.). SAGE Publications, Inc.

Zeng, Y., & Goh, C. C. M. (2018). A self-regulated learning approach to extensive listening and its impact on listening achievement and metacognitive awareness. *Studies in Second Language Learning and Teaching, 8*(2), 193–218. https://doi.org/10.14746/ssllt.2018.8.2.2

Zheng, J., Hao, L., Lu, K., Garg, A., Reese, M., Yap, M. J., Wang, I. J., Wu, X., Huang, W., Hoffman, J., Kelly, A., Le, M., Zhang, R., Lin, Y., Faayez, M., & Liu, A. (2025). *Do students rely on AI? Analysis of student-ChatGPT conversations from a field study*. arXiv. https://doi.org/10.48550/arXiv.2508.20244