

STUDENTS' RESPONSES TO SUPPLEMENTARY AI-AIDED LISTENING MATERIALS

Ho Thi Nguyen¹, Nguyen Thi Quyet²

Received: 09 March 2026/ Accepted: 20 July 2026/ Published: 30 July 2026

<https://doi.org/10.70117/hdujs.E11.7.2026.1182>

Abstract: *This study investigates EFL learners' responses to AI-aided listening materials at a private English center in Vietnam. Thirty-five B1-level students participated in a 12-week course integrating AI-supported materials with regular instruction. Data from a questionnaire and interviews revealed generally positive perceptions of the AI-aided materials, which helped reduce listening difficulty and anxiety. AI-assisted videos that combined audio and visual support helped students understand better and feel more confident when they were listening. Overall, the findings suggest that AI-aided listening materials can effectively support EFL instruction when thoughtfully integrated into teacher-guided practice.*

Keywords: *Artificial intelligence, EFL listening, AI-aided listening materials, classroom integration.*

1. Introduction

1.1. Background of the study

Listening is often considered one of the most challenging skills for learners of English as a foreign language (EFL) because spoken input must be processed in real time and cannot easily be reviewed. In many EFL classrooms in Vietnam, listening instruction mainly relies on pre-recorded audio materials followed by comprehension questions. This product-oriented approach often focuses on correct answers rather than helping learners develop listening processes and strategies [4], [14]. In the Vietnamese context, recent studies have shown growing interest in the application of artificial intelligence (AI) in language education. For example, Pham and Dang [12] found that both teachers and students in Vietnam perceive AI tools as useful for supporting language teaching and learning. Similarly, Nguyen and colleagues [11] reported increasing use of AI in foreign language education, highlighting its role in enhancing learning engagement and instructional practices.

Recent developments in Artificial Intelligence (AI) provide new opportunities for teachers to design more flexible listening materials [8], [13]. In this study, generative AI tools were used to create contextualized video-based listening materials with features such as visual context and captions. These supplementary materials were integrated into

¹ Faculty of Foreign Languages, Vinh university; Email: studyieltswithme9@gmail.com

² Faculty of Foreign Languages, Hong Duc University

classroom listening lessons under teacher guidance. Such multimedia support aligns with principles of multimedia learning, which emphasize the integration of auditory and visual information to support comprehension [9]. Despite growing interest in technology-enhanced language learning, limited research has examined students' responses to teacher-developed AI-aided listening materials in classroom settings. Therefore, this study investigates EFL students' perceptions of these materials, focusing on their cognitive and affective responses during listening activities [1], [6].

Specifically, this study addresses the following research question: *How do EFL students respond to supplementary AI-aided listening materials in classroom-based instruction?*

1.2. Theoretical framework

1.2.1. AI in teaching listening skills

Recent developments in artificial intelligence have created new opportunities to support language teaching. In education, AI is increasingly viewed as a tool that assists teachers in planning, designing, and adapting learning materials rather than replacing them [8], [11], [13]. AI technologies can generate texts, scripts, and multimedia resources, providing richer multimodal input for language learning [5]. In listening instruction, AI tools can help teachers create contextualized scripts, videos, and audio through generative platforms and text-to-speech technologies [11], [12]. These tools allow teachers to adjust speech rate, language complexity, and contextual support according to learners' proficiency levels, making listening input more accessible. From a sociocultural perspective, technological tools are effective only when implemented through guided instruction [7], [15]. In this study, AI was used for instructional preparation: the teacher employed generative AI (Google Gemini) to create supplementary listening videos with audio, visuals, and captions based on textbook content. These materials were integrated into teacher-led activities, with AI serving as a tool to support the design of multimodal listening materials.

1.2.2. Cognitive aspects of listening comprehension and the Theory of Multimedia Learning

Listening comprehension in a foreign language imposes substantial cognitive demands because spoken input is transient and must be processed in real time. Learners must decode phonological signals, access lexical meaning, and construct overall understanding simultaneously [4], [14]. When lower-level processes are not automatized, this simultaneity may lead to cognitive overload. From a cognitive perspective, successful listening involves the interaction of bottom-up processing (e.g., phoneme recognition, word segmentation, syntactic parsing) and top-down processing (e.g., using background knowledge, contextual cues, and expectations to construct meaning). Effective instructional design should therefore support both processes while preventing cognitive overload. The Cognitive Theory of Multimedia Learning (CTML) developed by Mayer [9], [10] provides a theoretical explanation for how multimedia-enhanced materials may support listening comprehension. According to CTML, learners process information through dual auditory and visual channels, each with limited capacity. Effective learning occurs when instructional design supports the active selection, organization, and integration of information across these channels. Multimedia materials that combine audio with meaningful visual support can therefore help manage cognitive load and facilitate

comprehension. Within this framework, multimedia materials that combine spoken input with meaningful visual support (e.g., contextualized video scenes, signaling cues, or limited captions) may distribute processing demands across channels and reduce overload in the auditory channel. Visual context can support top-down processing by activating schemata and clarifying situational meaning, while clear and well-segmented audio can facilitate bottom-up decoding.

1.2.3. Affective Filter and Self-Efficacy in Second Language Listening

In addition to cognitive factors, affective variables play a critical role in learners' engagement with second language listening (L2 listening). According to the Affective Filter Hypothesis, negative emotional states such as anxiety and low confidence can hinder learners' access to comprehensible input [6]. As a consequence, learners experiencing high listening anxiety may disengage from listening tasks even when instructional support is available. Closely related to this, self-efficacy refers to learners' beliefs in their ability to successfully perform specific tasks [1]. In the context of listening, learners with higher self-efficacy are more likely to persist in challenging listening activities and approach tasks with confidence. Conversely, learners with low self-efficacy may experience heightened anxiety and avoidance behaviors. From this perspective, AI-aided listening materials may help reduce anxiety and enhance self-efficacy by allowing learners to control task difficulty, revisit challenging input, and experience repeated success. Therefore, affective filter and self-efficacy theories offer an important lens for interpreting learners' emotional responses to AI-supported listening materials.

2. Methodology

2.1. Participants and setting

The participants were 35 EFL students (15 males and 20 females) enrolled in a private English center in Vietnam. All students were at the B1 level according to the CEFR and were taking a listening course as part of their regular program. They were approximately 16-18 years old and had similar English learning backgrounds. All 35 students completed the questionnaire, and five students voluntarily participated in follow-up semi-structured interviews. Participation was voluntary, and informed consent was obtained from all participants.

2.2. Implementation procedure

The AI-aided listening intervention was implemented as supplementary support alongside the prescribed course textbook, *Listening for IELTS* (Collins). The supplementary materials were developed using Google Gemini to generate video-based listening content, including scripted dialogues and contextual visual elements. All scripts were carefully reviewed and refined to ensure alignment with lesson objectives and textbook listening tasks. Each video was produced in two versions (with and without captions) to provide flexible instructional scaffolding and enhance student engagement. By incorporating audiovisual context, the videos enabled learners to activate background knowledge and become familiar with key vocabulary prior to listening to the original audio, thereby supporting comprehension.

Take one example: in Lesson 1 of Unit 4 ("Family Structure"), the textbook included five activities: (1) matching words with family trees, (2) completing words, (3) matching verbs with prepositions, (4) completing sentences, and (5) listening for missing words. However, the original material provided only one fixed audio recording (CD1-21), lasting 57 seconds, without visual support. To enrich the lesson, four AI-generated videos were created: two versions (captioned and non-captioned) supporting activities 1-4 and two versions supporting activity 5. These videos incorporated scripted dialogues and contextual visuals simulating communicative situations related to family structure.

In classroom implementation, first, the teacher introduced the lesson topic, clarified objectives, and activated students' background knowledge through discussion and vocabulary review. Then, the teacher played the AI-generated video (initially without captions), and students watched attentively to build contextual understanding. Next, students completed the textbook listening activities while listening to the original audio recording, with the teacher monitoring progress and replaying segments when necessary. Finally, the teacher reviewed answers and replayed selected captioned video segments to clarify misunderstandings and reinforce key language features. Students subsequently reflected on their performance and noted corrections.

2.3. Research instruments, data collection, and analysis

This study employed a mixed-methods descriptive design to explore EFL students' responses to supplementary AI-aided listening materials by combining quantitative questionnaire data with qualitative insights from semi-structured interviews [2]. Data were collected using a Likert-scale questionnaire and interviews. The questionnaire, consisting of eight items, was used to examine students' perceptions of the AI-aided listening materials. Section A focused on cognitive aspects of listening comprehension based on the Cognitive Theory of Multimedia Learning [9], [10], while Section B examined learners' emotional responses, such as confidence and anxiety, grounded in Affective Filter Hypothesis [6]. All items were rated on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) [3].

Data were collected at the end of the 12-week course. The questionnaire was administered to 35 students, and five students were randomly selected for semi-structured interviews, each lasting about five minutes. Quantitative data was analyzed using descriptive statistics (means and standard deviations), while interview data was transcribed and analyzed thematically to support and explain the questionnaire results.

3. Findings and discussions

3.1. Students' Responses to AI-Aided Listening Materials Based on Questionnaire Data

3.1.1. Cognitive Support of Multimedia learning

The items were theoretically grounded in the Cognitive Theory of Multimedia Learning (CTML) proposed by [9], [10]. Specifically, the items were designed to measure students' perceived cognitive support when engaging with AI-generated listening videos.

Table 1. *Students' Perceptions of the Cognitive Support of AI-Aided Multimedia Listening Materials*

Item	Statement	Mean (M)	Standard Deviation (SD)
A1	The combination of spoken audio and visual scenes of the AI-aided materials helped me understand the listening content more easily.	4.27	0.60
A2	The visual context supported my understanding of the spoken information in the videos.	4.34	0.55
A3	The integration of audio and visuals of the AI-aided materials helped me identify important information in the listening tasks.	4.38	0.53
A4	The audio-visual format of the AI-aided materials made the listening content clearer than audio-only materials.	4.25	0.57
Overall	Cognitive support (A1-A4)	4.31	0.56

As shown in Table 1, all four items received relatively high mean scores ranging from 4.25 to 4.38, indicating that students generally agreed that the multimedia format supported their listening comprehension. Item A1 ($M = 4.27$, $SD = 0.60$) shows that students felt the combination of spoken audio and visual scenes helped them understand the listening content more easily. This suggests that presenting information through both auditory and visual channels allowed learners to process the listening input more effectively. The visual scenes may have provided contextual clues that helped students understand the meaning of the spoken content. Item A2 ($M = 4.34$, $SD = 0.55$) also received a high mean score, indicating that students believed the visual context supported their understanding of the spoken information in the videos. Visual elements such as actions, settings, or background context may have helped learners interpret the meaning of the listening content more accurately. Item A3 obtained the highest mean score ($M = 4.38$, $SD = 0.53$), suggesting that students particularly valued the integration of audio and visual elements in helping them identify important information during listening tasks. This result indicates that the multimedia format may have helped students focus their attention on key points while listening. Item A4 ($M = 4.25$, $SD = 0.57$) also received a high rating, showing that students generally agreed that the audio-visual format made the listening content clearer than traditional audio-only materials.

The overall mean score for this construct was 4.31 ($SD = 0.56$), indicating that students perceived the AI-aided listening materials as providing strong cognitive support. The relatively low standard deviation values also suggest that students' responses were fairly consistent. Overall, these findings indicate that the multimedia features of the AI-generated listening materials helped students better understand the listening content and identify key information during listening activities.

3.1.2. Affective Responses

The Affective Responses construct examined students' emotional reactions to the AI-aided listening materials, particularly in terms of listening anxiety, confidence, and comfort during listening activities. This section consisted of four items (B1-B4) that were adapted from established measures of foreign language listening anxiety. The construct is theoretically grounded in Affective Filter Hypothesis [6], which emphasizes the role of emotional variables such as anxiety and confidence in second language acquisition. All items were rated on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree).

Table 2. *Students' Affective Responses to AI-Aided Multimedia Listening Materials*

Item	Statement	Mean (M)	Standard Deviation (SD)
B1	I enjoy listening to these AI-aided materials because of the colorful and authentic videos.	3.98	0.71
B2	I feel more confident when listening to these AI-aided materials with visuals.	4.05	0.66
B3	I feel calm when I am sure that I can understand these materials.	4.01	0.69
B4	I can overcome my nervousness when listening to these materials with the text (captions)	4.07	0.65
Overall	Affective Responses (B1-B4)	4.03	0.68

As shown in Table 2, the overall mean score for this construct was 4.03 (SD = 0.68), indicating generally positive emotional responses toward the AI-aided listening materials.

Item B4 recorded the highest mean score (M = 4.07, SD = 0.65), suggesting that captions and textual support helped students reduce nervousness during listening activities. This indicates that the availability of captions may have provided learners with additional support when they encountered difficulties in understanding spoken input. Item B2 also received a relatively high mean score (M = 4.05, SD = 0.66), indicating that the visual elements in the AI-aided materials helped increase students' confidence when completing listening tasks. Similarly, Item B3 (M = 4.01, SD = 0.69) shows that students tended to feel calmer when they were confident that they could understand the listening materials. This suggests that comprehension plays an important role in reducing anxiety during listening activities. Item B1 recorded the lowest mean score (M = 3.98, SD = 0.71), although the score still reflects a generally positive level of enjoyment toward the AI-aided materials. The slightly higher standard deviation for this item suggests that students' enjoyment levels varied more compared to the other affective responses. Overall, the relatively moderate standard deviation values across the items (0.65-0.71) indicate some variation in students' emotional reactions. While most students reported positive affective experiences, individual differences in listening anxiety or confidence may have influenced their responses. These

findings support the Affective Filter Hypothesis [6], which suggests that reducing anxiety can facilitate language acquisition by lowering emotional barriers to input processing. In addition, the positive ratings related to confidence are also consistent with Self-Efficacy Theory [1], which emphasizes the importance of learners' perceived ability in maintaining engagement and persistence in challenging tasks such as second language listening.

3.2. Post-experiment interview results

After the experiment, to further understand students' views on learning AI-aided materials provided, two questions were given to five randomly selected students. The interview data were analyzed thematically, focusing on cognitive support and affective response.

3.2.1. Cognitive support from audio-visual integration

All five participants (5/5) reported that the combination of audio and visual elements made the listening content easier to understand. Four out of five students (4/5) emphasized that visual context helped them infer meaning when they were unable to catch certain words in the audio. As one participant explained *"When I don't hear clearly, I can look at the video and guess what the speaker is talking about"* (S1, personal communication, February 3, 2026). This suggests that elements functioned as meaningful contextual support rather than merely decorative features.

In addition, all participants (5/5) indicated that visuals helped them identify key information in listening tasks. One student share, *"The video helps me see the important parts, so I know what I should focus on"* (S2, personal communication, February 3, 2026). This reinforces the role of multimodal input in facilitating more efficient processing auditory information.

3.2.2. Affective responses: confidence, engagement, and anxiety

Regarding emotional response, most students (4/5) reported increased confidence when engaging with AI-aided listening materials. The availability of visual context reduced uncertainty and supported comprehension. As one student stated, *"When I can see what is happening in the video, I feel more confident because I understand the context better"* (S2, personal communication, February 3, 2026).

All participants (5/5) agreed that the use of colorful and contextualized videos made listening activities more engaging compared to traditional audio-only materials, although slight variations in enjoyment reflected individual preferences. Importantly, three participants (3/5) reported reduced anxiety during listening tasks. They noted that the ability to watch and listen simultaneously, along with teacher-controlled repetition, lowered the pressure associated with one-time listening. One participant commented, *"When I listen and watch at the same time, I feel more confident because the visuals support what I hear"* (S4, personal communication, February 3, 2026). Additionally, captions were identified as a useful feature by four participants (4/5), as they allowed learners to verify their understanding and feel less nervous. As another student explained, *"When I see the captions, I feel less nervous because I can check if I heard correctly"* (S5, personal communication, February 3, 2026).

3.3. Discussion

The findings indicate that students perceived the supplementary AI-aided listening materials as both cognitively and affectively supportive in teacher-guided classroom instruction. The interview data supported the questionnaire results, providing triangulation and strengthening the reliability of the findings. From a cognitive perspective, the high mean scores for Items A1-A4 ($M = 4.25-4.38$) suggest strong agreement that the combination of audio and visual input supported listening comprehension. Interview responses confirmed that visual context helped students infer meaning when they missed certain words. These results are consistent with the Cognitive Theory of Multimedia Learning [9], [10], which suggests that coordinating auditory and visual information can reduce cognitive load and support comprehension. The affective results also showed positive responses (overall $M = 4.03$), particularly in terms of increased confidence and reduced nervousness. Interview participants explained that captions and visual context helped them check understanding and feel more confident when listening. This supports the Affective Filter Hypothesis [6] and reflects aspects of Self-Efficacy Theory [1].

Finally, interview data highlighted the important role of teacher mediation. Students emphasized that teacher pauses, repetition, and explanations helped them process the listening materials. This suggests that AI-generated materials are most effective when used as supportive tools within structured teacher-guided instruction rather than as independent learning systems.

4. Conclusion

This study examined EFL students' responses to supplementary AI-aided listening materials used in teacher-led classroom instruction at a private English center in Vietnam. The findings show generally positive perceptions, with students reporting that audio, visuals, captions helped them understand listening content more easily and feel more confident during activities. Students also emphasized the importance of teacher guidance in effectively using these materials. However, the study is limited by its small sample size and reliance on self-reported data. Future research could involve larger samples and performance-based measures to examine the impact of AI-aided materials on listening proficiency. Overall, the results suggest that AI-generated listening materials can support EFL instruction when integrated with teacher guidance.

References

- [1] Bandura, A. (1997), *Self-efficacy: The exercise of control*. W. H. Freeman.
- [2] Brown, J. D. (2014), *Mixed methods research for TESOL*. Edinburgh University Press.
- [3] Dörnyei, Z., & Taguchi, T. (2010), *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). Routledge.
- [4] Field, J. (2008), *Listening in the language classroom*. Cambridge University Press.
- [5] Godwin-Jones, R. (2023), *Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse*. *Language Learning & Technology*, 27(2):6-27. <https://doi.org/10.64152/10125/73501>
- [6] Krashen, S. D. (1985), *The input hypothesis: Issues and implications*. Longman.

- [7] Lantolf, J. P., & Thorne, S. L. (2006), *Sociocultural theory and the genesis of second language development*. Oxford University Press.
- [8] Law, L. (2024), *Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review*. Computers and Education Open, 6:1-13. <https://doi.org/10.1016/j.caeo.2024.100174>.
- [9] Mayer, R. E. (2009), *Multimedia learning*. Cambridge University Press.
- [10] Mayer, R. E. (2021), *Multimedia learning* (3rd ed.). Cambridge University Press.
- [11] Nguyen, H. D., Pham, N. T., & Dinh, T. H. (2025), *Application of artificial intelligence in foreign language teaching and learning: A comparative study*. Ho Chi Minh City University of Education Journal of Science, 22(11):2060-2072. <https://doi.org/10.54607/hcmue.js.22.11.4752>
- [12] Pham, T. N., & Dang, T. X. (2025), *An investigation into the application of artificial intelligence for language teaching and learning in Vietnam*. Vietnam Journal of Education, 9(2):265-283. <https://doi.org/10.52296/vje.2025.557>
- [13] Pokrivčáková, S. (2019), *Preparing teachers for the application of AI-powered technologies in foreign language education*. Journal of Language and Cultural Education, 7(3):135-153. <https://doi.org/10.2478/jolace-2019-0025>
- [14] Vandergrift, L., & Goh, C. C. M. (2012), *Teaching and learning second language listening: Metacognition in action*. Routledge.
- [15] Vygotsky, L. S. (1978), *Mind in society: The development of higher psychological processes*. Harvard University Press.