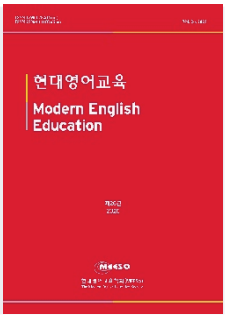
Modern English Education, Vol. 26, (2025) pp. 446-459
doi: https://doi.org/10.18095/meeso.2025.26.1.446

Modern English Education

ISSN: 1598-0782 / e-ISSN: 2586-6141
http://journal.meeso.or.kr



Opportunities, Challenges, and Perceived Effectiveness of AI-Generated Listening Materials in a University EFL Context

Jongbum Ha (Kumoh National Institute of Technology)

Received: 30 September 2025
Revised: 15 October 2025
Accepted: 24 October 2025

Ha, Jongbum. (2025). Opportunities, challenges, and perceived effectiveness of AI-generated listening materials in a university EFL context. *Modern English Education*, *26*, 446-459.

Keywords
AI, EFL, listening, students' perception, 인공지능, 영어 듣기, 학생 인식

Jongbum Ha
Professor
School of General Education
Kumoh National Institute of Technology
jbha@kumoh.ac.kr

*This paper was supported by Research Fund, Kumoh National Institute of Technology (2024-2026).

Abstract

This study explored university EFL learners' perceptions of AI-generated listening materials, emphasizing their instructional value, challenges, and effectiveness in enhancing listening skills. Conducted with 114 students in a Korean university EFL course, the research focused on three key questions regarding the opportunities, challenges, and effectiveness of these materials. Data were collected through a 12-item perception survey and open-ended responses during the final weeks of the semester.

The quantitative results indicated generally positive attitudes toward the structured integration of passages, dialogs, and summaries. Students acknowledged that repeated exposure helped reinforce vocabulary learning and supported comprehension development. They particularly valued the adjustable audio playback speed and the option to revisit material, which fostered autonomous learning. However, significant challenges were identified, including unnatural intonation patterns, robotic voice delivery, and pronunciation inaccuracies. These issues were especially problematic for lower-proficiency students, who rely more on prosodic cues for comprehension. The findings suggest that AI-generated materials can significantly enhance EFL listening instruction when implemented with thoughtful pedagogical design, targeted guidance, and attention to the quality of synthetic speech. The study also discusses implications for EFL curriculum development and offers recommendations for future research.

INTRODUCTION

The emergence of generative artificial intelligence (AI) technologies has opened new avenues for English as a Foreign Language (EFL) listening instruction. Tools such as ChatGPT, text-to-speech (TTS), and speech-to-text (STT) applications enable instructors to rapidly develop structured, customizable materials. When thoughtfully implemented, AI-generated content has the potential to enhance listening comprehension, increase learner engagement, and promote learner autonomy through relevant topics and flexible delivery options (Baxramova, 2025; Young & Shishido, 2023).

Despite these promising applications, significant concerns persist regarding the linguistic quality and delivery of AI-generated speech. Recent studies have documented various limitations including unnatural phrasing, monotonous intonation

patterns, and robotic voice quality (Meniado, 2023; Mak, 2021), factors that may impede comprehension and diminish student engagement. Furthermore, learner perceptions appear to vary considerably based on proficiency levels and their assessment of content clarity, naturalness, and pedagogical utility (Chan & Hu, 2023; Ngo et al., 2024).

Although several recent studies have examined AI-generated listening materials in language learning contexts, much remains to be understood about their instructional role. In particular, questions remain about how learners of different proficiency levels experience such materials, and how these tools function when integrated into structured classroom settings. To address these issues, the present study explores student perceptions of AI-generated listening materials in a university EFL course.

LITERATURE REVIEW AND RESEARCH QUESTIONS

AI-Generated Listening Materials in EFL Education

The advent of AI systems such as ChatGPT has enabled the generation of personalized, theme-based dialogs and passages that can be closely aligned with specific course objectives, thereby enhancing both relevance and learner engagement (Young & Shishido, 2023). Research by Dorgham (2024) demonstrated that AI-assisted listening tasks not only improved learners' proficiency but also reduced their anxiety levels, with the AI-supported group significantly outperforming control groups on both skill acquisition and affective measures. These findings align with those reported by Baxramova (2025), who observed enhanced comprehension and motivation when learners engaged with AI-generated listening materials specifically designed for topical relevance.

Collaborative approaches, particularly student-generated AI listening texts, have shown considerable promise for fostering learner autonomy and engagement (Jung, 2025). Nevertheless, certain limitations persist in current AI-generated materials. Both Meniado (2023) and Derakhshan (2025) have identified concerns regarding unnatural prosody, pragmatic inaccuracies, and limited contextual nuance that can negatively impact comprehension and perceived authenticity. These shortcomings suggest that a hybrid approach—one that combines the adaptability of AI-generated content with careful instructor-led refinement—may be most effective in leveraging AI's benefits while maintaining linguistic and pragmatic quality.

AI-generated listening materials thus offer scalable and customizable solutions for EFL listening instruction. However, their effectiveness remains heavily dependent on thoughtful pedagogical integration and systematic post-editing to ensure naturalness and contextual appropriateness.

Text-to-Speech (TTS) and Speech-to-Text (STT) Technologies in Second Language Listening

Text-to-Speech (TTS) technology provides learners with accessible listening input featuring adjustable playback speeds, enabling them to manage pacing according to their proficiency level and review content as needed. However, despite progress in synthesis quality, current systems still lack natural prosody and expressiveness, resulting in speech that sounds monotonous and less engaging (Peiró-Lilja et al., 2022; Yao et al., 2024). Such limitations may reduce learners' motivation and attentional focus during listening practice.

Speech-to-Text (STT) technology, based on automatic speech recognition, provides immediate transcripts that help learners notice pronunciation errors and improve articulation accuracy (Ngo et al., 2024). Nevertheless, recognition errors frequently occur with accented or low-proficiency speech, which can mislead learners. Therefore, teacher mediation remains crucial for interpreting inaccurate transcriptions and ensuring effective pedagogical use (Southwell et al., 2022).

When implemented together, TTS and STT technologies create a comprehensive input-output feedback loop: TTS provides natural-like speech input for comprehension practice, while STT facilitates productive practice with immediate corrective feedback. The pedagogical value of these technologies ultimately depends on the selection of high-quality processing engines and their integration within well-structured learning tasks designed to maximize their instructional impact.

Learner Perception and Affective Factors in AI-Integrated Listening Instruction

Student acceptance of AI-driven listening tools demonstrates strong correlations with perceived usefulness and ease of use,

consistent with predictions from the Technology Acceptance Model (Davis, 1989). Research indicates that when learners are able to control pacing and revisit materials at will, both engagement and motivation levels increase significantly (Chan & Hu, 2023). Additionally, adaptive systems that adjust difficulty levels in real time can effectively help manage cognitive load, particularly in independent study or blended learning contexts (Luque de la Rosa et al., 2025).

Despite these documented advantages, concerns regarding authenticity continue to emerge. Both Meniado (2023) and Aziza (2025) report that some learners find synthetic speech less engaging due to its mechanical quality, while others express discomfort about the reduction in human interaction or raise ethical considerations regarding AI use in education. Research suggests that appropriate teacher guidance can help mitigate these concerns by positioning AI tools as supplements to, rather than replacements for, human instruction.

Collectively, learner perception studies indicate that a balanced pedagogical approach—one that merges AI's inherent adaptability and efficiency with the authenticity and interpersonal benefits of human interaction—is most likely to maximize both skill development outcomes and positive affective responses among EFL learners.

The Present Study

While recent studies have explored AI-generated listening materials, important gaps remain. Existing research has focused primarily on general perceptions rather than systematic course integration, and insufficient attention has been given to how students at different proficiency levels experience these materials differently. Addressing these gaps, the present study examines AI-generated listening materials within a sustained, course-integrated EFL context, drawing on student perceptions to formulate the following three research questions:

Research Question 1. What opportunities do AI-generated materials offer in developing an EFL listening course?

Research Question 2. What challenges do learners perceive in using AI-generated listening materials?

Research Question 3. How do students perceive the effectiveness of AI-generated listening materials in improving their English listening skills?

METHOD

Participants

This study was conducted during the spring semester of 2025 with 114 students enrolled in three English listening courses taught by the researcher. Reflecting the university's engineering-focused profile, most participants were engineering majors. The course, a three-credit blended class, primarily targeted second-year students but was open to all academic years and counted as either a required or elective course depending on the student's major. The sample comprised 88 male and 26 female students, distributed by academic year as follows: 6 first-year, 74 second-year, 15 third-year, and 19 fourth-year students. Baseline English proficiency was self-assessed on a five-point scale ranging from Very Limited to Excellent. The results showed 11 students (9.6%) identifying as Very Limited users, 51 (44.7%) as Basic users, 43 (37.7%) as Moderate users, 6 (5.3%) as Good users, and 3 (2.6%) as Excellent users. These demographic and proficiency distributions are summarized in Table 1.

Materials and Instruments

The AI-integrated listening materials were developed using twelve engaging topics generated with assistance from the free version of ChatGPT, specifically curated for second-year EFL university students. Each topic originated from an authentic online article, which was adapted into three versions: a five-paragraph passage, a conversational dialog between two speakers, and a concise 120-word summary. This multi-format approach aimed to reinforce comprehension through repeated exposure to identical content presented in varied structural formats.

Audio production employed text-to-speech (TTS) technology via PlayHT, chosen for its natural-sounding voices and support for Speech Synthesis Markup Language (SSML) and adjustable parameters. To ensure efficiency and consistency, all recordings used PlayHT's default settings without SSML customization. The voice assignment strategy alternated

between two default American English voices for the passages (female for odd-numbered lessons, male for even-numbered) and incorporated both voices for the dialogs to simulate authentic exchanges.

TABLE 1
Demographic Data (n = 114)

Class	Gender	Year	English proficiency
Class 1 (*n* = 43)	Male (*n* = 88)	1st Year (*n* = 6)	Very Limited (*n* = 11)
Class 2 (*n* = 31)	Female (*n* = 26)	2nd Year (*n* = 74)	Basic (*n* = 51)
Class 3 (*n* = 40)		3rd Year (*n* = 15)	Moderate (*n* = 43)
		4th Year (*n* = 19)	Good (*n* = 6)
			Excellent (*n* = 3)

Note. Proficiency scale: 1. Very Limited (struggle with even simple English); 2. Basic (understand slow/simple speech, miss details); 3. Moderate (follow familiar topics, difficulty with fast/new input); 4. Good (understand most speech incl. lectures, occasional repetition needed); 5. Excellent (understand nearly all speech, including fast/complex)

Summary sections followed a rotation of four voices to provide balanced exposure to accent varieties:

- Lessons 1, 5, 9: American English female
- Lessons 2, 6, 10: American English male
- Lessons 3, 7, 11: British English female
- Lessons 4, 8, 12: British English male

Initially, *Papago* TTS was used for summaries under the assumption that slightly less clear audio would promote closer attention to linguistic details. However, classroom observations and midterm results revealed comprehension difficulties that hindered rather than enhanced learning. The summaries were subsequently regenerated with *Luvvoice*, which offered clearer, more intelligible output.

At first, only normal-speed audio (1.0x) was provided. After midterm results indicated challenges with native-speed comprehension, slower-speed versions (0.8x) were added using an online audio converter, allowing students to choose their preferred pace.

All materials—including both speed options—were hosted on a dedicated Google Sites course page, where students accessed weekly lessons, streamed audio, and submitted assignments.

The twelve topics and their source articles were as follows:

1. Politics and Governance – *Children and Youth Still Out of School*
2. Economics and Business – *The Future of the Semiconductor Industry*
3. Science and Technology – *How Technology Is Shaping the New Normal*
4. Environmental Issues – *Long-term Air Pollution Exposure Increases Asthma Risk*
5. Health and Medicine – *Second Round of Polio Vaccination in the Gaza Strip*
6. Education and Learning – *The Biggest Education Trends of the Next 10 Years*
7. Culture and Society – *How Social Media Turns Societies Upside Down*
8. History and World Events – *The History of Trick-or-Treating and How It Became a Halloween Tradition*
9. Arts and Entertainment – *The Evolution of Classical Music Sampling in K-pop*
10. Sports – *Reasons Why Soccer Is the Most Popular Sport in the World*
11. Travel and Tourism – *Why Travel Is Important to Build Life Experiences*
12. Innovation and Entrepreneurship – *How Startups Are Shaping the Future*

Course materials were accessible through the class website: https://sites.google.com/view/la0374-el.

Procedures

The study was conducted over a 15-week semester, beginning with an introductory session in Week 1, a midterm

examination in Week 8, and a final examination in Week 15. The course followed a three-credit blended learning model combining asynchronous digital instruction with synchronous face-to-face sessions.

Each week's asynchronous component consisted of a one-hour LMS-based lecture containing structured listening activities (keyword identification, main idea extraction, and content analysis) targeting both bottom-up (phonemic recognition, lexical access) and top-down (schema activation, inference-making) skills. The dialog component of the AI-generated materials was delivered in this format one-directionally.

Face-to-face sessions (two hours weekly) wrapped up and extended online learning. Nearpod was used for interactive instructor–student activities, allowing performance monitoring and instant feedback, though it did not support peer-to-peer interaction, even for dialog-related content. Additional class time included pronunciation instruction on English–Korean contrasts and demonstrations of supplementary listening resources (e.g., TED Talks, Arirang News, YouTube).

Students completed two weekly homework tasks: (A) comprehension questions on the listening materials and (B) productive speaking practice, recording the first paragraph of the passage after shadow-reading. Speech-to-text tools (e.g., Google Docs voice typing, Papago) were used for transcription, followed by self-analysis of errors submitted via Google Survey.

This cycle—online lecture, homework, and in-class consolidation—was followed for all twelve core lessons.

Data Collection and Analyses

Data were collected in the 14th week of the semester, immediately before the final examination period. A comprehensive written survey, administered during class, gathered both quantitative data (Likert-scale ratings) and qualitative data (open-ended responses) on student perceptions of the AI-integrated EFL listening course. The primary focus was on evaluating the instructional value and quality of AI-generated listening materials and the effectiveness of the three-format input sequence—passage, dialog, and summary—in supporting comprehension, vocabulary acquisition, and overall listening skill development. The survey questionnaire was provided in both English and Korean to ensure clarity and accessibility for all participants.

The survey began with demographic questions on academic year, gender, and self-rated English listening proficiency. The main section (Items 1–12) addressed content engagement and appropriateness, audio clarity at both normal and reduced speeds, and overall pronunciation and delivery. Several items also examined the impact of repeated content exposure across formats and identified unnatural or artificial voice qualities. The survey concluded with open-ended questions inviting qualitative feedback on learning support and suggestions for material improvement.

Quantitative data were analyzed using IBM SPSS Statistics. Internal consistency for the Likert-scale items was assessed with Cronbach's alpha: Items 1–6 (opportunities and benefits) showed good reliability (α = .863); Items 7–9 (perceived challenges) showed moderate reliability (α = .585); and Items 10–12 (perceived effectiveness) showed good reliability (α = .884). Descriptive statistics (means, standard deviations, and frequency distributions) summarized student perceptions, while one-way ANOVAs examined differences by self-rated English listening proficiency. Significant results were followed by Scheffé post hoc tests, and effect sizes (η^2) were reported to indicate practical significance.

Qualitative data from the open-ended responses were analyzed using thematic analysis through an inductive, iterative process. Responses were first read in full to establish familiarity, then segmented into meaningful units and coded to capture key ideas. Related codes were grouped into provisional categories, which were refined via constant comparison across the dataset. To improve reliability, coding was revisited multiple times and adjusted for consistency in interpretation. Final themes were consolidated into broader benefit- and challenge-related domains, and representative excerpts were selected for their clarity and relevance in illustrating patterns that contextualized the quantitative findings.

RESULTS AND DISCUSSION

The results are organized around the three research questions. Each section includes: (1) general perceptions, (2) differences by proficiency level, (3) qualitative insights, and (4) a summary and interpretation, integrating both quantitative and qualitative findings.

Opportunities of AI-Generated Materials in EFL Listening Courses

General Perceptions

In terms of opportunities, the survey data showed that students evaluated AI-generated materials positively. Internal

consistency reliability was good for the six relevant items (1–6), with Cronbach's α = .863, indicating strong agreement among measures of student perceptions. Table 2 presents descriptive statistics for these perceived opportunities.

TABLE 2
Descriptive Statistics for Perceived Opportunities of AI-Generated Materials (n = 114)

Item[1]	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	*Mean*	*SD*
1	2 (1.8)	8 (7.0)	20 (17.5)	52 (45.6)	32 (28.1)	3.90	0.87
2	5 (4.4)	14 (12.3)	33 (28.9)	42 (36.8)	20 (17.5)	3.58	1.01
3	16 (14.0)	26 (22.8)	27 (23.7)	28 (24.6)	17 (14.9)	2.84	1.10
4	7 (6.1)	14 (12.3)	33 (28.9)	42 (36.8)	18 (15.8)	3.50	1.01
5	5 (4.4)	13 (11.4)	29 (25.4)	43 (37.7)	24 (21.1)	3.91	0.89
6	4 (3.5)	10 (8.8)	28 (24.6)	43 (37.7)	29 (25.4)	3.96	0.85

Note. 1. Very unlikely, 2. Somewhat unlikely, 3. Neutral, 4. Somewhat likely, 5. Very likely

Overall, students responded positively to the AI-generated course materials, particularly noting how the structured sequence of passage, dialog, and summary formats enhanced their learning experience. The highest-rated item, Item 6 ("The combination of the passage, dialog, and summary helped reinforce key vocabulary and expressions"), received a mean score of 3.96 (*SD* = 0.846), suggesting that multi-format repetition was especially effective for vocabulary development and retention. Similarly, Item 5 ("The combination helped improve comprehension") achieved a high mean of 3.91 (*SD* = 0.888), reinforcing the value of presenting identical content through varied formats to support comprehension.

Differential Perceptions by Proficiency

When considering opportunities by proficiency level, analysis across five self-reported groups revealed that moderate- to good-level learners consistently rated the materials more positively than their lower-proficiency peers. Table 3 presents the descriptive statistics for each group, highlighting these differences.

TABLE 3
Statistical Analyses of Perceived Opportunities (n = 114)

Item	*F*(4, 109)	*P*	η^2	Significant Scheffé Comparisons
1	1.83	.128	.063	None
2	3.10	.018	.102	Good > Very limited
3	4.63	.002	.145	Good > Very limited; Excellent > Very limited
4	2.48	.049	.083	Moderate > Very limited
5	3.20	.016	.105	Good > Very limited
6	4.89	.001	.152	Good > Very limited; Excellent > Very limited

Note. η^2 *= effect size*

One-way ANOVAs revealed significant differences for Item 2, Item 3, Item 4, and Item 6, but not for Item 1 or Item 5. Effect sizes (η^2) ranged from small to large, and Scheffé post hoc tests confirmed significant contrasts for the same four items.

For content appropriateness (Item 2), Good-level learners rated the materials much higher (*M* = 4.33) than Very Limited learners (*M* = 2.91), *F*(4, 109) = 3.10, *p* = .018, η^2 = .102. For audio clarity (Item 3), Good and Excellent groups rated the materials as clearer than Very Limited learners (*M* = 2.49), *F*(4, 109) = 4.63, *p* = .002, η^2 = .145. For accent and style understanding (Item 4), Moderate learners scored higher (*M* = 3.80) than Very Limited learners (*M* = 2.91), *F*(4, 109) = 2.48, *p* = .049, η^2 = .083. For vocabulary reinforcement (Item 6), Good and Excellent learners rated the benefit higher than Very Limited learners, *F*(4, 109) = 4.89, *p* = .001, η^2 = .152.

[1] Refer to the Appendix for the full list of items.

Overall, higher-proficiency learners appeared to benefit more from linguistic content, structure, and audio delivery, while lower-proficiency learners may need additional guidance such as pre-teaching vocabulary, slower playback speeds, and additional visual support.

Qualitative Insights

Regarding opportunities identified through qualitative feedback, students' open-ended responses (Table 4) emphasized four main strengths of the AI-generated materials: dialogs, summaries, audio options, and readings/vocabulary.

Dialogs were appreciated for modeling authentic conversational patterns and offering exposure to different male and female voices with varied accents, which students felt improved their adaptability to diverse speaking styles. Summaries were viewed as effective tools for consolidating key information and reinforcing retention, especially when following the more detailed passages and dialogs.

Audio options—particularly the ability to slow playback—helped lower-proficiency learners build confidence and process information more comfortably. Repetition and segmentation features further reduced cognitive load, allowing learners to engage with content at their own pace.

Readings and targeted vocabulary support also emerged as important contributors to comprehension and skill development. While students across proficiency levels valued these features, those with moderate to good proficiency were more likely to highlight the benefits of shadowing, accent variation, and the structured sequence of passage–dialog–summary progression. This pattern suggests that stronger linguistic foundations may enable fuller engagement with the high-input-density format of the course.

TABLE 4

Most Helpful Aspects of the Materials (n = 42)

Dialog and Conversation Passages (Total 11)	
	Helped understand sentences and learn real-life expressions (5); supported overall comprehension (2); varied accents/voices improved adaptability (2); sentence ordering and shadowing reinforced expressions (2)
Summary (Total 11)	
	Organized content and identified key ideas (5); improved comprehension and structuring (3); listening to summary again aided retention (3)
Audio Materials and Speed Options (Total 11)	
	0.8× speed eased adjustment to faster speech (6); repetition and dictation built skills (3); paragraph-level audio segmentation improved comprehension (2); fill-in-the-blank tasks helped review (1)
Readings and Vocabulary (Total 9)	
	Weekly vocabulary PDFs, idiom explanations, and translations aided comprehension (3); interesting/diverse topics boosted motivation (2); multiple formats reduced cognitive load (2); few difficult words supported understanding (1); varied vocabulary improved reading skills (1)

Summary and Interpretation

Overall, opportunities centered on the structured integration of passages, dialogs, and summaries, which students rated most positively, particularly for vocabulary reinforcement (Item 6) and comprehension support (Item 5). Content interest (Item 1) was also rated highly, while audio clarity (Item 3) was the weakest area due to synthetic speech issues. Proficiency-based patterns indicated that higher-level learners derived greater benefits, while lower-level learners faced more difficulties, underscoring the need for differentiated instructional strategies when leveraging these opportunities in AI-generated listening materials.

Challenges of AI-Generated Materials in EFL Listening Courses

General Perceptions

In terms of challenges, analysis of the three survey items (Items 7–9) revealed modest reliability (Cronbach's $\alpha = .585$). This modest value is likely due to the small number of items and their coverage of distinct dimensions—unnatural language,

audio clarity, and limited voice or intonation variety. Unlike Items 1–6, which measured the positive opportunities of AI-generated materials, Items 7–9 were intentionally phrased to capture negative aspects. Accordingly, higher mean scores reflect stronger agreement that challenges were experienced, whereas lower scores indicate weaker agreement. Because each item addresses a distinct issue, they are interpreted separately. Table 5 presents descriptive statistics for these perceived challenges.

TABLE 5
Descriptive Statistics for Perceived Challenges of AI-Generated Materials (n = 114)

Item	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	*Mean*	*SD*
7	14 (12.3)	21 (18.4)	30 (26.3)	38 (33.3)	11 (9.6)	2.81	1.17
8	7 (6.1)	19 (16.7)	35 (30.7)	38 (33.3)	15 (13.2)	3.23	1.03
9	27 (23.7)	31 (27.2)	28 (24.6)	20 (17.5)	8 (7.0)	2.18	1.03

Note. 1. Very unlikely, 2. Somewhat unlikely, 3. Neutral, 4. Somewhat likely, 5. Very likely

The most prominent difficulty emerged from Item 8 ("The audio was hard to understand in terms of speed, pronunciation, and clarity"), which had the highest mean score (M = 3.23, SD = 1.03). This indicates that many students found clarity issues—including delivery speed and pronunciation—to be a significant obstacle. Item 7 ("Some content felt unnatural or artificial") also showed notable concern (M = 2.81, SD = 1.17), suggesting that awkward expressions and tone were perceived as problematic by a substantial group of learners. By comparison, Item 9 ("The audio lacked variety in voice or intonation, which reduced my engagement") received the lowest score (M = 2.18, SD = 1.03). While monotony was still recognized, it was not as widely endorsed as a major challenge compared to clarity or naturalness.

Overall, students' main reported difficulties were related to audio clarity and naturalness, whereas limited intonation variety appeared less disruptive to their learning experience.

Differential Perceptions by Proficiency

TABLE 6
Statistical Analyses of Perceived Challenges (n = 114)

Item	*F(4, 109)*	*P*	η^2	Significant Scheffé Comparisons
7	3.23	.015	.106	None
8	1.44	.227	.050	None
9	4.47	.002	.141	Moderate > Excellent (p = .010)

Note. η^2 *= effect size*

Looking at individual challenge items across proficiency groups, Table 6 shows the one-way ANOVA and Scheffé post hoc results. For Item 7 (unnatural language), overall group differences were statistically significant (F(4, 109) = 3.23, p = .015, η^2 = .106). Although no post hoc comparisons reached significance, Excellent-level learners tended to give lower ratings, suggesting they perceived fewer issues with authenticity compared to other groups. For Item 8 (audio clarity), no significant differences were found, indicating that problems with speed, pronunciation, and prosody were felt similarly across levels. For Item 9 (voice and intonation variety), group differences were statistically significant (F(4, 109) = 4.47, p = .002, η^2 = .141). Scheffé post hoc tests revealed that Moderate-level learners rated this aspect significantly more positively than Excellent-level learners (p = .010). However, this result should be treated with caution due to the very small number of Excellent-level participants (n = 3), which limits the reliability and generalizability of the finding.

Qualitative Insights

Qualitative responses further emphasized challenges tied to AI-generated audio (Table 7). Students most often mentioned voice quality and pronunciation issues (38 instances), describing delivery as monotonous or unnatural, with occasional mispronunciations, awkward intonation, and neglect of punctuation cues. Playback speed—especially in summaries—was also reported as too fast, leading to comprehension difficulties. Other concerns included slurred words, artificial speed adjustments, and lack of emotional nuance, all of which reduced perceived naturalness.

TABLE 7

Confusing or Unhelpful Aspects of the Materials (n = 55)

AI Voice and Pronunciation Issues (Total 38)	
	Unnatural/robotic tone (10); audio too fast (8); 1.0× speed faster than normal (1); mispronunciations (1); ignored punctuation (3); slurred words (4); artificially increased speed (1); summary section hard to follow (2); awkward intonation (2); negative views on AI listening (5); preference for human recordings (1)
Difficulties with Accents (Total 8)	
	British accent unfamiliar (2); mixed accents confusing (2); native speed too fast (2); preference for British/Australian English (1); AI lacks natural intonation vs. native speakers (1)
Dialog-Related Issues (Total 5)	
	Purpose unclear (1); grammar points hard to find (1); female voice too fast (1); unrealistic conversation style (1); poor pronunciation/intonation in dialog (1)
Other Material Concerns (Total 6)	
	Multiple vocabulary lists inconvenient (1); repetitive passages (1); no vocabulary preview (1); change of voice mid-way (1); occasional grammar errors (1); low audio quality (1)

Accent-related difficulties (8 mentions) included trouble with British or mixed accents and frustration with the speed of native-speaker recordings. Some students expressed a preference for focusing on specific varieties, such as British or Australian English, while others noted that AI still fell short of native speakers in prosody. Dialog sections also drew criticism (5 mentions). Respondents sometimes found their purpose unclear, pointed out unrealistic conversational patterns, or mentioned that certain voices were too fast. Others noted poor pronunciation or intonation in these segments. Finally, material-related concerns (6 mentions) ranged from inconvenient multiple vocabulary lists and repetitive passages to the absence of vocabulary previews, occasional grammatical errors, and inconsistent voice quality between segments.

Despite these challenges, many students responded positively to the course as a whole. A total of 40 comments such as "No issues," "It was helpful," and "Everything was useful" indicated that, for a significant proportion of learners, the benefits of the structured, multi-format design outweighed any drawbacks

Summary and Interpretation

Taken together, the item-level results suggest that audio delivery and language naturalness were the most salient challenges of AI-generated listening materials. Clarity-related issues were reported across all proficiency groups. However, higher-level learners generally perceived fewer problems with authenticity and expressiveness, indicating that they were less negatively affected by synthetic features of the audio. In contrast, moderate-level learners rated monotonous delivery as more problematic, suggesting they may have been more sensitive to limited variation in voice and intonation. These trends offer useful insights, though differences involving the Excellent group should be interpreted cautiously, given the very small sample size, which restricts the strength of statistical inference.

Effectiveness of AI-Generated Materials in EFL Listening Skill Development

General Perceptions

With regard to effectiveness, survey responses to Items 10–12 demonstrated high internal consistency (Cronbach's $\alpha = .884$), indicating strong reliability in measuring perceived learning gains. These questions covered bottom-up skills (e.g., vocabulary recognition), top-down skills (e.g., extracting main ideas), and overall listening development. Table 8 summarizes descriptive results for perceived effectiveness.

TABLE 8

Descriptive Statistics for Perceived Effectiveness of AI-Generated Materials (n = 114)

Item	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	*Mean*	*SD*
10	2 (1.8)	7 (6.1)	33 (28.9)	52 (45.6)	20 (17.5)	3.85	0.82
11	2 (1.8)	8 (7.0)	34 (28.9)	50 (43.9)	20 (17.5)	3.80	0.85
12	2 (1.8)	7 (6.1)	32 (28.1)	52 (45.6)	21 (18.4)	3.87	0.83

Note. 1. Very unlikely, 2. Somewhat unlikely, 3. Neutral, 4. Somewhat likely, 5. Very likely

Table 8 presents descriptive statistics for perceived effectiveness of AI-generated listening materials. Ratings across all three items were consistently positive. Item 10, which focused on bottom-up listening skills, received a mean score of 3.85 (*SD* = 0.82), indicating that students valued the materials for strengthening their ability to recognize vocabulary and phonemes. Item 11, centered on top-down skills such as extracting main ideas, was also rated favorably (*M* = 3.80, *SD* = 0.85). Item 12, asking about overall listening improvement from the structured sequence, received the highest rating (*M* = 3.87, *SD* = 0.83), suggesting that students appreciated the repeated, scaffolded format.

Overall, these results point to a strong perception that the multi-modal approach—cycling content through passage, dialog, and summary—helped students develop both detailed and global listening skills. The structure appeared to address a range of processing needs and provided balanced support for different aspects of listening comprehension.

Differential Perceptions by Proficiency

Effectiveness ratings showed no statistically significant differences across proficiency groups (Table 9). This indicates that learners, regardless of level, perceived the AI-generated materials as generally helpful for listening development. While mean scores varied slightly across groups, these differences were not consistent or large enough to reach statistical significance. Thus, the results suggest that the structured, repetitive format of the materials was broadly effective across proficiency levels, though individual learners may have experienced varying degrees of benefit depending on their listening needs and challenges.

TABLE 9

Statistical Analyses of Perceived Effectiveness (n = 114)

Item	*F(4, 109)*	*P*	η^2	Significant Scheffé Comparisons
10	2.30	.064	.078	None
11	2.17	.077	.074	None
12	1.30	.274	.046	None

Note. η^2 = *effect size*. No pairwise differences were statistically significant at $p < .05$.

Qualitative Insights

While the quantitative results focused directly on students' ratings of effectiveness, the qualitative findings were drawn from their open-ended suggestions for course improvement (Table 10). These comments nonetheless offered valuable insight into how effectiveness might be enhanced. Many students reaffirmed that repetition of the same content across multiple formats supported retention and comprehension, while also recommending expanded use of activities such as shadowing, sentence ordering, and dictation to strengthen listening and pronunciation. Others highlighted the need for greater flexibility in playback—slowing audio and replaying sections—to give learners more control over pacing and to support gradual adaptation to more natural speech.

TABLE 10

Suggestions for Course Improvement (n = 60)

Improving Audio Quality and AI Voice (Total 18)
Unnatural pronunciation or ignoring punctuation (6); audio too fast/speed adjustment needed (6); replace AI with human recordings (1); lack of natural intonation/emotion (1); fix AI errors (1); general/uncategorized (3)
Improving Lesson Structure (Total 16)
Reduce repetition, add variety (3); shorten or revise dialog/summary sections (2); focus on key-sentence explanations over full translations (2); improve content flow (1); add translations to readings (1); general/uncategorized (7)
Increasing Speaking and Interaction (Total 14)
Add group work and peer feedback (7); more speaking activities (6); use easy English videos/songs to build confidence (1)
Other Suggestions (Total 6)
Reduce workload (1); confidence-building for beginners (1); level-based content (1); simpler videos than TED Talks (1); more visual materials (1); LMS self-study guide (1)

However, several areas for improvement emerged. The most common involved AI-generated audio, with students noting unnatural pronunciation, awkward pauses, and a lack of natural intonation. Excessive speed in some recordings was also reported, prompting calls for more human-recorded audio to improve authenticity.

Lesson structure was another focal point for feedback. Some students found the sequence repetitive or overly rigid, recommending more variety, shorter dialog/summary sections, and key-sentence explanations instead of full translations.

Requests for increased interaction were also significant. While the course included real-time instructor–student exchanges during Nearpod-based face-to-face sessions, there were no structured peer-to-peer activities. Students suggested adding group work, peer feedback, and more speaking opportunities to create a collaborative learning environment.

Other suggestions included providing level-based content, simplifying materials for lower-level learners, adding more visual aids, and offering clear self-study guidance via the LMS.

Summary and Interpretation

Both quantitative and qualitative evidence highlight effectiveness as one of the clearest benefits of AI-generated materials. Structured repetition across passages, dialogs, and summaries was consistently perceived as supporting both bottom-up and top-down listening processes, and many students regarded it as the course's strongest feature. Importantly, no statistically significant differences were found across proficiency levels, suggesting that learners at all levels were able to benefit from the structured format. At the same time, students expressed a clear desire for improvements in audio naturalness and pacing to further enhance learning experiences.

CONCLUSION

This study examined university EFL learners' experiences with AI-generated listening materials, focusing on their instructional value, challenges, and perceived impact on listening development.

The findings revealed several key opportunities that AI-generated materials offer for EFL listening instruction. The structured three-part format—consisting of passage, dialog, and summary—provided systematic vocabulary reinforcement and supported both bottom-up and top-down listening processes. Additionally, features such as adjustable playback speed, unlimited repetition, and multimodal presentation formats enabled learners to engage in self-paced, independent study, making the materials particularly well-suited for individualized learning outside the classroom.

However, learners also identified multiple challenges that hindered their learning experience. Audio quality issues were most prominent, including robotic intonation, unnatural phrasing, and occasional pronunciation errors, which were especially problematic for lower-proficiency students. Beyond technical limitations, students expressed concerns about the lack of structural variety in lesson design and insufficient opportunities for peer interaction during class, both of which they felt could have enhanced engagement and communicative practice.

Despite these challenges, learners generally perceived AI-generated materials as effective in improving their listening comprehension, particularly through repeated exposure to vocabulary and structured content progression. However, their overall assessment was tempered by the limitations noted above. Students recognized the value of the materials but emphasized that their effectiveness was contingent not only on script and audio quality but also on how the materials were integrated into broader classroom pedagogy and supplemented with meaningful interactional opportunities.

Pedagogical Implications

The findings suggest that AI-generated listening materials can be a valuable addition to an EFL curriculum, but their impact depends greatly on how they are used in the classroom. While improving the quality of the AI output—through better scriptwriting, pronunciation checks, and post-production editing—is essential, this is only part of the equation. Learners benefit most when these materials are embedded in lessons that are varied, interactive, and responsive to their needs.

Instructors may wish to diversify lesson structures rather than relying too heavily on a single sequence of passage, dialog, and summary. Alternating between different activity formats can prevent monotony and maintain engagement. AI-generated audio should also be balanced with occasional human-recorded segments to provide more natural intonation and emotional nuance, giving learners a model of authentic speech patterns.

Another important consideration involves the representation of accent varieties in AI-generated materials. While this study utilized different voice options (American and British, male and female) primarily to address technical issues of

naturalness and clarity, accent variety carries broader pedagogical significance within a World Englishes framework. Exposing learners to diverse English accents can help cultivate awareness and acceptance of linguistic variation as a natural feature of global communication. Instructors should therefore view accent diversity not merely as a technical feature but as an opportunity to foster pluralistic attitudes toward English and prepare students for multilingual, multicultural contexts.

Support for lower-proficiency learners is another key consideration. Providing vocabulary previews, introducing slower playback speeds at first, and gradually moving toward faster, more authentic delivery can help these students keep pace without feeling overwhelmed.

Equally important is fostering peer interaction. Although this course included interactive exchanges with the instructor, students had limited opportunities to communicate directly with each other during class. Incorporating structured group listening tasks, role plays, or peer feedback sessions could extend listening practice into communicative contexts and make learning more collaborative.

Finally, integrating listening with speaking practice—such as pairing text-to-speech listening input with speech-to-text speaking tasks—could help students strengthen both skills in parallel. In this way, AI-generated listening content becomes not just a source of input, but a springboard for meaningful, output-oriented classroom activities.

Limitations and Further Study

This study was conducted within a single university context, which may limit the generalizability of its findings. The dialog component of the AI-generated materials was delivered only in the online class in a one-way format without interactive engagement, and because the effectiveness of listening materials is closely linked to instructional delivery, this may have influenced students' perceptions of that section. The reliance on self-reported perceptions, rather than objective measures of listening performance, also limits the ability to directly infer actual skill gains. In addition, the AI-generated audio reflected the specific TTS technology available at the time, and learner perceptions may change as these technologies advance.

Another limitation concerns the distribution of proficiency levels among participants. Although the sample size overall was adequate, the number of students in the Excellent group was very small (n = 3). As a result, statistical comparisons involving this group should be interpreted with caution, as subgroup imbalances reduce the stability and generalizability of findings. Furthermore, the participant pool was predominantly composed of engineering students, which limits the generalizability of the results to students from other academic disciplines. Given the potential differences in learning styles, linguistic backgrounds, and academic writing conventions across fields, future research should include more diverse samples representing various majors to enhance the external validity of the findings.

Future research should examine the impact of AI-generated listening materials on measurable listening outcomes over longer periods and across varied learning contexts. Studies could explore hybrid approaches combining AI-generated and human-recorded input, as well as adaptive AI systems that adjust complexity and delivery speed to individual learner profiles. Incorporating more interactive classroom uses of dialogs, rather than relying solely on one-way delivery, may provide a fuller understanding of their instructional value. Additionally, future studies should investigate how demographic variables such as gender, academic major, and prior technology experience influence learners' perceptions and use of AI-generated materials, as recent research (e.g., Alharbi, 2025) suggests that such factors may significantly shape learners' trust in and perceived effectiveness of AI tools.

References

Alharbi, A. (2025). Perceptions of ChatGPT's reliability and efficacy for academic tasks among Saudi university students. *Modern English Education*, *26*, 242-262. https://doi.org/10.18095/meeso.2025.26.1.242

Aziza, E. (2025). Learner perceptions and attitudes towards AI in English language learning. In I. Petrov, N. Koval, & L. Ivanova (Eds.), *Prospects of teaching English for professional purposes in non-philological higher education institutions: Problems and solutions* (pp. 703-709). Kyiv, Ukraine: National Technical University of Ukraine. https://doi.org/10.5281/zenodo.14981655

Baxramova, S. (2025). Exploring AI-generated content in EFL classrooms: Student engagement and comprehension outcomes. *Language Learning & Technology*, *29*(1), 45-62.

Chan, H., & Hu, B. (2023). Learner acceptance of AI-based English listening apps in university EFL settings. *System*, *113*, 102997.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-340.

Derakhshan, A. (2025). *EFL students' perceptions about the role of generative AI-mediated instruction in their learning* [Manuscript submitted for publication]. Department of English Language and Literature, Golestan University, Gorgan, Iran.

Dorgham, R. (2024). Using artificial intelligence tools for developing EFL students' listening skills and reducing their listening anxiety. *Faculty of Education Journal, Benha University*, *40*(140), 513-533.

Jung, H. (2025). AI-assisted student-generated listening materials in high school EFL: A sociomaterial perspective. *Multimedia-Assisted Language Learning*, *28*(2), 55-76. https://doi.org/10.15702/mall.2025.28.2.55

Luque de La Rosa, A., Fernández-Serrano, M. J., & García-Gómez, M. A. (2025). Reducing cognitive load in AI-assisted listening tasks: A study of Spanish EFL learners. *CALICO Journal*, *42*(1), 91-109.

Mak, B. (2021). Text-to-speech technology in ESL classrooms: Applications and limitations for listening instruction. *Computer Assisted Language Learning*, *34*(5-6), 523-540.

Meniado, J. C. (2023). Pedagogical affordances and linguistic limitations of AI-generated texts in ESL teaching. *TESOL Journal*, *14*(2), e00351.

Ngo, H. T., Tran, L. D., & Le, M. H. (2024). The impact of automatic speech recognition on pronunciation instruction: A meta-analytic review. *ReCALL*, *36*(1), 29-47.

Peiró-Lilja, À., Cámbara, G., Farrús, M., & Luque, J. (2022). Naturalness and Intelligibility Monitoring for Text-to-Speech Evaluation. In *Speech Prosody 2022* (pp. 445-449). International Speech Communication Association (ISCA). https://doi.org/10.21437/SpeechProsody.2022-91

Southwell, R., Pugh, S., Perkoff, E. M., Clevenger, C., Bush, J., Lieber, R., Ward, W., Foltz, P., & D'Mello, S. (2022). Challenges and feasibility of automatic speech recognition for modeling student collaborative discourse in classrooms. In *Proceedings of the 15th International Conference on Educational Data Mining (EDM 2022)* (pp. 282-293). International Educational Data Mining Society. https://doi.org/10.5281/zenodo.68531091

Yao, Y., Liang, T., Feng, R., Shi, K., Yu, J., Wang, W., & Li, J. (2024). SR-TTS: a rhyme-based end-to-end speech synthesis system. *Frontiers in Neurorobotics*, *18*, 1322312. https://doi.org/10.3389/fnbot.2024.1322312

Young, J., & Shishido, T. (2023). Enhancing EFL listening instruction with AI-generated texts: A classroom-based investigation. *Language Learning & Technology*, *27*(2), 23-40.

Appendix

Survey Questionnaire (Item 1-12)

Opportunities:

1. The content of the listening passages, dialog, and summary was interesting and engaging.
2. The language used in the listening passages, dialog, and summary was appropriate for my level.
3. The audio was easy to understand in terms of speed, pronunciation, and clarity.
4. The audio of the passage helped me understand various English accents or speaking styles.
5. The combination of the passage, dialog, and summary, which shared the same content, helped improve my comprehension.
6. The combination of the passage, dialog, and summary, which shared the same content, helped reinforce key vocabulary and expressions.

Challenges:

7. Some content (e.g., vocabulary, tone, or expressions) felt unnatural or artificial.
8. The audio was hard to understand in terms of speed, pronunciation, and clarity.
9. The audio lacked variety in voice or intonation, which reduced my engagement. Effectiveness:
10. The course materials helped me improve bottom-up listening skills, such as recognizing vocabulary and sounds.
11. The course materials helped me improve top-down listening skills, such as understanding main ideas and context.
12. The structured sequence (passage, dialog, summary) supported my overall listening development.

Open-Ended Questions:

- What aspects of the course materials helped your learning the most?
- Were there any parts of the course materials that you found confusing or unhelpful?
- How do you think this course could be improved?