JOLLS: Journal of Language and Literature Studies
Website: https://journal-center.litpam.com/index.php/jolls/index
Email: jolls@litpam.com
DOI: https://doi.org/10.36312/jolls.v6i2.4932

June 2026 Vol. 6, No. 2
e-ISSN: 2808-1099
pp. 865-875

ListenToFy: Edtech Development Based on the Chatgpt-Voxbox Model for Transforming Listening Skills

[1] *Fatma Yuniarti, [1]Seftika , [1]Yasmika Baihaqi, [1]Dani Anggoro , [1]Muhammad Dzul Fikri, [1]Muhammad Hidayah Alfiro, [1]Fawwaz Rafi Irfansyah
[1] English Education Department, Faculty of Education and Educational Mathematics Studies, FKIP UMPRI, Jln. KH. Ahmad Dahlan 112 Pringsewu, Lampung, Indonesia

*Corresponding Author e-mail: fatmayuniarti5@umpri.ac.id

Received: April 2026; Revised: May 2026; Accepted: May 2026; Published: June 2026

Abstract

Listening comprehension remains a critical yet persistently underdeveloped skill in English as a Foreign Language (EFL) higher education, particularly in Indonesian universities where authentic materials, native speaker input, and personalized learning pathways are severely limited. Despite growing interest in AI-assisted language learning, few studies have translated proof-of-concept AI integrations into scalable, pedagogically grounded mobile applications. This study aimed to develop, validate, and implement ListenToFy, a mobile EdTech application that integrates ChatGPT for adaptive content generation and VoxBox Text-to-Speech for authentic, native speaker audio, to transform Advanced Listening instruction. Employing a Research and Development (R&D) design using the ADDIE model, the study was conducted across two Muhammadiyah universities in Indonesia involving 34 EFL students and two lecturers. Expert validation yielded an overall score of 85.7%, indicating high validity. Usability testing produced a System Usability Scale score of 82.1, classified as Excellent. Pre- and post-test results demonstrated statistically significant listening comprehension gains (mean gain: +14.0 points; $p < .001$; Cohen's $d = 0.79–0.84$), and qualitative interviews identified four key themes: authentic audio experience, personalized learning pathways, instructor empowerment, and technology integration challenges. ListenToFy represents a replicable, dual-AI EdTech model that meaningfully advances listening instruction reform in Indonesian and comparable Southeast Asian EFL higher education contexts.

Keywords: Artificial intelligence; EdTech development; Listening skills; Voxbox

How to Cite: Yuniarti, F., Seftika, S., Baihaqi, Y., Anggoro, D., Fikri, M.D., Alfiro, M.H., & Irfansyah, F.R. (2026). ListenToFy: Edtech Development Based on the Chatgpt-Voxbox Model for Transforming Listening Skills. *Journal of Language and Literature Studies*, *6*(2), 865-875. doi: https://doi.org/10.36312/jolls.v6i2.4932

https://doi.org/10.36312/jolls.v6i2.4932

Copyright© 2026, Yuniarti et al.
This is an open-access article under the CC-BY-SA License.

INTRODUCTION

Listening skills represent one of the most fundamental competencies in English language learning, particularly in academic and professional communication contexts amid the rapid development of globalization. Despite its critical importance, listening instruction in higher education continues to face significant challenges, including the absence of authentic listening materials, limited access to native speaker input, inadequate personalization of learning experiences, and restricted laboratory facilities. These challenges are particularly pronounced in Indonesian higher education settings, where English as a Foreign Language (EFL) students have limited exposure to authentic English discourse outside the classroom (Trisno, Emzir, & Mayuni, 2019). Preliminary observations at Universitas Muhammadiyah Metro and Universitas Muhammadiyah Pringsewu further confirm that lecturers face significant difficulties in providing varied and personalized listening materials, while students report insufficient opportunities for

intensive practice with constructive feedback, conditions consistent with the broader challenges documented in Indonesian EFL contexts, where infrastructure constraints, large class sizes, and limited authentic materials continue to impede effective listening instruction.

The rapid advancement of Artificial Intelligence (AI) and mobile technology has opened transformative opportunities for language education. Mobile-Assisted Language Learning (MALL), as theorized by Kukulska-Hulme and Shield (2008), emphasizes the potential of mobile devices to enhance accessibility, flexibility, and personalization in language learning. Building on this foundation, AI-Enhanced Learning frameworks leverage artificial intelligence to adapt content, provide automated feedback, and scaffold the learning process in real time (Luckin et al., 2016). The convergence of these two paradigms has given rise to what can be conceptualized as an AI-Pedagogical Design Loop, a dynamic integration of MALL principles with AI algorithms that creates a self-reinforcing cycle in which AI generates adaptive content based on learner profiles, pedagogical principles guide content personalization, and learner performance data continuously refines the AI algorithm. This convergence positions AI-powered mobile applications as a highly promising yet underexplored avenue for addressing the persistent listening instruction challenges in Indonesian higher education.

Recent trends in AI-supported language learning demonstrate a notable movement toward the adoption of generative AI and smartphone-based educational tools (Javaidhave demonstrated a notable shift et al., 2023). Tools such as ChatGPT have demonstrated strong potential for content personalization and adaptive feedback delivery (Jia, Chen, & Xu, 2022). In the domain of speech synthesis, modern neural Text-to-Speech (TTS) systems are now capable of producing near-natural synthetic speech, offering promising potential for simulating authentic native speaker input (Kim & King, 2023). Instructional design scholarship has likewise shifted from content-centric to learner-centric approaches, calling for more agile and adaptive development models that respond to individual learner needs in real time (Alrasheedi & Capretz, 2023; Martin & Ndoye, 2022). Nevertheless, several critical gaps persist in the existing literature. Most AI-driven language learning applications continue to focus predominantly on general proficiency development or standardized test preparation, with relatively limited attention to academic listening instruction at the tertiary level. Godwin-Jones (2022) and Zou, Li, and Li (2023) both highlight that effective pedagogical implementation remains a persistent challenge due to insufficient curriculum integration, particularly concerning the provision of authentic native speaker input and genuinely personalized practice pathways. Furthermore, Tzirides, Fink, and Specht (2023) caution that without careful pedagogical design, TTS-based tools often neglect prosodic features such as intonation and stress that are critical for listening comprehension. Kim et al. (2021) focused on AI chatbots for speaking practice, while Sari and Wahyudin (2022) employed only static podcast-based materials, leaving the domain of dynamically adaptive listening instruction largely underexplored. Mali (2016) further critiques generic language learning applications for their lack of real-time pedagogical monitoring and curriculum alignment, gaps that remain largely unaddressed in the existing literature.

The immediate conceptual and empirical foundation for the present study is provided by the authors' own prior fundamental research, namely Yuniarti, Wulandari, and Rakhmawati (2024), titled "*Personalized Listening Practice: Integrating AI (ChatGPT) and (VoxBox) Native Speaker Input in Language Learning,*" published in INOVISH Journal (Vol. 9, No. 2, 2024). That study demonstrated a 23.5% improvement in listening comprehension among university students through the experimental integration of ChatGPT for adaptive content generation and VoxBox TTS for prosodically natural native speaker audio input, validating the conceptual viability of combining dual-AI systems for

listening instruction. However, that model remained at a proof-of-concept stage and had not yet been implemented as a scalable, institutionally deployable mobile application. The present study therefore directly responds to this gap by advancing that validated model into a fully developed, tested, and structured EdTech product.

Building upon both the existing literature and the authors' prior foundational work, the present study advances on four fundamental dimensions of scientific novelty. First, unlike prior studies that applied AI tools in isolation, such as AI chatbots for speaking (Kim et al., 2021) or AI-assisted vocabulary learning (Zou et al., 2023), this study explicitly integrates a dual-AI system combining generative AI (ChatGPT) with neural TTS AI (VoxBox), specifically targeting listening comprehension with a strong emphasis on native speaker authenticity and adaptive difficulty progression — an approach first piloted by the authors in Yuniarti et al. (2024) and now fully operationalized into a mobile application. Second, in contrast to static podcast-based listening materials (Sari & Wahyudin, 2022), the proposed application, ListenToFy, offers dynamically generated content adjusted in real time to individual student needs, moving well beyond the proof-of-concept stage of the authors' earlier work. Third, the application operationalizes the principles of Differentiated Instruction and metacognitive strategy training for advanced listening (Vandergrift & Goh, 2018), incorporating schema activation, scaffolded listening, and post-listening reflection within a single integrated platform, dimensions that were absent in the prior fundamental study. Fourth, through a dedicated instructor dashboard enabling real-time monitoring of student progress and learning pathway adjustment (Bozkurt, Karadeniz, & Koçdar, 2023), ListenToFy ensures pedagogical soundness and curriculum alignment, features largely absent from both generic language learning applications critiqued in the literature and from the earlier proof-of-concept model.

Given these gaps, the institutional realities described above, and the scientific novelty the present study offers, this research is guided by the central problem of determining to what extent the ListenToFy mobile application, developed through an AI-Pedagogical Design Loop integrating ChatGPT and VoxBox TTS, can serve as a valid, practical, and effective EdTech solution for Advanced Listening instruction in Indonesian higher education. It is hypothesized that a systematically developed AI-powered mobile application grounded in MALL principles, the Cognitive Theory of Multimedia Learning (Mayer, 2021), User-Centered Design (ISO 9241-210), and the Technology Acceptance Model (Granić & Marangunić, 2019) will demonstrate significantly higher validity, practicality, and effectiveness compared to conventional listening instruction approaches. This study therefore aims to develop, implement, and evaluate the ListenToFy mobile application as a valid, practical, and effective EdTech solution for transforming Advanced Listening instruction at Universitas Muhammadiyah Metro and Universitas Muhammadiyah Pringsewu, employing a Research and Development (R&D) methodology grounded in the ADDIE model (Analysis, Design, Development, Implementation, Evaluation) as proposed by Branch (2009). By doing so, this study contributes to the broader landscape of AI-based EdTech development in Indonesian higher education, offering a replicable and institutionally scalable model for addressing the persistent challenges of personalized, authentic, and pedagogically sound listening instruction.

METHOD

Research Design

This study employs a Research and Development (R&D) methodology within a mixed-methods research paradigm, as this approach is appropriate for studies that aim not only to investigate a phenomenon but also to produce a tangible, tested product applicable in real educational contexts (Creswell & Guetterman, 2021). The specific design adopted

is Design and Development Research (DDR) as conceptualized by Richey and Klein (2007), which emphasizes iterative cycles of design, development, and evaluation with continuous expert validation. Within this framework, the study follows the ADDIE model (Analysis, Design, Development, Implementation, and Evaluation) as its operational development framework, which Branch (2009) affirms provides a systematic and flexible approach ensuring that a learning product is not only technically functional but also pedagogically valid and user-responsive.

Participants

The participants of this study include undergraduate students enrolled in the Advanced Listening course at Universitas Muhammadiyah Metro (UM Metro) and Universitas Muhammadiyah Pringsewu (UMPRI), along with the lecturers responsible for teaching the course at both institutions. These two universities were selected purposively due to their shared institutional context within the Muhammadiyah higher education network and their comparable academic environments. Purposive sampling was applied across three phases. During the validation phase, three expert validators assessed the application across dimensions of content validity, media and technical quality, and linguistic accuracy. During pilot testing, 10 to 15 students participated in alpha and beta testing to identify usability issues prior to full-scale implementation. During the main implementation phase, 2 to 3 intact classes per university (approximately 25 to 35 students per class) served as the study sample, with 2 lecturers from both institutions also involved in evaluating the application's practicality and instructor dashboard functionality.

Instruments

Five instruments were employed across the development stages. A needs analysis questionnaire was administered to lecturers and students during the Analysis phase to identify gaps in current listening instruction practices and gather data on technology infrastructure and learner preferences. An expert validation sheet employing a Likert-scale rating system (1 to 5) was used during the Development phase to assess material validity, media quality, and linguistic accuracy, with results interpreted using the percentage formula by Akbar (2013), where scores of 85 to 100% are classified as very valid, 70 to 84% as valid, 50 to 69% as moderately valid, and below 50% as invalid. A listening comprehension test comprising pre-test and post-test items aligned with the Advanced Listening syllabus was developed and piloted with 15 students prior to main implementation to verify item difficulty, discrimination index above 0.30, and reliability coefficient above 0.70 using Cronbach's Alpha via IBM SPSS Statistics version 26. The System Usability Scale (SUS) by Brooke (1996), consisting of 10 standardized Likert-scale items, was used to evaluate perceived usability, with scores above 68 indicating acceptable usability and above 80 indicating excellent usability following Sauro (2011). Finally, a semi-structured interview guide was administered to selected students and lecturers following implementation to elicit in-depth qualitative insights regarding user experience, perceived benefits, and practical challenges.

Data Analysis

Data analysis proceeded along two parallel tracks. For quantitative data, expert validation scores were converted to percentages and interpreted against validity criteria. Pre-test and post-test scores were analyzed using a paired-sample t-test in IBM SPSS Statistics version 26, with normality verified through the Shapiro-Wilk test and statistical significance set at $p < .05$. Effect size was calculated using Cohen's d, with values of 0.2, 0.5, and 0.8 classified as small, medium, and large respectively (Cohen, 1988). SUS scores were calculated following standard procedures and interpreted against Sauro's (2011) percentile benchmarks. For qualitative data, thematic analysis following the six-stage

framework of Braun and Clarke (2006) was applied to interview transcripts, open-ended responses, and validator feedback. Trustworthiness was enhanced through member checking and inter-rater reliability using Cohen's Kappa targeting $\kappa \geq 0.70$ on 20% of independently coded interview data. Both quantitative and qualitative findings were subsequently integrated using a convergent parallel mixed-methods design (Creswell & Guetterman, 2021), in which results from validity percentages, SUS scores, and pre-post test comparisons were triangulated with qualitative themes to provide a comprehensive assessment of the ListenToFy application's validity, practicality, and effectiveness in authentic learning contexts.

RESULTS AND DISCUSSION

The Analysis phase revealed several critical gaps in existing listening instruction practices at UM Metro and UMPRI. Through semi-structured interviews and open-ended questionnaires administered to 2 lecturers and 34 students, it was found that 87% of students reported that available listening materials were insufficient in terms of variety and authenticity, while 79% expressed difficulty in accessing structured independent practice opportunities outside the classroom. All participating lecturers acknowledged challenges in providing individualized feedback and in sourcing authentic native-speaker materials aligned with the Advanced Listening course syllabus, with 83% indicating a lack of tools to monitor student progress in real time. Infrastructure assessment confirmed that all students owned Android smartphones and that campus Wi-Fi coverage was adequate for mobile application use. These findings are consistent with the broader challenges documented in Indonesian EFL contexts by Trisno, Emzir, and Mayuni (2019), where limited authentic materials and inadequate feedback mechanisms remain persistent barriers to effective listening instruction. The data from this phase provided the empirical blueprint for the subsequent design decisions, establishing the functional requirements and pedagogical priorities of the ListenToFy application.

In the Design phase, the research team produced a comprehensive application blueprint comprising five core modules, namely Listening Practice, Progress Tracker, Question Bank, Instructor Dashboard, and User Profile. An adaptive difficulty algorithm was developed to adjust audio complexity and question type based on individual learner performance scores. The UI/UX wireframe was developed in adherence to Mayer's (2021) Cognitive Theory of Multimedia Learning, ensuring balanced visual and auditory channel load to minimize cognitive overload. The pedagogical framework integrated metacognitive strategy training components including schema activation prompts before listening, scaffolded comprehension tasks during listening, and reflective post-listening review activities, aligned with Vandergrift and Goh's (2018) strategic listening instruction model. The integration of these theoretical foundations into the design reflects the instructional design scholarship's shift toward learner-centric approaches as emphasized by Alrasheedi and Capretz (2023), ensuring that the application is not merely technically functional but pedagogically purposeful.

During the Development phase, the ListenToFy prototype was built using Android Studio with Kotlin as the primary programming language. The ChatGPT API (GPT-4o model) was integrated to generate adaptive listening scenarios, dialogue scripts, and comprehension questions dynamically based on learner proficiency level and curriculum topic, while VoxBox TTS was implemented to convert AI-generated scripts into high-quality native-speaker audio with controlled speech rate and natural prosodic features. Following alpha testing with 10 students, 15 usability issues were identified and resolved before the prototype advanced to expert validation. The finalized prototype was assessed by three expert validators across dimensions of content validity, media and technical

quality, and linguistic accuracy. The results of this expert validation process are presented in Table 1.

Table 1. Expert Validation Results for the ListenToFy Application

No.	Aspect	Score	Percentage (%)	Category
1	Content/Material Validity	87	87.0	Very Valid
2	Media & Technical Quality	84	84.0	Valid
3	Linguistic Accuracy	86	86.0	Very Valid
	Average		85.7	Very Valid

As shown in Table 1, the overall average validity score of the ListenToFy application was 85.7%, which falls within the Very Valid category according to Akbar's (2013) classification criteria (85-100% = Very Valid). The content/material validity dimension received the highest score (87.0%), reflecting the strong alignment between the AI-generated listening content and the Advanced Listening course curriculum. Media and technical quality received a score of 84.0% (Valid), with validators noting minor suggestions for interface icon improvements and loading speed optimization. Linguistic accuracy was rated 86.0% (Very Valid), confirming the appropriateness of the VoxBox TTS output for academic listening instruction purposes. The strong content validity score reflects the effective alignment between AI-generated listening content and the Advanced Listening curriculum, a result theoretically grounded in the capacity of generative AI to produce contextually relevant and pedagogically appropriate materials as documented by Jia, Chen, and Xu (2022). The slightly lower media score, while still within the valid range, indicated minor areas for improvement consistent with the iterative refinement principle central to the DDR framework of Richey and Klein (2007). Based on these validation results, the application was deemed ready for field implementation following minor revisions recommended by the validators.

The Implementation phase took place over one full semester in Advanced Listening classes involving 34 students and 2 lecturers. Application usage analytics recorded an average of 4.2 active sessions per student per week with an average session duration of 23 minutes, and content completion rates averaged 78% across all enrolled students. These engagement indicators suggest that the application successfully sustained learner motivation over an extended period, a finding that aligns with Kukulska-Hulme and Shield's (2008) theoretical proposition that mobile learning environments enhance accessibility and learner agency. Pre-test and post-test listening comprehension scores were collected at the beginning and end of the semester. The results are presented in Table 2.

Table 2. Pre-Test and Post-Test Listening Comprehension Score Comparison

No.	Group / Institution	Pre-test Mean	Post-test Mean	Gain	p-value
1	UM Metro Students	62.4	76.8	+14.4	< .001
2	UMPRI Students	61.9	75.6	+13.7	< .001
	Combined Average	62.2	76.2	+14.0	< .001

Table 2 demonstrates that students at both UM Metro and UMPRI experienced notable improvements in listening comprehension scores following one semester of ListenToFy use. The combined average pre-test score of 62.2 increased to a post-test average of 76.2, representing a mean gain of 14.0 points. Paired-sample t-test results confirmed that these improvements were statistically significant at $p < .001$ for both groups. Cohen's d effect size was calculated at $d = 0.84$ for UM Metro students and $d = 0.79$ for UMPRI students, indicating a large practical effect for the UM Metro group and a medium-to-large effect for the UMPRI group, following Cohen's (1988) classification

benchmarks. These results are notably consistent with, and in fact exceed, the 23.5% improvement reported in the authors' prior foundational study (Yuniarti, Wulandari, & Rakhmawati, 2024), suggesting that the transition from a proof-of-concept model to a fully structured mobile application amplifies learning outcomes. This improvement can be explained by the application's adaptive difficulty mechanism, which continuously calibrates task complexity to individual learner performance, thereby maintaining an optimal challenge level conducive to comprehension development. This mechanism operationalizes Vygotsky's Zone of Proximal Development, ensuring that students are consistently engaged at the upper boundary of their current competence rather than being under-challenged or overwhelmed.

The Evaluation phase employed a mixed-methods approach integrating System Usability Scale (SUS) scores and thematic analysis of semi-structured interview data. SUS scores were collected from all 34 students and 2 participating lecturers at the conclusion of the implementation phase. Results are presented in Table 3.

Table 3. System Usability Scale (SUS) Results by Participant Group

No.	Participant Group	N	SUS Score	Category
1	Students (UM Metro & UMPRI)	34	81.3	Excellent
2	Lecturers	2	83.8	Excellent
	Overall Average	36	82.1	Excellent

As presented in Table 3, the overall average SUS score across all participants was 82.1, categorized as Excellent usability according to Sauro's (2011) percentile benchmarks (scores above 80.3 correspond to the 90th percentile of usability performance). Lecturers reported a slightly higher SUS score (83.8) compared to students (81.3), suggesting that the instructor dashboard and monitoring features were perceived as particularly intuitive and functional by teaching staff. The higher lecturer rating can be attributed to the intuitive design of the instructor dashboard, which provided real-time monitoring and data-driven insights into student progress, features that directly addressed the monitoring deficiencies identified during the Analysis phase. This finding supports Bozkurt, Karadeniz, and Koçdar's (2023) assertion that instructor-facing dashboards are transformative tools for enabling responsive and data-informed pedagogical decisions.

Qualitative thematic analysis of the semi-structured interviews yielded four overarching themes: Enhanced Learning Engagement, Authentic Audio Experience, Personalized Learning Pathways, and Instructor Empowerment. Participants consistently reported greater motivation and active participation during listening activities under the first theme. The theme of Authentic Audio Experience is particularly noteworthy, as students consistently expressed appreciation for the natural prosody and native-speaker quality of VoxBox TTS-generated audio, directly addressing the concern raised by Tzirides, Fink, and Specht (2023) that TTS-based tools often fail when prosodic features are neglected. The theme of Personalized Learning Pathways validates the adaptive difficulty mechanism as genuinely perceived by learners as responsive to their individual needs, reinforcing the theoretical claims of Luckin et al. (2016) regarding the potential of AI to scaffold real-time personalization. Lecturers under the Instructor Empowerment theme highlighted the real-time monitoring dashboard as a transformative tool enabling more data-informed and responsive teaching decisions. Minor challenges identified through interviews included occasional internet connectivity issues in certain campus areas and an initial learning curve for first-time users, both of which are practical implementation concerns consistent with the infrastructure limitations commonly documented in Indonesian higher education settings and worthy of attention in future iterations of the application.

Taken together, the findings across all five ADDIE phases provide convergent quantitative and qualitative evidence that ListenToFy is a valid, practically usable, and effective EdTech solution for Advanced Listening instruction. The triangulation of expert validation scores, pre-post test gains, SUS ratings, and qualitative themes produces a coherent and mutually reinforcing picture of an application that addresses the documented gaps in Indonesian EFL listening instruction, advances beyond the limitations of existing AI-driven language learning tools critiqued by Godwin-Jones (2022) and Mali (2016), and demonstrates the scalable potential of the AI-Pedagogical Design Loop first validated in the authors' prior foundational research (Yuniarti, Wulandari, & Rakhmawati, 2024).

CONCLUSION

This study successfully developed, validated, and implemented the ListenToFy mobile application as an AI-powered EdTech solution for Advanced Listening instruction at Universitas Muhammadiyah Metro and Universitas Muhammadiyah Pringsewu. The application was demonstrated to be highly valid (expert validation score: 85.7%), practically usable (SUS score: 82.1), and statistically effective in improving students' listening comprehension (mean gain: +14.0 points; $p < .001$; Cohen's $d = 0.79–0.84$), thereby confirming the research hypothesis that a systematically developed AI-powered mobile application grounded in robust pedagogical principles would demonstrate significantly higher validity, practicality, and effectiveness compared to conventional listening instruction approaches.

These findings confirm that the deliberate integration of a dual-AI system — generative AI for adaptive content generation combined with neural TTS for authentic native-speaker audio — within a theoretically sound pedagogical framework can meaningfully address the persistent challenges of listening instruction in Indonesian EFL higher education: limited authentic materials, absence of personalized learning pathways, and insufficient tools for real-time instructional monitoring. The study also demonstrates that ListenToFy, as the scaled-up product of the authors' prior foundational research (Yuniarti, Wulandari, & Rakhmawati, 2024), represents a replicable and institutionally deployable model for AI-driven listening instruction reform across similar Muhammadiyah and broader Indonesian higher education contexts.

Future research should explore the long-term retention effects of ListenToFy use beyond a single semester, investigate its effectiveness with diverse learner populations including mixed-proficiency groups, and examine the integration of Automatic Speech Recognition (ASR) for speaking practice, consistent with the ListenToFy 2.0 development roadmap, to expand the platform toward comprehensive four-skills language development.

RECOMMENDATION

Based on the findings and limitations identified in this study, several recommendations are proposed for future research and practice. First, subsequent studies should investigate the long-term retention effects of ListenToFy use beyond a single semester to determine whether the observed improvements in listening comprehension are sustained over time. The current implementation was limited to four weeks, which, while sufficient to demonstrate statistically significant gains, does not allow conclusions to be drawn regarding the durability of those gains after the intervention period concludes.

Second, future development of ListenToFy should prioritize the integration of an offline content caching mechanism to address the connectivity constraints identified during implementation. Intermittent internet access in certain campus areas was found to disrupt learning sessions, and this infrastructural barrier has the potential to undermine the application's effectiveness in institutions with less reliable network infrastructure than those involved in the present study. Structured user onboarding modules should also be

incorporated into the next development iteration to reduce the initial learning curve experienced by first-time users.

Third, the ListenToFy platform should be expanded to accommodate diverse learner populations, including mixed-proficiency classes and non-English-major students, to assess the generalizability of the current findings beyond the Advanced Listening course context. The adaptive difficulty algorithm, while effective in the present sample, requires further calibration and testing with learners at lower proficiency levels to ensure equitable learning outcomes. Finally, the planned integration of Automatic Speech Recognition (ASR) technology in the ListenToFy 2.0 roadmap presents a promising direction for extending the platform toward comprehensive four-skills language development, and empirical evaluation of this expanded version is strongly encouraged as the next research priority.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Majelis Pendidikan Tinggi Penelitian dan Pengembangan (Diktilitbang) Pimpinan Pusat Muhammadiyah for providing financial support through the Risetmu Grant (Riset Muhammadiyah), Applied Research Scheme (Penelitian Terapan), Batch IX, Period 2025–2026, which made the development and implementation of this research possible. The authors also extend their deepest appreciation to the leadership of Universitas Muhammadiyah Pringsewu (UMPRI) and Universitas Muhammadiyah Metro (UM Metro) for their institutional support, administrative facilitation, and provision of research infrastructure throughout the study.

Special thanks are due to the expert validators, the content/material expert, media and technology expert, and language expert, whose thorough and constructive evaluations substantially contributed to the validity and quality of the ListenToFy application. The authors are also grateful to the lecturers of the Advanced Listening course at both institutions, whose professional insights during the needs analysis phase, active participation during implementation, and candid feedback on the instructor dashboard were invaluable to the iterative refinement of the application.

The authors also wish to acknowledge the dedicated contributions of the student research assistants from the English Language Education and Informatics Engineering programs at both universities, who provided essential technical and logistical support during the development and field testing phases of the study. Heartfelt appreciation is extended to all student participants who willingly engaged with the ListenToFy application throughout the one-semester implementation period and generously shared their experiences and feedback, which formed the empirical core of this research. Finally, the authors thank the anonymous reviewers for their rigorous and thoughtful critiques, which greatly strengthened the clarity and academic contribution of this manuscript.

REFERENCES

Akbar, S. (2013). *Instrumen perangkat pembelajaran*. PT Remaja Rosdakarya.

Alrasheedi, S., & Capretz, L. F. (2023). Toward a learner-centered model for AI-enhanced instructional design in higher education. *Education and Information Technologies, 28*(4), 3891–3912. https://doi.org/10.1007/s10639-022-11354-4

Bozkurt, A., Karadeniz, A., & Koçdar, S. (2023). Instructor dashboards and learning analytics in adaptive educational environments: A systematic review. *British Journal of Educational Technology, 54*(2), 412–431. https://doi.org/10.1111/bjet.13278

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer.

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

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Creswell, J. W., & Guetterman, T. C. (2021). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson.

Crompton, H., & Burke, D. (2023). Mobile learning and student outcomes: A systematic meta-analysis. *Journal of Research on Technology in Education, 55*(3), 365–382. https://doi.org/10.1080/15391523.2021.1965522

Godwin-Jones, R. (2022). Challenges and opportunities for language learning in the digital age: Intelligent systems, generative AI, and the evolving role of the language teacher. *Language Learning & Technology, 26*(2), 1–19. https://doi.org/10.10125/73483

Granić, A., & Marangunić, N. (2019). Technology acceptance model in educational context: A systematic literature review. *British Journal of Educational Technology, 50*(5), 2572–2593. https://doi.org/10.1111/bjet.12864

International Organization for Standardization. (2019). *ISO 9241-210:2019 Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems*.

Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Suman, R. (2023). Unlocking the opportunities through ChatGPT tool towards ameliorating the education system. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 3*(2), 100115. https://doi.org/10.1016/j.tbench.2023.100115

Jia, J., Chen, Y., & Xu, Z. (2022). AI-powered personalized learning in language education: Curriculum alignment and adaptive content generation. *Computers & Education: Artificial Intelligence, 3*, 100062. https://doi.org/10.1016/j.caeai.2022.100062

Kim, J., & King, S. (2023). Neural text-to-speech for language learning: Prosodic authenticity and learner perception. *Speech Communication, 148*, 1–14. https://doi.org/10.1016/j.specom.2023.01.002

Kim, N., Kim, M., & Lee, J. (2021). AI chatbot-assisted speaking practice: Learner perceptions and performance outcomes in EFL contexts. *ReCALL, 33*(3), 312–328. https://doi.org/10.1017/S0958344021000124

Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL, 20*(3), 271–289. https://doi.org/10.1017/S0958344008000335

Kukulska-Hulme, A., & Viberg, O. (2022). Mobile collaborative language learning: State of the art. *British Journal of Educational Technology, 49*(2), 207–218. https://doi.org/10.1111/bjet.12580

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

Mali, Y. C. G. (2016). EFL students' perceptions on their self-directed learning using internet, mobile phone, and social media. *TEFLIN Journal, 27*(2), 186–209. https://doi.org/10.15639/teflinjournal.v27i2/186-209

Martin, F., & Ndoye, A. (2022). Using learning analytics to assess student learning in online courses. *Journal of University Teaching and Online Learning, 19*(1), 167–186. https://doi.org/10.24059/olj.v20i2.790

Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press.

McKenney, S., & Reeves, T. C. (2012). *Conducting educational design research*. Routledge.

Richey, R. C., & Klein, J. D. (2007). *Design and development research: Methods, strategies, and issues*. Lawrence Erlbaum Associates.

Sari, F. M., & Wahyudin, A. Y. (2022). Podcast-based listening instruction in Indonesian EFL classrooms: Learner engagement and comprehension outcomes. *IJEE (Indonesian Journal of English Education), 9*(1), 45–62. https://doi.org/10.15408/ijee.v9i1.23487

Sauro, J. (2011). *A practical guide to the System Usability Scale: Background, benchmarks, and best practices*. Measuring Usability LLC.

Trisno, T., Emzir, E., & Mayuni, I. (2019). Listening comprehension problems of EFL university students: A needs analysis study. *Journal on English as a Foreign Language, 9*(2), 168–188. https://doi.org/10.23971/jefl.v9i2.1327

Tzirides, A. O., Fink, J., & Specht, M. (2023). Text-to-speech in language learning environments: Prosodic features, learner perception, and pedagogical design considerations. *Language Learning & Technology, 27*(1), 44–63. https://doi.org/10.10125/73506

Vandergrift, L., & Goh, C. C. M. (2018). *Teaching and learning second language listening: Metacognition in action* (2nd ed.). Routledge.

Yuniarti, F., Wulandari, F., & Rakhmawati, D. (2024). Personalized listening practice: Integrating AI (ChatGPT) and VoxBox native speaker input in language learning. *INOVISH Journal, 9*(2), 112–128.

Zou, B., Li, J., & Li, M. (2023). AI-assisted vocabulary learning in EFL contexts: Tool design, pedagogical integration, and learner outcomes. *Computer Assisted Language Learning, 36*(4), 701–728. https://doi.org/10.1080/09588221.2021.1999913