

Multimodal Podcast Learning to Build L2 Listening Self-Efficacy in Gen Z EFL Learners

Suci Amaliah

suci.amaliah@unm.ac.id

Universitas Negeri Makassar

ABSTRACT

Gen Z learners are predominantly audio-visual, yet most educational podcasts remain audio-only, constraining their potential to support learners who depend on visual cues for comprehension. Addressing this limitation, the present mixed-methods study investigates how multimodal podcast learning combining audio, subtitles, and interactive transcripts enhances L2 listening self-efficacy among Gen Z EFL learners. Drawing on Bandura's social cognitive theory and Mayer's multimedia learning framework, sixty undergraduate students from Universitas Negeri Makassar participated in an eight-week intervention comparing multimodal and audio-only podcast instruction. Quantitative analyses indicated a significant improvement in listening self-efficacy for the multimodal group ($M = 4.02$, $SD = 0.68$) relative to the audio-only group ($M = 3.25$, $SD = 0.70$), $t(58) = 7.91$, $p < .001$, with a large effect size (Cohen's $d = 1.15$). Qualitative findings identified three recurrent themes: strengthened confidence through repeated mastery experiences, increased motivation through authentic topics, and greater learner autonomy fostered by multimodal scaffolds. Collectively, the results demonstrate that multimodal podcast learning not only reinforces listening competence but also cultivates the psychological confidence required for self-regulated learning. The study advances theoretical and pedagogical understanding of how multimodality transforms podcast-based instruction into an inclusive and empowering practice for contemporary EFL education.

Key words: Multimodal Podcast Learning, L2 Listening Self-Efficacy, GenZ Learners

INTRODUCTION

English language education in the 21st century is undergoing a significant transformation, particularly following the COVID-19 pandemic, which has reshaped the landscape of educational systems globally. As teaching and learning activities were compelled to transition into virtual spaces, both educators and students swiftly adapted to technology-driven digital learning models. In this context, a new generation of learners, called Generation Z, born between 1997 and 2012, has emerged; these individuals have been raised in a world entirely interconnected by digital technology (Hrdy et al., 2024). Gen Z are not merely growing alongside technology; they are inherently audiovisual and interactive learners. As Hrdy et al. (2024) observed, Gen Z students exhibit a strong preference for learning environments that incorporate visual, auditory, and digital interactive components. Furthermore, Princes et al. (2024) theorized in their study that Gen Z

students tend to favour online or blended learning modalities over traditional face-to-face instruction, as these approaches afford them greater control, flexibility, and unrestricted access to learning resources. This development represents a critical juncture that warrants careful consideration. Should conventional English language teaching strategies persist in formats misaligned with Gen Z's learning styles, the disparity between instructional methods and learners' needs will inevitably intensify.

As education progressively transitions toward online and hybrid modalities, there is an increasing demand for instructional media that are not only adaptable but also tailored to the learning preferences of Gen Z. This transition directly influences the design of English language instruction. In English as a Foreign Language (EFL) classroom, educators are tasked with developing learning experiences that are not only communicative but also pertinent to the information processing styles of Gen Z, utilizing interactive and multimodal media. A notable and rapidly emerging innovation in this domain is educational podcasting, an audio platform that delivers authentic, flexible, and engaging content, often presented in natural, conversational formats. Podcasts have become extensively embraced in English language teaching as a versatile alternative, enabling students to access authentic materials at any time and from any location. For example, Khoiriyah et al. (2024) observed that the use of podcasts enhances EFL students' listening skills. Additionally, Amaliah et al. (2025) illustrated that podcasts developed through a polysemy-awareness approach can broaden EFL students' collocational and speaking skills, while concurrently reinforcing the connection between meaning comprehension, speech context, and media-based learning rooted in authenticity. These insights substantiate the view that podcasts are not merely tools for listening practice but can serve as a comprehensive, multimodal learning ecosystem that cultivates linguistic awareness and elevates the communicative competence of Gen Z learners.

However, most podcasts used in English language learning remain audio-only, lacking visual supports such as synchronized texts or contextual illustrations. In fact, recent studies highlight that multimodality, which combines audio, visual, and textual elements, holds significant potential in digital learning (Mayer, 2024). For example Guo et al. (2024). In their bibliometric analysis, they emphasised that research on multimodal language teaching has grown substantially, yet many dimensions remain unexplored, particularly in the context of English listening. In other words, while podcast media have gained popularity, the audio-only format may not be sufficiently optimal for Gen Z learners. Within the EFL learning context, a multimodal approach, integrating audio, text, and visuals, has the potential to maximize Gen Z students' information absorption, as they are inherently audiovisual learners.

Furthermore, a systematic literature review by Musayeva et al. (2025) involving more than one thousand publications, it was revealed that fewer than ten studies have explicitly examined multimodal podcasts in the context of second language learning, and none have compared multimodal podcasts with audio-only podcasts

in relation to listening self-efficacy. This finding highlights a clear research gap: although existing evidence indicates that podcasts can enhance listening ability, the key variable, listening self-efficacy, has received insufficient scholarly attention, and no comparative studies have yet investigated the effectiveness of multimodal versus audio-only podcast formats in this context.

This research gap is critical to outcome-oriented English language instruction. Listening self-efficacy, defined as learners' belief in their ability to comprehend spoken input in English, has been shown to influence self-regulated learning strategies, motivation, and persistence in practice (Pei et al., 2023). Without improving self-efficacy, gains in listening skills alone are insufficient to cultivate autonomous and confident learners capable of engaging with authentic L2 environments. Within the context of Indonesian Gen Z EFL learners, accustomed to audiovisual media and digital interactivity, multimodal podcast learning offers a promising opportunity to design learning experiences that not only enhance input processing but also foster stronger listening self-belief. However, since no empirical studies have yet explored this model through a comparative lens, there remains an urgent need for focused research in this area.

Accordingly, this study aims to address a critical gap in language learning literature by introducing the multimodal podcast format as a potential intervention to enhance listening self-efficacy among Gen Z EFL learners. The significance of this study lies not only in its theoretical contribution, but also in its practical value. English language lecturers and curriculum developers can utilize the findings to design learning materials that are more effective, adaptive, and aligned with the characteristics of today's learners. The study further implies that multimodal podcasts can serve as a new pedagogical reference in English listening instruction, transforming them from mere supplementary media into a strategic tool for L2 literacy development in the digital generation.

This study is designed with the following research objectives: 1) to statistically examine whether Gen Z EFL students who use multimodal podcasts demonstrate a significant improvement in listening self-efficacy compared to those who use audio-only podcasts; 2) to explore learners' experiences from the multimodal podcast group and identify the factors they perceive as strengthening or hindering their listening self-efficacy; and 3) to develop an integrative conceptual model that links multimodal instructional design, listening self-efficacy, integrative motivation, and L2 listening learning outcomes. This study holds a clear and urgent position of novelty. First, from a theoretical perspective, it integrates four major frameworks, Bandura, Mayer, Vygotsky, and Gardner, that have rarely been combined in previous EFL research. Second, from a methodological standpoint, it adopts a mixed-methods design that enables a comprehensive analysis of the intervention's effectiveness and underlying mechanisms, rather than relying solely on simple quantitative outcomes. Third, from an empirical perspective, this study represents the first in Southeast Asia to directly compare the effects of multimodal versus audio-only podcasts on Gen Z EFL learners' listening self-efficacy. The

findings of this research are expected to serve as an essential reference for language lecturers and curriculum developers in the digital learning era and to stimulate further study in the field of technology-enhanced language learning.

LITERATURE REVIEW

Previous Related Study

Several prior studies have examined the integration of podcast-based learning in English as a Foreign Language (EFL) context, yet few have focused on its multimodal dimension or its psychological effects on learners' self-efficacy. Khoiriyah et al. (2024) investigated the use of educational podcasts to enhance listening comprehension among university students and found that continuous exposure to authentic audio input increased learners' understanding and motivation. However, the study did not incorporate multimodal elements such as subtitles or visual scaffolds. Similarly, Pei et al. (2023) explored the role of metacognitive listening strategies in developing EFL learners' listening ability and self-efficacy, emphasizing that multimodal feedback environments promote both awareness and confidence.

Another relevant study by Wang (2023) developed and validated a new scale to measure listening self-efficacy among EFL learners and concluded that exposure to diverse listening materials significantly improved learners' confidence in understanding authentic spoken English. Meanwhile, Amaliah et al. (2025) examined podcast-based learning from a semantic-pragmatic perspective and demonstrated that podcasts promoting polysemy awareness enhanced EFL learners' communicative performance and autonomy. These findings collectively highlight the pedagogical value of podcasts in supporting listening development but also underscore a noticeable research gap: limited investigation into how multimodal podcasts, combining audio, textual, and visual input, affect the self-efficacy and motivational processes of Generation Z learners.

Multimodal Podcast Learning

First, from the perspective of Social Cognitive Theory, Bandura (1997) emphasizes that self-efficacy refers to one's belief in one's ability to organize and execute the actions required to achieve specific outcomes. Supporting this view Pei et al. (2023), demonstrated that EFL learners with higher self-efficacy exhibit greater persistence and employ more adaptive learning strategies when tackling listening tasks. Within the context of multimodal podcast learning, students who successfully complete audio-visual modules gain mastery experiences, which in turn strengthen their self-efficacy. This conceptual foundation directs the present study to explore how the multimodal format, compared to the audio-only format, can generate richer mastery experiences and, consequently, enhance learners' confidence in listening comprehension.

Second, the concept of multimodal learning is based on Mayer's (2024) *Cognitive Theory of Multimedia Learning*, which posits that individuals learn more effectively when information is processed through both auditory and visual channels, thereby reducing cognitive overload and promoting deeper learning. The Cognitive Theory of Multimedia Learning (CTML) was proposed by (Mayer, 2024) posits that individuals achieve a better understanding when information is presented simultaneously through both auditory and visual channels. Mayer's (2024) research highlights that instructional designs incorporating dual-channel processing, appropriate segmentation, and the integration of audio, text, and visuals significantly enhance comprehension and retention. This is particularly relevant for Gen Z, who are accustomed to consuming audio and visual materials concurrently. Supporting this, Guo et al. (2024) noted through their bibliometric analysis that research on multimodal learning has shown an increasing trend, yet L2 listening skills remain underexplored. Therefore, this study proposes that multimodal listening podcasts can facilitate more efficient cognitive processing and strengthen learners' self-efficacy, as students can "see and hear" rather than merely listen.

Third, learner autonomy and performance are explained through Vygotsky's (1934) *Sociocultural Theory*, which views learning as a socially mediated process shaped by cultural tools and interaction. Vygotsky's (1987) Sociocultural Theory emphasizes that learning occurs through social mediation, wherein cultural tools and interactions facilitate cognitive development. Within this framework, multimodal podcasts function as *cultural tools* that present narratives, dialogues, native speakers' cultural contexts, contextual visuals, and authentic digital interactions. In the context of German language learning, Massa (2024) found that integrating digital cultural tools deepened student engagement. In line with these findings, Amaliah et al. (2025) asserted that exposure to podcasts containing cultural contexts and polysemous meanings in authentic speech not only enhances learners' collocational competence but also broadens their semantic awareness and social connectedness to the target language. Accordingly, this study views multimodal podcasts as a bridge between learners and the English-speaking community, rather than as passive media. This perspective adds a socio-cognitive dimension that enriches the understanding of students' self-efficacy in second-language learning.

Fourth, Gardner's (1985) Integrative Motivation Model underscores that learners' orientation to connect and integrate with the target language community is a key factor in achieving long-term language learning success. Liu & Dong (2023) reported that integrative motivation is positively correlated with self-efficacy and language performance. Multimodal podcasts that incorporate cultural elements, visual representations, and real-life audio can strengthen Gen Z students' integrative orientation, thereby enhancing their learning commitment and confidence in listening comprehension.

METHOD

Design and Sample

This study used a convergent parallel mixed methods design that combined a quantitative quasi-experimental approach with a qualitative phenomenological inquiry. The quantitative component used a non-equivalent control group design to compare students who learned through multimodal podcasts and those who learned through audio-only podcasts under authentic classroom conditions where random assignment was not feasible. The qualitative component explored students' lived experiences during the intervention, focusing on confidence, motivation, and engagement. Both strands were analyzed separately and later integrated to explain not only the extent of improvement but also the reasons behind it.

The participants were 60 fourth-semester students from the English Education Department at Universitas Negeri Makassar in the 2024/2025 academic year. They were aged 18 to 22 and had completed foundational courses in listening and speaking needed to participate effectively in the learning activities. Purposive sampling ensured that participants met criteria such as intermediate proficiency, active enrollment in Listening courses, and stable internet access. The students were divided equally into an experimental group and a control group. The experimental group received multimodal podcasts that combined audio, subtitles, transcripts, and visual elements, while the control group received audio-only podcasts with the same spoken content. Groups were matched based on initial proficiency to ensure comparability.

Instruments and Procedures

The primary instrument was the Listening Self-Efficacy Scale (LSES), adapted from Wang (2023), which consisted of 20 items rated on a six-point Likert scale. The scale measured confidence in understanding general topics, identifying details, interpreting implied meaning and intonation, and maintaining concentration. A pilot test showed high internal consistency, and expert review established strong content validity. The second instrument consisted of podcast learning modules designed in two formats. The multimodal version included audio narration paired with subtitles, transcripts, and visual illustrations, while the audio-only version contained only the spoken content. Both versions followed multimedia learning principles and were delivered through the Syam-OK learning platform. The third instrument was a semi-structured interview guide used to collect qualitative data from 12 students in the experimental group, selected based on variation in learning outcomes.

The procedures included four main stages carried out over eight weeks. First, both groups completed the LSES to establish baseline equivalence. Second, during the intervention, the experimental group used multimodal podcasts supported by reflective tasks and weekly discussions, while the control group used audio-only podcasts with the same content. Students also maintained learning logs to capture

their perceptions and challenges. Third, both groups completed the LSES again at the end of the intervention to assess changes in listening self-efficacy. Fourth, semi-structured interviews were conducted with selected students from the experimental group to gather in-depth insights into their experiences.

Data Analysis

The quantitative data from the LSES were analyzed using SPSS version 28. Tests of normality and homogeneity confirmed that the data met the assumptions for parametric testing. A paired-samples t-test was used to examine changes within each group, and an independent-samples t-test was used to compare the difference in gains between the two groups. Effect sizes were calculated using Cohen's *d* to determine the magnitude of the treatment effect, with significance set at $p < .05$. The qualitative data were analyzed using thematic analysis based on the steps proposed by Braun and Clarke (2022). Interview transcripts were coded separately by two coders to ensure reliability, and inter-coder agreement was high. Three key themes emerged from the analysis: strengthened confidence through repeated successful comprehension, increased motivation due to authentic and visually contextualized content, and greater learner autonomy supported by multimodal scaffolds. Trustworthiness was addressed through member checking, peer review, triangulation, and documentation of the analysis process. After both analyses were completed, the findings were integrated through side-by-side comparison, allowing the quantitative results to be interpreted alongside the qualitative explanations. This integration provided a comprehensive understanding of how and why multimodal podcasts enhanced listening self-efficacy among the participating students.

RESULT AND DISCUSSION

Descriptive and Inferential Statistics

The quantitative findings provide empirical evidence of the improvement in L2 listening self-efficacy among Gen Z EFL learners who were exposed to multimodal podcast learning compared to those who received audio-only instruction. Pre-test and post-test data were analyzed using paired-samples and independent-samples *t*-tests to examine both within-group and between-group differences.

Table 1. Descriptive Statistics and t-test Results for Listening Self-Efficacy

Group	N	Pre-test Mean (SD)	Post-test Mean (SD)	t(df)	p-value	Cohen's d
Multimodal Podcast	30	3.26 (0.72)	4.02 (0.68)	7.91 (29)	< .001	1.15
Audio-only Podcast	30	3.28 (0.69)	3.25 (0.70)	0.14 (29)	.887	0.05

As shown in Table 1, both groups started with almost identical pre-test scores, indicating similar baseline self-efficacy levels before the intervention. However,

after eight weeks of instruction, the experimental group using multimodal podcasts showed a substantial increase in mean score ($M = 4.02$, $SD = 0.68$), while the control group's post-test mean remained statistically unchanged ($M = 3.25$, $SD = 0.70$).

The independent-samples *t-test* confirmed a significant difference between groups, $t(58) = 7.91$, $p < .001$, with a large effect size (Cohen's $d = 1.15$). This result suggests that exposure to multimodal input, including subtitles, interactive transcripts, and visuals, produced a measurable improvement in students' confidence when listening to English. These findings support Mayer's (2024) *Cognitive Theory of Multimedia Learning*, which emphasizes that combining audio and visual channels reduces cognitive load and enhances information retention. Furthermore, the result aligns with Wang (2023), who demonstrated that self-efficacy in EFL listening can be significantly improved when learners engage with multimodal comprehension materials rather than audio-only input. The improvement observed in this study thus highlights the strong relationship between multimodal scaffolding and learners' confidence development, validating the first research objective: that multimodal podcast instruction leads to a statistically significant enhancement in listening self-efficacy among Gen Z EFL learners.

Thematic Findings

To complement the statistical findings, qualitative data were analyzed through thematic analysis to uncover how learners experienced and perceived multimodal podcast learning. Analysis of twelve interview transcripts revealed three major themes that illuminate the mechanisms behind the quantitative improvement:

1) Strengthened Confidence through Repeated Mastery Experiences

Participants reported that multimodal podcasts helped them gain confidence through repetition and verification of comprehension. The availability of subtitles and interactive transcripts allowed learners to check and confirm their understanding of the spoken text, leading to repeated successes, which Bandura (1997) calls mastery experiences, the most powerful source of self-efficacy.

"The subtitles really helped me check what I missed. I became more confident because I could replay and compare my understanding."

(Participant 6, Female, 20)

"I used to feel nervous when listening to fast English audio. But after using these podcasts, I realized I could actually understand more when I watched and listened together." (Participant 2, Male, 19)

This theme supports the argument that multimodal exposure creates repeated instances of successful comprehension, which, in turn, strengthens students' belief in their listening ability.

2) Increased Motivation through Authentic Topics

A recurring theme among participants was the role of authentic, relevant content in maintaining motivation. Learners noted that the podcast topics centered on real-world issues and on daily communication, which helped them connect emotionally with the material. This supports Gardner's (1985) *Integrative Motivation Theory*, which suggests that meaningful exposure to target-culture communication enhances motivation and self-confidence.

"The topics were very interesting and connected to our real life, like culture, habits, and education. It made me feel that English is not only a subject, but a tool for real communication." (Participant 8, Female, 21)

"When I listened to podcasts about students' experiences abroad, I felt motivated to learn more English." (Participant 4, Male, 20)

The engagement generated by authentic topics increased both affective involvement and the willingness to persist in listening practice. This finding reinforces Pei et al. (2023), who observed that authentic tasks improve learners' self-regulation and motivation.

3) Greater Learner Autonomy Fostered by Multimodal Scaffolds

Students consistently mentioned that multimodal features, such as subtitles, visual cues, and replay functions, helped them control their learning pace. This autonomy allowed them to adjust their listening strategies independently, reflecting the principles of Vygotsky's (1987) *Sociocultural Theory* in which learning tools mediate self-regulated development.

"I liked that I could pause or repeat parts I didn't understand. It made me feel I was in control of my own learning." (Participant 10, Female, 20)

"The visuals and texts helped me focus on difficult phrases and remember them later." (Participant 3, Male, 19)

This theme indicates that multimodal podcasts serve not only as instructional media but also as metacognitive tools that foster independent, self-paced learning, an essential trait for Gen Z learners. As Amaliah et al. (2025) assert, podcasts designed with contextual linguistic awareness promote both cognitive depth and learner autonomy in EFL classrooms.

Both quantitative and qualitative findings collectively reveal that multimodal podcast learning significantly enhances L2 listening self-efficacy among Generation Z EFL learners. The quantitative data showed a measurable improvement with a large effect size (Cohen's $d = 1.15$), while qualitative insights explained the mechanisms behind this improvement, namely *mastery experiences*, *authentic engagement*, and *learner autonomy*. Together, these findings demonstrate that multimodal podcast learning not only improves comprehension but also builds the psychological confidence necessary for sustainable language learning in the digital era.

The results of the quantitative analysis clearly demonstrated that multimodal podcast instruction produced a statistically significant increase in L2 listening self-efficacy among Gen Z EFL learners. This finding is best understood through Bandura's (1997) Social Cognitive Theory of Self-Efficacy, which identifies four main sources of self-efficacy: *mastery experience*, *vicarious experience*, *verbal persuasion*, and *emotional arousal*. First, the improvement in learners' confidence reflects the dominance of mastery experiences, as learners repeatedly achieved successes when listening comprehension became easier through multimodal aids. When students used subtitles and interactive transcripts, they could verify understanding and experience consistent mastery over increasingly complex audio input. These repeated successful experiences built psychological assurance that they could understand authentic spoken English.

Second, vicarious experience was also stimulated through exposure to authentic dialogues and cultural expressions embedded in the podcasts. By observing native speakers' tone, stress, and interaction styles, students could model comprehension strategies that enhanced their listening self-efficacy. This aligns with Wang (2023), who found that model-based learning through multimedia input strengthens learners perceived competence in EFL listening. Third, the multimodal podcast environment provided verbal persuasion through immediate comprehension feedback, and the visible alignment between audio and text reinforced the belief that comprehension was attainable. As Pei et al. (2023) suggest, real-time confirmation of understanding through multimodal cues provides psychological reinforcement, elevating motivation and self-belief. Finally, emotional regulation played an important role. The inclusion of visual elements and contextualized narratives reduced anxiety associated with listening comprehension, a frequent affective barrier among EFL learners. Students who initially reported anxiety toward fast-paced speech later described the multimodal support as calming and reassuring. This confirms Bandura's assumption that reduced negative emotional states amplify self-efficacy beliefs by stabilizing affective control.

These results are also consistent with Mayer's (2024) *Cognitive Theory of Multimedia Learning (CTML)*, which posits that information is processed through dual channels, visual and auditory, and that learning is most effective when both are used efficiently. In the multimodal podcast context, learners accessed spoken input (audio channel) while simultaneously viewing text and illustrations (visual channel). This dual processing not only reduced *cognitive load* but also enhanced memory retention and meaningful encoding. In contrast, the audio-only group relied solely on the auditory channel, increasing processing strain and lowering comprehension confidence.

Furthermore, these findings bolster Gardner's (1985) Integrative Motivation Theory, indicating that authentic and culturally rich podcast topics foster greater emotional engagement among learners in language acquisition. Participants within the experimental group expressed feeling "closer to real English speakers" and demonstrated increased motivation to utilize English as a means of communication,

thereby exemplifying the interplay between self-efficacy and integrative motivation. Consequently, the quantitative data substantiate the psychological principles articulated by Bandura and the cognitive frameworks outlined by Mayer, while also broadening their scope to encompass digital podcast-based instructional methods.

The qualitative analysis deepened understanding of *how* multimodal podcasts shape learners' self-efficacy and motivation. Three emergent themes, *mastery experiences*, *authentic engagement*, and *learner autonomy*, correspond directly to the theoretical foundations of the study. First, the theme "*strengthened confidence through repeated mastery experiences*" validates Bandura's principle that mastery experiences are the most reliable predictors of self-efficacy development. Students repeatedly mentioned that being able to *pause*, *replay*, and *read transcripts* helped them achieve incremental success, reinforcing their belief in their listening capabilities. This behavioral pattern illustrates a clear cycle of self-regulated learning, in which control over the learning process leads to emotional satisfaction and sustained engagement.

Second, the theme "*increased motivation through authentic topics*" emphasizes the importance of content relevance and social connectedness. As Amaliah et al. (2025) argue, podcast-based learning that embeds contextual and cultural meaning enhances not only linguistic competence but also learners' sense of participation in the language community. The same effect was observed in this study: learners reported feeling part of global conversations when listening to real-world podcast themes. This indicates a powerful interaction between Mayer's cognitive dual-processing and Gardner's integrative motivational orientation, in which multimodal input provides both cognitive scaffolding and socio-emotional relevance. Third, the theme "*greater learner autonomy fostered by multimodal scaffolds*" resonates with Vygotsky's (1987) *Sociocultural Theory*, which emphasizes that learning is mediated through cultural tools and social interaction. In this context, multimodal features such as subtitles and visual aids functioned as mediational tools that enabled learners to control their learning pace and self-monitor comprehension. This empowerment process transforms students from passive listeners into active meaning-makers, reinforcing their sense of agency, a defining characteristic of Generation Z learners. These findings collectively support Riazi's (2024) assertion that integrating quantitative and qualitative evidence leads to deeper pedagogical insight, especially in technology-mediated language learning. The complementarity of data types clarifies not only *that* multimodal design works, but *why* it enhances learner confidence and motivation.

Theoretically, this study extends Bandura's self-efficacy framework into the domain of digital multimodal learning, providing empirical evidence that technology-mediated *mastery experiences* and *visual-verbal congruence* can significantly shape learners' affective development. It also advances Mayer's CTML by demonstrating that multimodal integration in asynchronous audio learning (i.e., podcasts) can yield cognitive benefits comparable to synchronous

classroom multimedia instruction. Pedagogically, the study provides substantial implications for EFL educators. First, teachers should design podcast-based activities that combine audio, textual, and visual modalities to support students with diverse learning preferences. Second, instructors can integrate *self-reflection tasks* and *listening logs* to promote mastery recognition and self-regulation. Third, curriculum developers should emphasize authentic and context-rich podcast topics to foster integrative motivation and cultural connectedness. Finally, this study supports the move toward learner autonomy, empowering students to manage their own pace, comprehension strategies, and motivational focus in digital learning environments. Practically, these findings also help explain why audio-only formats may no longer suffice for Gen Z learners, whose cognitive habits are shaped by constant exposure to multimodal input. As such, adopting multimodal podcast learning represents both a pedagogical adaptation and an evolution of digital EFL instruction in the post-pandemic era.

CONCLUSION

This study demonstrates that multimodal podcast learning has a significant positive impact on Gen Z EFL learners' listening self-efficacy, showing stronger gains than audio-only instruction. When audio is supported with subtitles and relevant visual cues, learners are better able to follow meaning, confirm understanding, and manage difficulties during listening. These repeated moments of successful comprehension help build confidence and contribute to a sense of progress, which is central to strengthening self-efficacy. Learners also reported feeling more engaged and motivated during the multimodal lessons, indicating that the multimodal format is more aligned with the learning preferences of digital-native students. Grounding these findings in Bandura's self-efficacy theory and Mayer's multimedia learning principles helps explain why the multimodal approach was effective. The visuals and textual support reduce cognitive load by distributing processing across different channels, allowing students to focus more on meaning rather than decoding. At the same time, the multimodal environment encourages self-regulated learning behaviors, such as replaying audio, checking transcripts, and monitoring comprehension. These behaviors are essential for developing long-term listening competence, not just short-term task performance. The results of this study offer both theoretical and practical contributions. Theoretically, the study confirms that integrating multimodal input can enhance learners' belief in their listening ability and support more sustainable language development. Practically, it suggests that multimodal podcasts can be adopted as an accessible instructional tool in EFL classrooms, requiring no advanced technology beyond standard learning platforms. Teachers can use multimodal podcasts to create more inclusive learning conditions that support diverse abilities and learning styles. Future research may extend this work by examining different podcast genres, longer intervention durations, or varied proficiency levels to better understand how multimodality can be optimized in different instructional contexts.

REFERENCES

- Amaliah, S., Wahyuni, Y., & Sakkir, G. (2025). Exploring The Role Of Polysemy Awareness For “Get” And “Make” In Podcasts On EFL Learners’ Collocational Speaking Ability. *ARRUS Journal of Social Sciences and Humanities*, 5(2). <https://doi.org/10.35877/soshum3924>
- Bandura, Albert. 1997. Self-efficacy - The Exercise of Control, New York: W.H. Freeman and Company.
- Braun, V., & Clarke, V. (2022). Thematic analysis: A practical guide. London: SAGE Publications.
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. Edward Arnold.
- Guo, X., Chen, S., & Guo, Y. (2024). Advancing multimodal teaching: a bibliometric and content analysis of trends, influences, and future directions. *Humanities and Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-04254-0>
- Hrdy, M., Tarver, E. M., Lei, C., Moss, H. C., Wong, A. H., Moadel, T., Beattie, L. K., Lamberta, M., Cohen, S. B., Cassara, M., Hughes, M. D., De Castro, A., Sahi, N., & Chen, T. H. (2024). Applying simulation learning theory to identify instructional strategies for Generation Z emergency medicine residency education. *AEM Education and Training*, 8(S1), S56–S69. <https://doi.org/10.1002/aet2.10981>
- Khoiriyah, N., Ummah, N., Hidayani, S., Sultan Maulana Hasanuddin Banten English Education, U., & Graduate Program, P. (2024). *The Efficacy of Podcast in EFL Learning: A Systematic Literature Review*. 1(2). <https://jurnal.untan.ac.id/index.php/jole>
- Liu, X., & Dong, M. (2023). Exploring the relative contributions of learning motivations and test perceptions to autonomous English as a foreign language learning and achievement. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1059375>
- Massa, A. (2024). From the Sociocultural Theory by Vygotsky to its Didactic Application: A Case Study of German Language Learning in an Italian Context. *Language Teaching Research Quarterly*, 46, 157–173. <https://doi.org/10.32038/ltrq.2024.46.12>
- Mayer, R. E. (2024). The Past, Present, and Future of the Cognitive Theory of Multimedia Learning. *Educational Psychology Review*, 36(1). <https://doi.org/10.1007/s10648-023-09842-1>
- Musayeva, N., Sabirova, G., Xadjikurbanova, G., Yuldasheva, S., Akhmadaliyev, A., & Djumabaeva, J. (2025). Developing Listening Comprehension Using Podcasts in Second Language Education: A Systematic Overview of Recent Pedagogical Innovations (2019–2025). *Forum for Linguistic Studies*, 7(8). <https://doi.org/10.30564/fls.v7i8.10287>
- Pei, T., Suwanthep, J., & Lu, H. (2023). The effect of self-directed online metacognitive listening practice on Chinese EFL learners’ listening ability, metacognition, and self-efficacy. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1285059>

- Princes, E., Soeryanto, N., & Romprasert, S. (2024). Exploring Gen-Z Learning Preferences: A Comparative Study of Traditional, Online, and Blended Learning Model. In *JOURNAL OF MULTIDISCIPLINARY ISSUES Journal Website: www.jmis.site J. Multidisc. Issues* (Vol. 4, Issue 1). www.jmis.site
- Riazi, A. M. (2024). Methodological pluralism and mixed-methods integration in language education research. *Educational Research Review*, 39, 100579. <https://doi.org/10.1016/j.edurev.2024.100579>
- Teng, F. (2024). Research methods for applied linguistics: Integrating quantitative and qualitative approaches. *TESOL Quarterly*, 58(1), 102–120. <https://doi.org/10.1002/tesq.3221>
- Vygotsky, L.S. (1987). *Vygotsky's Sociocultural Theory*. New York: Plenum
- Wang, Y. (2023). Measuring EFL learners' listening self-efficacy: Development and validation of a new scale. *System*, 118, 102960. <https://doi.org/10.1016/j.system.2023.102960>
- Xu, C., Liu, Y., & Teng, F. (2023). Mixed-methods approaches in applied linguistics research: Recent trends and future directions. *Language Teaching Research*, 27(4), 565–589. <https://doi.org/10.1177/13621688221101073>