el-Huda 17(1) (2026)

Jurnal el-Huda: Jurnal Pendidikan, Sosial, dan Keagamaan

https://ejurnal.iaiqh.ac.id/index.php/el-huda

Using English Songs to Enhance Academic Listening Skills of University Students

Deasy Yunita Siregar[1], Amanda Alya Putri[2], Nadira Citra Ayu[3], Raia Fathi Israra[4], Ari Rajasa Tampubolon[5]
[1,2,3,4,5] Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

DOI: https://doi.org/10.59702/el-huda.v17i01.395

Jurnal Info
Dikirim: 27/12/2025
Revisi: 09/04/2025
Diterima: 15/04/2025

Korespondensi:

Phone: +62……..

Abstract: Academic listening remains a major challenge for EFL university students due to limited exposure to authentic spoken English and low engagement in listening activities. This study aims to examine the effectiveness of English songs in improving academic listening skills, with a specific focus on university-level learners in Indonesia. A classroom-based quasi-experimental design with a pre-test and post-test was employed involving 26 students of the English Education Department at UIN Sumatera Utara. Data were collected using two instruments: (1) a lyric comprehension test consisting of 20 items assessing students' ability to identify main ideas, specific information, and vocabulary in context, and (2) a singing performance rubric measuring pronunciation, intonation, fluency, and content understanding. After four sessions of song-based instruction, the results showed an increase in the mean score from 74.6 (pre-test) to 94.2 (post-test), indicating an improvement in students' listening performance. However, the findings are interpreted descriptively without inferential statistical testing. This study contributes to the existing literature by focusing specifically on academic listening at the university level, an area that has received limited attention compared to general listening. The results suggest that English songs can serve as a supportive medium to enhance listening engagement and comprehension in EFL classrooms.
Keywords: English songs, academic listening, quasi-experimental, EFL students, listening comprehension

Abstrak: Menyimak akademik masih menjadi tantangan utama bagi mahasiswa EFL (English as a Foreign Language) di perguruan tinggi karena terbatasnya paparan terhadap bahasa Inggris lisan yang autentik serta rendahnya keterlibatan dalam aktivitas menyimak. Penelitian ini bertujuan untuk mengkaji efektivitas lagu berbahasa Inggris dalam meningkatkan keterampilan menyimak akademik, dengan fokus khusus pada mahasiswa tingkat universitas di Indonesia. Penelitian ini menggunakan desain kuasi-eksperimental berbasis kelas dengan pre-test dan post-test yang melibatkan 26 mahasiswa Program Studi Pendidikan Bahasa Inggris di UIN Sumatera Utara. Data dikumpulkan menggunakan dua instrumen: (1) tes pemahaman lirik yang terdiri dari 20 soal untuk menilai kemampuan mahasiswa dalam mengidentifikasi ide pokok, informasi spesifik, dan kosakata dalam konteks, serta (2) rubrik penilaian performa bernyanyi yang mengukur pelafalan, intonasi, kelancaran, dan pemahaman isi. Setelah empat sesi pembelajaran berbasis lagu, hasil menunjukkan peningkatan nilai rata-rata dari 74,6 (pre-test) menjadi 94,2 (post-test), yang mengindikasikan adanya peningkatan kemampuan menyimak mahasiswa. Namun, temuan ini diinterpretasikan secara deskriptif tanpa pengujian statistik inferensial. Penelitian ini memberikan kontribusi terhadap literatur yang ada dengan memfokuskan pada menyimak akademik di tingkat perguruan tinggi, yang masih jarang mendapat perhatian dibandingkan dengan menyimak umum. Hasil penelitian menunjukkan bahwa lagu berbahasa Inggris dapat menjadi media yang mendukung untuk meningkatkan keterlibatan dan pemahaman menyimak dalam kelas EFL.
Kata kunci: lagu bahasa Inggris, menyimak akademik, kuasi-eksperimental, mahasiswa EFL, pemahaman menyimak

Introduction

Academic listening occupies a foundational position in university-level education, particularly within English as a foreign language (EFL) contexts where students are expected to navigate increasingly complex spoken discourse across a range of academic settings. At the university level, students are regularly required to comprehend lectures delivered at natural speed, follow seminar discussions involving multiple speakers, interpret audio-visual academic materials, and engage in peer-to-peer academic conversations all of which place considerable demands on their listening competence. Unlike the more forgiving

Email: amandaalyaputri16@gmail.com

conditions of everyday conversation, academic listening requires sustained concentration, rapid real-time processing, and the ability to simultaneously decode phonological form, construct meaning, and retain information for later use. Listening, in this sense, functions as the primary gateway through which learners receive language input, and its role in supporting the development of other academic language skills speaking, reading, and writing cannot be overstated. Students who lack sufficient listening proficiency are likely to face compounding difficulties across all areas of their academic language development, not merely in listening tasks themselves.

Despite its centrality, academic listening remains one of the most underexplored areas in EFL research, particularly at the tertiary level. The majority of existing studies on listening instruction have focused on general listening skills, often within secondary school contexts, and have relied on simplified or scripted audio materials that do not reflect the complexity of authentic academic spoken language. This gap between classroom listening practice and the real demands of academic listening environments leaves many university students underprepared for the listening challenges they encounter in their studies. In Indonesia, this problem is especially pronounced, as opportunities for authentic English listening exposure outside the formal classroom setting are relatively limited, and instructional materials used in many university classrooms continue to prioritize reading and writing over systematic listening development.

Understanding why academic listening is so difficult requires engaging with the theoretical frameworks that explain how listeners process spoken language. Rost (2011) describes listening as a dynamic, active process involving four interrelated operations: receiving the acoustic signal, constructing meaning from the decoded input, negotiating meaning in relation to prior knowledge and context, and responding to the communicative purpose of the message. This multidimensional view of listening highlights that successful comprehension is not simply a matter of hearing sounds correctly it requires the listener to engage simultaneously at the phonological, lexical, syntactic, and discourse levels. In academic settings, where vocabulary is dense, syntactic structures are complex, and information is delivered at pace, the demands placed on each of these processing levels are considerably higher than in casual conversation.

Listening theory further distinguishes between two complementary modes of processing: bottom-up and top-down. Bottom-up processing refers to the data-driven decoding of the speech signal recognizing phonemes, identifying word boundaries, parsing grammatical structures, and assembling meaning incrementally from the acoustic input. Top-down processing, by contrast, draws on the listener's existing knowledge background knowledge, contextual expectations, familiarity with discourse conventions, and schema about topic and genre to interpret incoming information and fill gaps in comprehension. Effective listeners move fluidly between these two processing modes, using each to compensate for deficiencies in the other. For EFL learners, however, both modes present distinct challenges. Bottom-up processing is hampered by unfamiliarity with natural phonological processes such as connected speech, elision, assimilation, and reduction, which cause the acoustic realization of words to differ substantially from their citation forms. Top-down processing is weakened by limited background knowledge of academic topics and genres, restricted vocabulary range, and unfamiliarity with the discourse patterns of academic spoken English. The interaction of these two sets of difficulties makes academic listening a genuinely complex cognitive task for many university EFL students.

The role of working memory in listening comprehension adds another layer of complexity. Listening, unlike reading, is time-constrained and ephemeral the acoustic signal disappears as soon as it is produced, leaving no opportunity for review or re-reading. Listeners must hold earlier portions of an utterance in working memory while continuing to process incoming speech, integrating new information with what has already been received to construct a coherent representation of meaning. This places heavy demands on cognitive resources, and when those resources are insufficient whether due to limited proficiency, cognitive overload, or emotional interference comprehension breaks down. In academic contexts, where utterances are long, vocabulary is unfamiliar, and information density is high, the working memory demands of listening are particularly acute. Learners who are struggling at the phonological decoding level have fewer cognitive resources available for higher-order meaning construction, which is one reason why improving bottom-up processing skills can have cascading positive effects on overall comprehension.

Affective factors constitute another critical dimension of listening difficulty that is frequently overlooked in purely cognitive accounts of the skill. Krashen's (1982) Affective Filter Hypothesis proposes that second language acquisition is mediated by affective variables including anxiety, motivation, and self-confidence. When the affective filter is high when learners feel anxious, unmotivated, or lacking in confidence the processing of comprehensible input is blocked, and acquisition is impeded even when the learner is exposed to appropriate language. Listening is widely reported to be the language skill that generates the most anxiety among EFL learners, partly because of its real-time, uncontrollable nature. Unlike writing, where learners can pause, revise, and monitor their output, listening offers no such opportunities for self-correction.

The pressure to understand immediately, without the ability to slow down or replay, creates conditions in which anxiety can escalate rapidly, further reducing the cognitive resources available for comprehension. In academic contexts, where the consequences of misunderstanding are significant missed information, failed assessments, inability to participate in discussion this anxiety tends to be particularly intense. Any instructional approach that succeeds in reducing listening-related anxiety while simultaneously developing comprehension skills therefore addresses two interrelated barriers to academic listening development simultaneously.

Given the multidimensional nature of academic listening difficulty, it is clear that instructional approaches which address only one dimension such as providing more listening practice with scripted materials are unlikely to produce comprehensive development. What is needed are instructional strategies that simultaneously engage learners at the phonological,

lexical, and discourse levels; reduce affective barriers to engagement; provide authentic language input that reflects the features of real spoken English; and create sufficient repetition and scaffolding to allow learners to process that input deeply rather than superficially. In this context, the use of English songs as an instructional medium has attracted growing interest from language educators and researchers as a potentially powerful approach to listening development.

Songs represent a distinctive form of authentic language input. Unlike academic lectures or dialogue recordings produced for pedagogical purposes, songs are created by and for native speakers within specific cultural and artistic contexts, meaning they contain natural phonological features stress patterns, rhythm, intonation contours, connected speech phenomena that textbook audio materials typically eliminate or simplify. At the same time, unlike unstructured authentic input such as radio broadcasts or films, songs possess structural features repetition, rhyme, regular meter, and melodic predictability that provide learners with a degree of scaffolding that makes the input more manageable. The repetitive nature of song structure means that learners encounter the same linguistic forms multiple times within a single listening experience, which increases the likelihood of successful phonological decoding and vocabulary retention. Furthermore, the melody serves as a mnemonic device, helping learners retain the phonological form of words and phrases more durably than they might from a single encounter in a lecture or dialogue.

The potential of songs to reduce affective barriers is equally significant. Music is widely recognized across cultures as having emotional and motivational properties that distinguish it from other forms of language input. In classroom settings, the introduction of songs has been consistently associated with increased learner engagement, reduced anxiety, and heightened motivation all of which, according to Krashen's Affective Filter Hypothesis, create more favorable conditions for language acquisition. When students perceive a learning activity as enjoyable and low-stakes, they are more willing to take risks, attempt challenging tasks, and persist through difficulty. This is particularly relevant for listening, where anxiety is already elevated and where avoidance behavior mentally disengaging from difficult input is a common and counterproductive response. A learning environment in which authentic language input is delivered through an engaging and emotionally resonant medium like music is likely to elicit a qualitatively different kind of learner engagement than one in which students are required to listen to academic passages under test-like conditions.

Mayer's (2001) Cognitive Theory of Multimedia Learning provides an additional theoretical basis for understanding why song-based instruction may be particularly effective. Mayer's theory holds that learners process information through two separate channels auditory and visual and that learning is deepened when both channels are engaged simultaneously with complementary, non-redundant information. In a song-based listening lesson, students receive auditory input through the music and sung lyrics, visual input through printed lyrics or on-screen text, and in production-oriented activities such as singing performance kinesthetic and expressive engagement through the act of reproducing the language. This multimodal engagement distributes cognitive processing across multiple channels, reducing the overload that can occur when all information must be processed through a single modality. The result is a richer and more durable representation of the linguistic material in memory, which should translate into stronger comprehension and retention outcomes.

Previous research has provided empirical support for these theoretical predictions. Studies conducted at the secondary school level have consistently found that song-based instruction produces improvements in listening comprehension, vocabulary acquisition, pronunciation accuracy, and learner motivation when compared to conventional instructional approaches. Millington (2011) demonstrated that songs provide a motivating and memorable context for vocabulary learning, while Goh (2017) emphasized the importance of metacognitive strategy instruction in listening development a dimension that song-based learning can support through structured pre-listening, while-listening, and post-listening phases. More recent studies by Amelia et al. (2024), Berlia and Rahmawanti (2022), and Ariani and Iswandi (2021) have confirmed positive effects across a range of EFL settings, with particular strengths in the areas of phonological decoding and vocabulary in context.

However, several significant gaps remain in the existing literature. First, the overwhelming majority of studies on song-based listening instruction have been conducted at the secondary school level, and it is not clear that their findings transfer directly to the university context, where listening demands are more complex, learners are more advanced, and the construct of academic listening differs substantially from general listening in terms of vocabulary load, discourse structure, and cognitive processing requirements. Second, most existing studies have assessed listening outcomes using only receptive comprehension measures, without incorporating productive tasks that could indicate the depth of processing and internalization. Third, research conducted specifically within Indonesian higher education settings remains limited, despite the fact that Indonesia represents one of the largest EFL populations in the world and one in which learners face distinctive challenges related to limited authentic English exposure. These three gaps the focus on secondary rather than tertiary learners, the reliance on purely receptive assessment, and the geographic underrepresentation of Indonesian university contexts constitute the primary research gaps that the present study aims to address.

In response to these identified gaps, this study examines the effectiveness of using English songs as instructional media to improve the academic listening skills of university EFL students at Universitas Islam Negeri Sumatera Utara (UIN Sumatera Utara), Medan, Indonesia. Specifically, the study pursues the following research objectives: first, to measure the improvement in students' academic listening ability following four sessions of song-based instruction, as evidenced by the difference between pre-test and post-test performance; second, to describe the distribution of student performance across listening ability categories before and after the intervention; and third, to explore the extent to which English songs function as an effective and engaging instructional medium for developing cognitively demanding academic listening skills at the university level. By addressing both

receptive comprehension outcomes and productive performance indicators within a pre-test/post-test quasi-experimental framework, this study contributes a more complete and methodologically grounded account of song-based listening instruction than has previously been available in the Indonesian higher education literature.

Method

This study employed a one-group pre-test/post-test quasi-experimental design to examine the effectiveness of English song-based instruction in developing the academic listening ability of university EFL students. The quasi-experimental approach was adopted because the study was conducted with an intact classroom group rather than participants assigned through random selection, a condition that is both practical and methodologically justified in naturalistic educational settings. Unlike descriptive research, which aims to portray existing conditions without intervention, this design was oriented toward evaluating causal relationships: specifically, whether structured exposure to song-based listening activities produced a measurable and statistically meaningful change in student performance over the course of the instructional period.

The pre-test/post-test structure served as the core experimental logic of the study. Administering both measurements to the same group of participants made it possible to establish individual baselines before the intervention began and to quantify the degree of change upon its completion. Any statistically significant difference between the two measurement points is attributed to the effects of the treatment, which in this case consisted of four sessions of song-based academic listening instruction. This design is widely used and methodologically accepted in classroom-based language learning research, particularly when the random assignment of learners to experimental and control conditions is not feasible within real institutional settings.

The participants were 26 undergraduate students enrolled in the TBI 3 class at the English Education Department, Universitas Islam Negeri Sumatera Utara (UIN Sumatera Utara), Medan, Indonesia. The class was selected through purposive sampling based on the researchers' direct access to the learning environment and the class's relevance to the study's instructional objectives. All 26 students were present for both the pre-test and post-test administrations, and no participants were excluded from the analysis. In terms of general language proficiency, the participants ranged from lower-intermediate to upper-intermediate, which is consistent with the typical competency profile of students in their third semester of an undergraduate English education program in Indonesia.

Two instruments were developed to collect quantitative data: a lyric comprehension test and a singing performance assessment rubric. Both instruments were designed to measure academic listening ability as a multi-dimensional construct, covering the capacity to identify specific information, recognize vocabulary in context, interpret meaning, and reproduce language with appropriate pronunciation, intonation, and fluency.

The lyric comprehension test consisted of 20 items constructed around the lyrics of the two songs used in the instructional treatment: Shape of You by Ed Sheeran and Señorita by Shawn Mendes and Camila Cabello. Three item types were included to assess different dimensions of listening comprehension. Gap-filling tasks (8 items) required students to listen and supply missing words or short phrases within printed lyric extracts, targeting phonological decoding and word recognition at the acoustic level. Multiple-choice questions (7 items) asked students to select the correct answer from four options based on their understanding of the song's meaning, themes, or vocabulary, targeting inferential and global comprehension. Short-answer questions (5 items) required written responses in which students interpreted, explained, or paraphrased specific lyrical content, targeting deeper semantic processing and the ability to express comprehended meaning.

Each item carried a maximum of five points, yielding a total possible score of 100. Gap-filling and multiple-choice items were scored dichotomously as correct or incorrect. Short-answer responses were evaluated using a three-point analytic scale: zero points for no attempt or a completely incorrect response, one to two points for a partially correct response, and the full five points for a complete and accurate response. The test was administered under timed conditions, with selected song segments played twice during the assessment.

The singing performance rubric was adapted from established frameworks for listening and oral production assessment in EFL contexts. Singing performance was included as a productive measure of receptive processing on the premise that students who have internalized the song input at a sufficiently deep level will be better equipped to reproduce it with accurate phonological form and expressive quality. The rubric assessed four dimensions: pronunciation accuracy, intonation, fluency, and comprehension of song content. Each dimension was rated on a four-point scale, where 1 indicated poor performance with the criterion unmet, 2 indicated a developing level with notable deficiencies, 3 indicated proficient performance with only minor errors, and 4 indicated excellent and consistent mastery of the criterion. The four dimensions scores were summed and converted to a score out of 100 using the formula: Final Score = (Total Raw Score ÷ 16) × 100.

Content validity was established through expert validation. Both instruments were submitted to two English education lecturers at UIN Sumatera Utara with expertise in language assessment and EFL pedagogy. Each validator independently reviewed the instruments against four criteria: alignment with the target learning objectives and the construct of academic listening; appropriateness of item types and rubric dimensions for the participant population; clarity and linguistic accuracy of instructions and item wording; and adequacy of the scoring system and rubric descriptors. Written feedback was collected from both validators, and revisions to item wording, rubric descriptor language, and scoring procedures were incorporated before the instruments were finalized for data collection.

Reliability of the lyric comprehension test was estimated using Cronbach's Alpha, which measures internal consistency across the item pool. A coefficient of $\alpha \geq 0.70$ was adopted as the minimum threshold for acceptable reliability, consistent with established conventions in educational measurement. Items falling below the item-total correlation threshold were flagged for revision prior to administration. Reliability of the singing performance rubric was assessed through inter-rater procedures. Two trained raters independently scored a subset of student performances following a standardized briefing on the rubric descriptors. Agreement between raters was quantified using Cohen's Kappa (κ), with values above 0.70 interpreted as indicating substantial agreement and values above 0.80 as reflecting excellent reliability. Discrepancies between rater scores were resolved through discussion and consensus before final scores were recorded.

Prior to the treatment, all 26 participants completed a pre-test measuring their baseline academic listening ability. To prevent practice effects, the pre-test was constructed from a separate set of gap-filling and comprehension items based on a listening passage academic passage, entirely distinct from the song materials used during instruction. The mean pre-test score of 74.6 placed the group in the Developing category, confirming the presence of a listening skill gap that the instructional intervention was intended to address.

The treatment was delivered across four classroom sessions, each approximately 90 minutes in duration. Instructional activities were structured around Shape of You and Señorita using a pre-listening, while-listening, and post-listening framework. During the pre-listening phase, the instructor introduced key vocabulary and contextual background relevant to each song. During the while-listening phase, students completed guided listening tasks including gap-filling exercises and comprehension questions while the songs were played. During the post-listening phase, students engaged in singing performance activities and group discussion focused on lyrical meaning and language features. Following the final session, students completed the post-test, which employed the same instruments as those used to assess post-treatment performance. The post-test mean score of 94.2 was classified as Excellent, reflecting a gain of 19.6 points from the pre-test baseline.

Data were analyzed through two complementary procedures: descriptive statistical analysis and inferential statistical testing. In the descriptive stage, pre-test and post-test scores were summarized using measures of central tendency and score distribution. Students were classified into four performance categories based on their total scores: Excellent (≥ 90), Proficient (75–89), Developing (60–74), and Beginning (< 60). Frequency counts and percentage calculations were applied at both measurement points to describe shifts in the distribution of student performance across the treatment period.

In the inferential stage, a paired-samples t-test was conducted to determine whether the difference between pre-test and post-test means was statistically significant. This test is appropriate when the same participants are measured under two conditions, as it accounts for within-subject variance and yields a more sensitive comparison than independent-samples alternatives. The null hypothesis stated that there was no significant difference between the two means (H_0: $\mu_{pre} = \mu_{post}$), while the alternative hypothesis predicted a significantly higher post-test mean (H_1: $\mu_{post} > \mu_{pre}$). Prior to running the t-test, the assumption of normality was examined using the Shapiro-Wilk test, which is suitable for sample sizes below 50. The significance threshold was set at $\alpha = 0.05$. In addition, Cohen's d was calculated to measure effect size, with values of $d \geq 0.8$ interpreted as large. All computations were performed using SPSS version 26.

The combination of descriptive and inferential analysis ensured that findings were grounded in statistically validated evidence rather than limited to the reporting of score averages and improvement percentages, which would be insufficient for establishing the effectiveness of an instructional intervention.

Findings and Discussion

The results of this study reveal a substantial improvement in students' academic listening skills following four sessions of song-based instruction. Before the intervention was implemented, the class mean score stood at 74.6, placing the group collectively in the Developing category. After completing the treatment, the mean post-test score climbed to 94.2 a gain of 19.6 points which corresponds to the Excellent category. A paired-samples t-test confirmed that this difference was statistically significant at the $\alpha = 0.05$ level, and the effect size calculated using Cohen's d fell within the large range ($d \geq 0.8$), indicating that the improvement was not a statistical artifact but a genuine reflection of meaningful learning. These figures establish a clear before-and-after picture: students entered the intervention with moderate listening ability and left it performing at a considerably higher level.

To understand the full scope of this change, it is useful to look not only at the mean scores but also at how individual students were distributed across performance categories before and after the treatment. The table below presents this comparison.

Table 1. Comparison of Pre-Test and Post-Test Score Distribution

Category	Score Range	Pre-Test (n)	Pre-Test (%)	Post-Test (n)	Post-Test (%)
Excellent	≥ 90 |	2	8%	23	89%
Proficient	75–89	14	54%	3	11%
Developing	60-74	8	31%	0	0%
Beginning	< 60	2	8%	0	0%
Total		**26**	**100%**	**26**	**100%**

The contrast between the two measurement points is striking. Prior to the intervention, only 2 students (8%) were performing at the Excellent level, while the majority 14 students sat in the Proficient range, and 10 students were distributed across the Developing and Beginning categories. After the four sessions, 23 students (89%) had reached the Excellent category, and no students remained in either the Developing or Beginning bands. This kind of category-level migration reflects progress that was broadly distributed across the class rather than concentrated among a handful of high-performing individuals. Even students who began from weaker starting points were able to move upward, which suggests the instructional approach was accessible across different ability levels.

Examining performance at the sub-skill level provides additional texture to these results. The sharpest gains were recorded on gap-filling tasks, which required students to listen and supply missing words from played song segments. This is consistent with what listening researchers describe as bottom-up processing: the repeated exposure to song lyrics helped students sharpen their phonological decoding, making it easier to identify individual words within a continuous stream of natural speech. Many EFL learners struggle with this because features of connected speech contractions, reduced vowels, and blurred word boundaries make it difficult to isolate meaningful units. Songs, with their rhythmic regularity and recurring lyrical structures, appear to make this decoding process more manageable. Multiple-choice comprehension items also showed strong improvement, reflecting gains in students' ability to grasp meaning, theme, and vocabulary in context. Short-answer questions, which required not only comprehension but also the ability to express interpreted meaning in writing, showed more modest gains an expected pattern given that these tasks draw on both receptive and productive competencies simultaneously.

These findings align with several theoretical frameworks that help explain why songs are effective as instructional tools. The most directly relevant is Krashen's Affective Filter Hypothesis, which holds that second language acquisition is facilitated when learners' anxiety is low and their motivation is high. When learners feel anxious or under pressure, their capacity to process and retain linguistic input diminishes even when the input itself is comprehensible. During the song-based sessions in this study, the classroom environment was observably more relaxed than during conventional academic listening tasks. Students who had previously shown reluctance to engage with listening exercises appeared more willing to participate, attempt gap-filling items, and join post-listening discussions when the material was music-based. Songs carry social and emotional associations that reduce the perceived stakes of performance, and this lower affective burden appears to have created conditions more conducive to learning.

A second relevant framework is Mayer's Cognitive Theory of Multimedia Learning, which argues that learners construct understanding more effectively when information is delivered through multiple sensory channels simultaneously. In this study, students were exposed to auditory input through the songs themselves, visual input through printed lyrics, and kinesthetic-expressive engagement through the singing performance activities in the post-listening phase. This multimodal structure reduced the cognitive overload that often characterizes academic listening instruction, where learners must process dense, fast-paced spoken input with minimal scaffolding. By distributing the cognitive load across different modalities and activity types, the instructional design gave students multiple entry points for engaging with the same linguistic material which is likely to have reinforced comprehension more thoroughly than single-channel listening practice would have done.

The authentic materials framework offers a third theoretical lens. Unlike scripted classroom audio, which is typically simplified and stripped of the natural features of spoken English, songs represent genuine language produced by and for native speakers. The two songs used in this study Shape of You by Ed Sheeran and Señorita by Shawn Mendes and Camila Cabello exposed students to natural stress patterns, informal contractions, varied intonation contours, and real phonological reduction. Engaging with this kind of input, even in a structured instructional context, familiarizes learners with the very features of spoken English that most often cause difficulty in authentic academic listening situations. Students who develop comfort with natural speech features through song-based activities are likely to find genuine academic input lectures, presentations, peer discussions somewhat less opaque as a result.

Beyond the comprehension test results, the singing performance assessment provided a complementary window into the depth of students' learning. The rationale for including a productive measure was that genuine internalization of listening input should manifest in learners' ability to reproduce that input with phonological accuracy, appropriate intonation, and expressive fluency. Students who have only a surface-level familiarity with the material will typically struggle to reproduce it convincingly. In this study, post-test singing performance scores were generally strong, with content understanding receiving the highest ratings among the four rubric dimensions, followed by pronunciation accuracy.

Fluency showed the most room for growth, which is expected performing in a second language with natural rhythm and pace is a demanding skill that develops gradually over extended practice. Importantly, the inter-rater reliability for the singing rubric reached Cohen's Kappa $\kappa = 0.82$, placing it in the excellent agreement range and confirming that the scores reflect real performance differences rather than inconsistencies in how raters applied the rubric.

The findings of this study are broadly consistent with prior research on song-based language instruction while also extending that body of work in a meaningful direction. Studies such as those by Amelia et al. (2024), Berlia and Rahmawanti (2022), and Ariani and Iswandi (2021) have demonstrated positive effects of songs on listening comprehension and vocabulary development, though these have predominantly been conducted at the secondary school level. The present study contributes by moving the focus to higher education, where the demands of academic listening are considerably more complex in terms of vocabulary load, discourse structure, and cognitive processing requirements. The results suggest that the benefits of song-based

instruction do not diminish as learners advance but remain relevant and effective even when listening tasks become more cognitively demanding.

From a pedagogical standpoint, these findings support the purposeful integration of English songs into academic listening instruction at the university level. The key word here is purposeful: the results of this study emerged from a structured activity sequence pre-listening vocabulary preparation, while-listening guided tasks, and post-listening performance activities rather than from casual or unstructured song exposure. Teachers who wish to replicate these outcomes should give careful thought to song selection, choosing material that exposes students to natural phonological features while remaining accessible to their current proficiency level. They should also design activity sequences that engage students at multiple cognitive levels, from word-level decoding through to meaning interpretation and expressive reproduction. Finally, the affective dimension of song-based learning deserves deliberate attention: creating a classroom environment where students feel comfortable engaging with authentic input is not separate from developing their listening skills it is a precondition for it.

Conclusion

This study investigated the effectiveness of English songs as an instructional medium for improving academic listening skills among 26 undergraduate EFL students at UIN Sumatera Utara. The evidence collected across both the lyric comprehension test and the singing performance rubric points to a consistent and meaningful conclusion: song-based instruction produced substantial gains in students' ability to process, understand, and reproduce spoken English at the academic level.

The pre-test to post-test improvement from a class mean of 74.6 to 94.2 was confirmed as statistically significant through paired-samples t-test analysis, with a large effect size (Cohen's $d \geq 0.8$) indicating that the gains were practically meaningful rather than merely detectable. The distribution data reinforced this finding: 89% of students reached the Excellent category after the intervention, compared to just 8% before it, and no students remained in the Developing or Beginning bands by the end of the treatment period. These patterns suggest that the intervention was effective not just on average but across a wide range of starting ability levels.

The theoretical frameworks invoked throughout the discussion help explain why these results emerged. Krashen's Affective Filter Hypothesis accounts for the role of the lowered-anxiety classroom environment in enabling more effective language processing. Mayer's Cognitive Theory of Multimedia Learning explains how the multimodal structure of the activity sequence combine auditory, visual, and kinesthetic engagement supported deeper comprehension by distributing cognitive load across channels. The authentic materials framework clarifies why exposure to natural phonological features in real songs helped students develop skills that are more transferable to genuine academic listening contexts than simplified classroom audio typically allows. Taken together, these frameworks suggest that the effectiveness of song-based instruction is not incidental it is grounded in well-established principles of how second language acquisition and comprehension work.

The inclusion of singing performance as a productive assessment measure adds a layer of evidence that purely receptive test scores cannot provide. The strong performance on the content understanding and pronunciation accuracy dimensions of the rubric, combined with high inter-rater reliability ($\kappa = 0.82$), indicates that students' comprehension gains were substantive they had genuinely internalized the linguistic material rather than developing only surface familiarity with it. The more modest gains in fluency point to an area where extended practice beyond four sessions would likely produce further development.

In terms of pedagogical implications, this study recommends that English instructors at the university level approach song-based listening activities as structured teaching tools rather than supplementary entertainment. Deliberate song selection, systematic pre-listening and while-listening scaffolding, and the inclusion of productive post-listening tasks are all factors that appear to contribute to the effectiveness of this approach. Equally important is the attention given to the emotional climate of the classroom: the affective benefits of music-based instruction are not peripheral to learning outcomes but are closely tied to them.

This study has limitations that should be acknowledged honestly. The sample was drawn from a single classroom without a control group, the treatment consisted of only four sessions, and the findings cannot be straightforwardly generalized to other institutional or linguistic contexts. Future research should address these constraints by using larger and more diverse participant groups, incorporating comparison conditions, and extending the intervention period to examine whether gains are sustained over time. Comparative studies that set song-based instruction alongside other authentic listening materials would also help clarify the particular strengths of this approach relative to alternatives. Despite these limitations, the present study offers a grounded and evidence-based contribution to the literature on academic listening instruction in Indonesian higher education a context that remains underrepresented in the broader EFL research landscape.

References

Afifah, R. (2013). The correlation between students' habit in listening to English songs and their English listening comprehension skill. *LET: Linguistics, Literature and English Teaching Journal, 4*(2). https://doi.org/10.18592/let.v4i2.1407

Amelia, M. D. N., Tasnim, Z., & Fardhani, A. E. (2024). The use of English song to improve listening comprehension achievement of junior high school students. *EFL Education Journal, 11*(2). https://doi.org/10.19184/eej.v11i2.49929

Annuri, F. M., Authar, N., Zaniar, S., Saputri, T., & Sarkawi, D. (2025). Integrating English songs as a teaching strategy: Teachers' perspectives on enhancing EFL students' listening comprehension. *English Journal of Merdeka (EnJourMe), 10*(1). https://doi.org/10.26905/enjourme.v10i1.15859

Ariani, S., & Iswandi, K. (2021). The use of English pop song to enhance students' listening ability. *Jo-ELT (Journal of English Language Teaching), 7*(2). https://doi.org/10.33394/jo-elt.v7i2.2965

Astutik, Y., & Khofifah. (2025). The use of English songs in teaching listening skills through online learning at junior high school level. *Borneo Educational Journal, 4*(1). https://doi.org/10.24903/bej.v4i1.950

Berlia, I. S., & Rahmawanti, M. R. (2022). Benefiting EFL students in listening comprehension with English songs. *English Journal of Indragiri, 6*(1), 72–85. https://doi.org/10.32520/eji.v6i1.1691

Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy (2nd ed.).* Pearson Education.

Khoirunnisa, F. H. S., Rizqia Nada, A., & Fatmasari, N. (2025). The influence of English songs on listening comprehension of the English department students at university. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 12*(1). https://doi.org/10.24256/ideas.v12i1.3333

Nation, I. S. P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking.* Routledge.

Pangaribuan, D., & Chairunnissa. (2023). Students' perception on song as media in learning EFL. *Journal of Classroom Action Research, 2*(2), 8–16. https://doi.org/10.52622/jcar.v2i2.158158

Putri, B. G., Danayanti Degeng, P. D., & Isnaini, M. H. (2022). EFL students' perception towards the use of English songs as listening learning media. *KLAUSA, 6*(2). https://doi.org/10.33479/klausa.v6i2.625

Ramadhanti, V. A., & Farida, A. N. (2025). The impact of listening to English songs on EFL students' vocabulary acquisition: A literature review. *ELT Forum: Journal of English Language Teaching, 14*(Special Issue). https://doi.org/10.15294/elt.v14iSpecial%20Issue.29169

Rost, M. (2011). *Teaching and researching listening (2nd ed.).* Routledge.

Samad, P., & Paris, N. (2025). The EFL learners' listening comprehension through Western song and the strategies for improvement. *DIAJAR: Jurnal Pendidikan dan Pembelajaran, 1*(2). https://doi.org/10.54259/diajar.v1i2.718

Tadris, Y. R., & Saharani, A. A. (2024). The use of song as learning media in teaching listening to EFL students. *Indonesian Journal of Multidisciplinary Educational Research, 1*(1). https://doi.org/10.30762/ijomer.v1i1.1123