

Islamic English Pop Songs in EFL Instruction: A Mixed-Method Study of Sixth-Grade Students' Listening and Speaking Development at an Islamic Elementary School

Mohamad Irfan Sahroni
MI Alkhoiriyah 02 Semarang, Indonesia
sahroniirfan666@students.unnes.ac.id

Seful Bahri
Universitas Negeri Semarang, Indonesia
seful.bahri@mail.unnes.ac.id

Abstract

This study examined the effectiveness of Islamic English pop songs as a learning medium for developing sixth-grade students' listening and speaking skills at an Islamic elementary school and explored students' perceptions of this medium. A sequential explanatory mixed-methods design was employed involving 42 sixth-grade students, divided into an experimental group (n=30) taught using Maher Zain's songs "One Big Family" and "Hold My Hand," and a control group (n=12) receiving conventional textbook-based instruction, across six treatment sessions. Quantitative data from pre-tests and post-tests were analyzed using the Wilcoxon Signed-Rank Test and Mann-Whitney U Test, while qualitative data from classroom observations and semi-structured interviews were analyzed thematically. Within-group analysis showed significant improvements in both listening and speaking skills for the experimental group. However, between-group comparison indicated a significant difference in listening post-test scores favoring the control group, while no significant difference was found in speaking post-test scores between groups. Qualitatively, students responded positively to song-based learning, reporting increased motivation, enjoyment, and appreciation of the moral values embedded in the songs. These findings suggest that Islamic English pop songs hold pedagogical value as an engaging and culturally appropriate complementary medium, particularly for fostering motivation and speaking practice, although their comparative advantage over conventional instruction for listening skills remains inconclusive.

Keywords: English pop songs; Islamic elementary school; listening skills; mixed-method; speaking skills

INTRODUCTION

Songs are considered one type of authentic language input which contains various vocabulary, natural pronunciation patterns and rhythmic structures that provide students with authentic listening input, and which also can be used to help students in their speaking production and listening comprehension (Lengari, 2024). The song media is an alternative to teaching students in the elementary level because the use of songs can be used in teaching that is entertaining,

interactive, and fun, where young learners are able to pay attention to the teaching. The old method can not retain students' attention, while young students are naturally drawn to dynamic, interactive, and enjoyable teaching input, so that song media is appropriate for teaching elementary students (Mawarni & Hamdani, 2024; Ndayisenga & Laos Mbato, 2024). The need is particularly urgent for elementary school students in Indonesia, because students are still often struggling with the development of oral communication

skills and learning English is still passive because they have limited exposure to authentic input (Andreanto, 2021).

Based on this need, previous studies suggest the pedagogical significance of songs in both speaking and listening. From the speaking point of view, Hilmy et al. (2024) could be seen as having an effect of the use of English song media on the accuracy and fluency of the participants' pronunciation, while Kitjaroonchai and Sukman (2025) reported that the song integrated speaking intervention for fifteen weeks resulted in significant improvement in accuracy and fluency. For the listening part, Aini et al. (2025) concluded that there was a high correlation between the song engagement with the retained memory of junior secondary students, $r=0.82$, and Trisnawati et al. (2023) showed that the post-test listening comprehension of a song taught group was significantly higher than the control group. These effects are theoretically related to Krashen's Input Hypothesis and Schmidt's Noticing Hypothesis, the latter stating that the melodic and rhythmic salience of songs decreases the affective filter to the comprehensible input and increases the learners' noticing of phonological features like stress and connected speech (Mulyani et al., 2026).

In addition to these general benefits, the relevance of English pop songs to listening development extends beyond motivation alone. Songs provide repeated, contextualized exposure to pronunciation, intonation, and stress patterns that reflect how native speakers produce connected speech. Consequently, when learners actively engage with lyrics by singing, repeating, and performing, they practice oral output within a comparatively low anxiety environment, which is considered particularly suitable for young learners (Hilmy et al., 2024). This claim is reinforced by Simanjuntak et al.'s (2025) experimental finding that learners

taught pronunciation through song-based tasks, such as gap filling and connected speech identification, showed greater gains in both pronunciation accuracy and listening comprehension than learners who practiced only through conventional, isolated pronunciation drills. They attributed this advantage to songs' steady rhythmic structure and predictable stress patterns, which help learners internalize the stress-timed rhythm of English, a feature that often feels unfamiliar to learners from syllable-timed language backgrounds such as Indonesian.

Moreover, finding a method that is engaging and suitable for Islamic perspective in elementary schools is more urgent, because Islamic elementary schools are schools that emphasize the combination of moral education and intellectual education (Halisnawati et al., 2024). In this regard, songs by artists like Maher Zain whose lyrics are totally in line with Islamic values can provide opportunities for language instruction and character education. Previous research has validated that the use of carefully selected English Islam related songs can stimulate learners' interest and become vehicles for language learning in a culturally appropriate manner (Hilmy et al., 2024; Nurdin et al., 2023).

In theory, the role of learning media in the classroom is a bridge that carries the teaching content from the teacher to the student, the function of learning media is in the classroom has three functions, namely attention, affective, and compensatory. Theoretically, learning media as a mediating bridge that carries the teaching content from the teacher to the student, learning media has three functions in the classroom, namely attention function, affective function, and compensatory function. In this context, songs are a special type of language-learning media that has been explained using two complementary hypotheses on its impact on language

acquisition. According to one view, Krashen's Input Hypothesis suggests that songs offer comprehensible and repeated exposure to authentic phonological patterns in a low anxiety context. However, according to Schmidt's Noticing Hypothesis, the melodic and rhythmic salience of songs makes learners notice the phonological aspects of the songs, which are seen as a prerequisite for song acquisition to take place (Mulyani et al., 2026). In the light of this structure, two hypotheses were tested with this study. The first hypothesis (Ha1) suggested that the use of English pop songs as a learning medium would lead to students' speaking skills improving significantly, because English pop songs give natural pronunciation models and rhythms for students to learn, and the pronunciation is repeated orally, which helps students to speak fluently. The second hypothesis (Ha2) assumed that the inclusion of English pop songs would lead to a significant improvement in students' listening skills because listening to songs would help them improve the pronunciation, intonation and rhythm of English.

Of these previous studies, four are particularly relevant with complete sources documented. Hilmy et al. (2024) used a quasi-experimental research design with quantitative only data and without an Islamic context, involving 72 students in grade 10 of SMAN 2 Situbondo, to determine the effect of pop song media on students' speaking ability, which concluded that the use of pop song media had a significant effect on the increase in students' speaking ability with significance value of .000. Similarly, Mawarni and Hamdani (2024) found that songs facilitated vocabulary and grammar retention, as well as participation in learning, through qualitative observation and interviews of the young learners, but their study was not conducted in an Islamic elementary school context; hence, they did

not quantitatively measure the impact of songs. Aini et al. (2025) identified that there is a strong quantitative correlation between song engagement and memory retention between the 42 seventh graders of Islamic junior secondary school (JSS1) with r value of .82, but this research was on Islamic Religious Education content not English listening but not having a measure of speaking. Meanwhile, Kamil (2024) revealed that there was a significant correlation between listening to pop songs and mastery of vocabulary skills of Islamic junior secondary students, and was conducted by only using one method design which is vocabulary.

Despite its potential, four gaps are repeated in the broader literature. First, most work on songs as a tool for instruction has been done at the secondary or university level and has not been much focused on contexts at the elementary level. Second, the studies so far have focused on studying speaking, listening, and or vocabulary separately from each other rather than studying the simultaneous development of speaking and listening. Third, most studies use a purely quantitative or purely qualitative research design and not a mixed method research design that can explain both outcomes and learner experiences. Fourthly, there is a lack of research that focuses on Islamic English pop songs as media for developing speaking and listening in Islamic elementary school which requires screening for religious and cultural appropriateness in song content. A preliminary observation at the research site confirmed these gaps: students in grade six were found to be not very enthusiastic in speaking English and were found to find difficulties in following simple English instructions, while the teacher indicated that the instruction based on the use of an English textbook, which was aligned with the curriculum, was not being able to keep the students engaged in the learning process,

because of the repetitive read translate answer pattern in the English textbook.

Based on these gaps, three research questions were developed in this study. The first question was about the effectiveness of Islamic English Pop Songs as a learning media to improve the speaking ability of SMT students in grade 6. The next question was how effective Islamic English pop songs are as a learning medium in improving listening ability in the sixth-grade students. The third question was about students' opinions regarding the Islamic English pop songs as a medium in English learning.

METHODOLOGY

This research used a sequential explanatory mixed-methods design (Creswell & Plano Clark, 2018), with the quantitative phase being quasi-experimental to provide the foundation for the next phase of the research, which was qualitative, to explain and contextually understand the quantitative findings. Quantitative was conducted on two intact sixth grade classes at the Islamic Elementary School in Semarang as the quantitative experimental group and control group. Thirty students in an experimental group were taught six sessions of instruction that included themes of "One Big Family" (Sessions 1-3) and "Hold My Hand" (Sessions 4-6) and 12 students in a control group were taught six sessions of instruction based on Maher Zain's texts of "One Big Family" (Sessions 1-3) and "Hold My Hand" (Sessions 4-6). For listening and speaking, both groups received the same pre-test and post-test measures before and after the intervention, which was six sessions, two sessions weekly for three weeks, limited by the thematic unit provided in the school's current English curriculum schedule. The entire study period lasted for 8-weeks – baseline data was collected in Week 1, intervention was conducted in Weeks 2-4, post-test was conducted in Week 5 and

qualitative data collection was carried out in Weeks 6-8. The different group sizes were chosen due to the different natural make-up of the existing two intact classes at the school, not because of a planned allocation ratio. As there were two classes, these were assigned intact to the experimental and control conditions, as this method would have led to breaking up classes and disruption of the regular school timetable if random assignment to condition had been used. This unequal-sized random assignment is recognized as a weakness in the design and discussed further in the Limitations section.

The session structure for each of the experimental groups was the same and consisted of five stages. The warm-up discussion took 5 minutes and was used to tap into background knowledge of the theme of the song, with students following the written lyrics on the first listening stage, which lasted for 10 minutes. Then the teacher led the students through a 15-minute speaking and expression focus period where they were given examples of phrases to be used and practised in pairs, followed by a gap fill listening exercise in the second listening period, lasting 10 minutes. This session ended with a 10-minute speaking task where students explained or paraphrased the message of the song. The control group was exposed to a textbook text and parallel five-stage structure, with the corresponding texts in song replaced by textbook audio and text at the respective stages.

The subject of the study was 42 students in the sixth grade with the age range of 11-12 years, which is in the Fase C level of the Kurikulum Merdeka. There were five inclusion criteria for participants in this study. First, they had to attend the teaching school at the research site, and they had to be in the sixth grade. Second, all participants should have had a minimum of 2 years of English prior to the study. Thirdly, participants could not have received

systematic song-based English teaching in the past. Fourth, parents/Guardians had to give permission to participate. Fifth, participants had to be 80% or more present in the study period. The six students in the experimental group with a wide range of pre-test performance levels, and their English teacher were selected to represent the various performance levels for the purpose of the qualitative phase of the study.

Two songs were used as the basis for the 20 MC items used to assess listening; this resulted in a Cronbach's alpha of 0.874 based on a pilot study of 30 students, and a structured oral task measuring pronunciation accuracy, fluency and comprehensibility, adapted from the task in Brown (2004) was used to assess speaking resulting in a Cronbach's alpha of 0.768. The full speaking assessment rubric and the listening test blueprint are included in Appendix. All the speaking post tests were independently scored by two scorers, the classroom English teacher and the researcher, who used the same rubric, to prevent subjectivity in scoring. Then interrater reliability was verified by Pearson correlation and a final score was calculated as the average of the two scores. In the pilot study, the analysis of item difficulty of the 20 listening items revealed that a few were in an easy range for the pilot participants and this demonstrated the ceiling effect that later occurred on the control post-test with a mean score of 100.00 and a standard deviation of 0.000, which is discussed below. Classroom observations consisted of a nine-aspect rating sheet with the following rating scales: 1 (low) to 4 (high); administered in all six experimental sessions; and, semi-structured interviews were used to gain insights into students' and the teacher's experience of the intervention.

Data analysis was done using SPSS version 26 (Quantitative). Several score distributions showed deviations from normality as indicated by Shapiro-Wilk tests, which led to

the use of non-parametric alternatives throughout the hypothesis testing portion. A Mann-Whitney U Test was completed to ensure that the experimental and control groups were not different on the pre-test scores before examining post-test differences; the results of this baseline check are reported at the beginning of the Findings section. For within-group pre-test/post-test comparisons, Wilcoxon Signed-Rank Test was used and effect size was calculated as.

For between group post test comparisons, the Mann-Whitney U Test was used. The qualitative analysis of the data centered on thematic analysis based on the six phases described by Braun and Clarke (2006) and triangulation of observation notes and both types of interviews, member checks, and thick description was used to establish trustworthiness. The quantitative and qualitative results were merged during the interpretation, where the qualitative results were used to interpret and place the statistical results into context.

Both instruments have been judged as content valid by experts. The English teacher of the class and an English Education lecturer reviewed the listening test items and the speaking rubric before the piloting, and items and/or rubric descriptors that were found ambiguous or that were inconsistent with the content of the songs were revised. Construct validity was established by correlating listening items and speaking rubric descriptors directly with the words, phrases, and concepts used in the two chosen songs used as the input to the students during treatment, and not with their general ability of listening or speaking.

Songs for the study were screened against five criteria specified in the thesis. First, songs needed to align with the Kurikulum Merdeka capaian pembelajaran for Fase C. Second, songs needed thematic relevance to students' everyday life. Third, songs needed to be compatible with Islamic values,

including appropriate content and moral messaging. Fourth, songs needed linguistic accessibility at a level appropriate for sixth-grade learners. Fifth, songs needed a suitable musical tempo and clarity of pronunciation for listening tasks. "One Big Family" and "Hold My Hand" by Maher Zain were judged to meet all five criteria.

Ethical procedures included obtaining written consent from parents or guardians prior to data collection, securing permission from the school's principal and the class's English teacher, and informing student participants in age-appropriate language that participation was voluntary and would not affect their course grades. All data were anonymized prior to analysis, and the pseudonymous student labels used in the qualitative findings below, ranging from Student A through Student F, do not correspond to the students' actual names or the order in which they were interviewed.

RESULT AND DISCUSSION

Effectiveness for Listening and Speaking

Skills

Prior to examining post-test differences, a Mann-Whitney U Test was conducted to verify baseline equivalence between the experimental and control groups on the pre-test scores. For listening, no significant difference was found between the experimental group (Mean Rank = 19.52) and the control group (Mean Rank = 26.46), $U = 120.5$, $p = .090$. Similarly, for speaking, no significant difference was found between the experimental group (Mean Rank = 19.57)

and the control group (Mean Rank = 26.33), $U = 122.0$, $p = .100$. These results indicate that the two groups did not differ significantly at baseline, providing a reasonable basis for interpreting the subsequent post-test comparisons as reflecting the effect of instructional treatment rather than pre-existing group differences.

The descriptive statistics presented and show the listening pre-test and post-test scores for both the experimental and control groups. In the experimental group, the mean listening score increased from 87.17 in the pre-test to 96.17 in the post-test, and the standard deviation decreased from 6.654 to 2.520.

This indicates that students' average listening ability improved following the treatment using Islamic English pop songs, and the scores also became more homogeneous among students. In the control group, the mean listening score increased from 90.83 in the pre-test to 100.00 in the post-test, and the standard deviation decreased to .000, meaning that all students in this group achieved a perfect score in the post-test. The increase observed in the control group was slightly higher than that of the experimental group, with a mean gain of 9.17 points compared to 9.00 points in the experimental group. These results suggest that both instructional approaches were associated with an improvement in students' listening performance from pre-test to post-test. However, the fact that the control group reached a ceiling score of 100 for all students raises the possibility of a ceiling effect, which is discussed further below.

Table 4. 1 Descriptive Statistics of Students' Listening Pre-Test and Post-Test Scores

Variable	N	Minimum	Maximum	Mean	Std. Deviation
<i>Group = Experimental</i>					
Listening_Pre	30	75	95	87.17	6.654
Listening_Post	30	90	100	96.17	2.520
<i>Group = Control</i>					

Listening_Pre	12	80	100	90.83	6.686
Listening_Post	12	100	100	100.00	.000

Descriptive statistics were also calculated for students' speaking scores in both the experimental and control groups. In the experimental group, the mean speaking score increased from 62.2227 in the pre-test to 86.9450 in the post-test, and the standard deviation decreased slightly from 10.88370 to 8.30227. This indicates a substantial improvement in students' speaking ability following the treatment using Islamic English pop songs. In the control group, the mean speaking score increased from 68.7500 in the pre-test to 92.3617 in the post-test, and the standard deviation decreased from 10.12809 to 8.30227.

Table 4. 2 Descriptive Statistics of Students' Speaking Pre-Test and Post-Test Scores

Variable	N	Minimum	Maximum	Mean	Std. Deviation
Group = Experimental					
Pre_FinalScore	30	41.67	83.33	62.2227	10.88370
Post_FinalScore	30	66.67	100.00	86.9450	10.42315
Group = Control					
Pre_FinalScore	12	50.00	83.33	68.7500	10.12809
Post_FinalScore	12	75.00	100.00	92.3617	8.30227

This also indicates a notable improvement in speaking ability among students who received conventional textbook-based instruction. When comparing the gain scores, the experimental group showed an increase of 24.72 points, while the control group showed an increase of 23.62 points. Unlike the listening results, the gain in speaking scores between the two groups was relatively similar, with the experimental group showing a slightly larger improvement. Whether this difference is statistically significant is examined through the subsequent normality and hypothesis testing.

The normality of the listening pre-test and post-test data was examined using the Shapiro-Wilk test, because the sample size was relatively small. The results indicated that the listening scores for both the experimental and control groups were not normally distributed. In the experimental group, the significance values were 0.001 for the pre-test and 0.000

for the post-test, both below the 0.05 threshold. In the control group, the pre-test significance value was 0.028, also below 0.05, while the post-test data could not be computed due to zero variance, because all students obtained the same score of 100. Since the significance values for all variables were below 0.05, the assumption of normality was violated. Therefore, in accordance with the data analysis procedure outlined in the Methodology, non-parametric alternatives were employed for the subsequent hypothesis testing. The Wilcoxon Signed-Rank Test was used in place of the paired sample t-test to examine the difference between pre-test and post-test scores within each group, and the Mann-Whitney U Test was used in place of the independent sample t-test to compare post-test scores between the experimental and control groups.

Normality testing for the speaking scores was also conducted using the Shapiro-Wilk test. In the experimental group, the pre-test score showed a significance value of 0.062,

indicating a normal distribution, whereas the post-test score showed a significance value of 0.007, indicating a non-normal distribution. In the control group, the pre-test score showed a significance value of 0.495, indicating a normal distribution, while the post-test score showed a significance value of 0.027, indicating a non-normal distribution. Since the post-test scores in both groups did not meet the normality assumption, non-parametric tests were employed for the subsequent hypothesis testing on the speaking data, consistent with the approach used for the listening data.

The Wilcoxon Signed-Rank Test was conducted to examine whether there was a significant difference between the pre-test and post-test listening scores within each group. In the experimental group, the results showed a Z value of -4.522 with an Asymp. Sig. (2-tailed) value of 0.000, which is below the 0.05 significance level. Of the 30 students, 26 showed an increase in their listening scores

from pre-test to post-test; no students showed a decrease, and 4 students showed no change. This indicates that the increase in listening scores following the treatment using Islamic English pop songs was statistically significant. In the control group, the results showed a Z value of -2.989 with an Asymp. Sig. (2-tailed) value of 0.003, also below the 0.05 significance level. This indicates that the increase in listening scores from pre-test to post-test in the control group, which received conventional textbook-based instruction, was also statistically significant. These findings suggest that both instructional methods were associated with a significant improvement in students' listening skills. Because both groups showed significant improvement, the Mann-Whitney U Test was subsequently conducted to determine whether the magnitude of improvement differed significantly between the experimental and control groups.

Table 4. 3 Results of Wilcoxon Signed-Rank Test for Listening Pre-Test and Post-Test Scores

Listening_Post – Listening_Pre	
<i>Group = Experimental</i>	
Z	-4.522
Asymp. Sig. (2-tailed)	.000
<i>Group = Control</i>	
Z	-2.989
Asymp. Sig. (2-tailed)	.003

The Wilcoxon Signed-Rank Test was conducted to examine whether there was a significant difference between the pre-test and post-test speaking scores within each group. In the experimental group, the results showed a Z value of -5.396 with an Asymp. Sig. (2-tailed) value of 0.000, which is below the 0.05 significance level. This indicates that the increase in speaking scores following the treatment using Islamic English pop songs was statistically significant. In the control group, the results showed a Z value of -3.276 with an

Asymp. Sig. (2-tailed) value of 0.001, also below the 0.05 significance level. This indicates that the increase in speaking scores from pre-test to post-test in the control group was also statistically significant. These findings indicate that both instructional methods were associated with a significant improvement in students' speaking skills. As with the listening results, the Mann-Whitney U Test was subsequently conducted to determine whether the magnitude of improvement in speaking scores differed significantly between the experimental and control groups.

Table 4. 4 Results of Wilcoxon Signed-Rank Test for Speaking Pre-Test and Post-Test Scores

Post_FinalScore – Pre_FinalScore	
Group = Experimental	
Z	-5.396
Asymp. Sig. (2-tailed)	.000
Group = Control	
Z	-3.276
Asymp. Sig. (2-tailed)	.001

To provide a more complete picture of the treatment effect beyond statistical significance, effect size was calculated for the Wilcoxon Signed-Rank Test using the formula $r = Z/\sqrt{N}$, where Z is the standardized test statistic and N is the total number of observations. For the experimental group's listening skills, $r = -4.522/\sqrt{30} = -0.826$, indicating a large effect. For speaking skills, $r = -5.396/\sqrt{30} = -0.985$, also indicating a large effect. These effect size values suggest that, within the experimental

group, the use of Islamic English pop songs had a practically significant impact on both listening and speaking skill development, even though the between-group comparison did not yield statistically significant superiority over conventional instruction.

The Mann-Whitney U Test was conducted to examine whether there was a significant difference in listening post-test scores between the experimental and control groups.

Table 4. 5 Results of Mann-Whitney U Test for Listening Post-Test Scores

Ranks	N	Mean Rank	Sum of Ranks
1.00 (Experimental)	30	17.10	513.00
2.00 (Control)	12	32.50	390.00
Total	42		

Test Statistics	Listening_Post
Mann-Whitney U	48.000
Wilcoxon W	513.000
Z	-4.195
Asymp. Sig. (2-tailed)	.000
Exact Sig. [2×(1-tailed Sig.)]	.000

The results showed a Mann-Whitney U value of 48.000 with a Z value of -4.195 and an Asymp. Sig. (2-tailed) value of 0.000, which is below the 0.05 significance level, indicating a statistically significant difference between the two groups. However, the mean rank for the control group was 32.50, which was higher

than the mean rank for the experimental group at 17.10. This indicates that the control group, which received conventional textbook-based instruction, obtained higher listening post-test scores than the experimental group, which received treatment using Islamic English pop songs. This finding is contrary to the initial

hypothesis that the use of Islamic English pop songs would result in greater improvement in listening skills compared to conventional instruction. Possible explanations for this finding, including the disproportionate sample sizes between groups and the potential ceiling effect observed in the control group, are elaborated further in the discussion that follows.

Table 4. 6 Results of Mann-Whitney U Test for Speaking Post-Test Scores Between Experimental and Control Groups

Ranks	N	Mean Rank	Sum of Ranks
1.00 (Experimental)	30	19.72	591.50
2.00 (Control)	12	25.96	311.50
Total	42		

Test Statistics	Post_FinalScore
Mann-Whitney U	126.500
Wilcoxon W	591.500
Z	-1.540
Asymp. Sig. (2-tailed)	.124
Exact Sig. [2×(1-tailed Sig.)]	.138

The Mann-Whitney U Test was also conducted to examine whether there was a significant difference in speaking post-test scores between the experimental and control groups. The results showed a Mann-Whitney U value of 126.500 with a Z value of -1.540 and an Asymp. Sig. (2-tailed) value of 0.124, which is above the 0.05 significance level, indicating no statistically significant difference between the two groups. Although the mean rank for the control group was higher than that of the experimental group, at 25.96 compared to 19.72, this difference was not statistically significant. This finding indicates that, in terms of speaking ability, the use of Islamic English pop songs as a learning medium produced improvement that were not significantly different from those achieved through conventional textbook-based instruction.

To further examine the relative improvement of each group beyond raw post-test scores, a gain score analysis was conducted by calculating the difference between each student's post-test and

pre-test scores. For listening skills, the mean gain score for the experimental group was 9.00 points, compared to 9.17 points for the control group, reflecting a comparable level of improvement between the two groups despite the statistically significant between-group difference identified by the Mann-Whitney U Test, which was largely attributable to the ceiling effect in the control group rather than a genuine superiority of either instructional approach. For speaking skills, the mean gain score for the experimental group was 24.72 points, while the control group gained 23.62 points, suggesting broadly similar levels of improvement between the two groups. These gain score results further support the interpretation that song-based instruction produced learning gains comparable to those of conventional instruction within the scope of this study, and that the absence of a statistically significant between-group difference in speaking scores reflects a genuine similarity in the magnitude of

improvement rather than a failure of the song-based treatment to produce meaningful gains.

Discussion

The statistical findings presented above reveal a complex picture of how Islamic English pop songs influenced students' listening and speaking skills compared to conventional textbook-based instruction. Regarding listening skills, the Wilcoxon Signed-Rank Test demonstrated that both the experimental and control groups experienced statistically significant improvements from pre-test to post-test, with significance values of 0.000 and 0.003, respectively. This finding is consistent with Supriyani et al. (2025), who found that the use of English songs as a learning medium significantly improved students' listening comprehension, with mean scores increasing substantially from pre-test to post-test. The improvement observed in the experimental group therefore lends support to the theoretical claim that songs serve as a genuine source of language input, allowing students to internalize natural intonation, rhythm, and pronunciation patterns embedded in the lyrics. This view is further reinforced by Torras-Vila (2021), who concluded that singing and listening to songs in a foreign language boosts students' linguistic capacity through enchanting melodies, varying rhythm, and image-evoking lyrics that appeal to the multidimensional development of human intelligence.

However, the Mann-Whitney U Test produced a result that diverges considerably from the initial hypothesis. The test revealed a statistically significant difference between the two groups, but the mean rank for the control group, at 32.50, was higher than that of the experimental group, at 17.10, indicating that the control group achieved higher listening post-test scores. This outcome contradicts the assumption that Islamic English pop songs would be more effective than conventional instruction in enhancing listening ability. A

closer examination of the descriptive statistics offers a plausible explanation for this unexpected result. The control group reached a mean post-test score of 100.00 with a standard deviation of .000, meaning that every student in this group obtained a perfect score. This pattern strongly suggests the presence of a ceiling effect, a measurement limitation described by Staus et al. (2021) as occurring when participants begin or end a study with scores so high that no further improvement can be detected. Cox (2022) similarly found that when participants performed at the upper boundary of a proficiency level, assessment instruments were less able to detect meaningful differences in performance between individuals, since all participants clustered at or near the maximum score regardless of their actual underlying ability, a principle that directly applies to the present study's listening post-test. Because baseline equivalence testing confirmed no significant pre-existing difference between groups ($p = .090$), the higher post-test performance of the control group is more plausibly attributable to the ceiling effect described above than to an initial advantage in listening ability.

This finding also resonates with the conclusions drawn by Widyaningtyas and Bambang (2024), who examined the relationship between students' habits of listening to English songs and their listening proficiency among university students and reported a weak and statistically non-significant negative correlation, with a Spearman's rho value of -0.125 and a significance value of 0.571. They concluded that exposure to English songs as an informal listening activity may not directly translate into improved academic listening performance, suggesting that the relationship between song-based exposure and listening achievement is more complex than commonly assumed. The result of the present study, in which the experimental group receiving song-based treatment did not outperform the control

group on the listening post-test, appears to align with this conclusion. It is also worth noting that the control group's own improvement should not be interpreted as evidence that conventional instruction is inherently superior. Wahyudi and Masduqi (2026), in a mixed-method study exploring Spotify as an informal learning tool, reported that students perceived improvements in listening comprehension through exposure to both songs and podcasts, with podcasts offering particularly authentic input due to their natural speech rate. This suggests that authentic audio input in general, whether delivered through songs or other audio media, has the potential to enhance listening skills, and that songs do not hold a unique or exclusive advantage over other forms of input. Turning to speaking skills, the Wilcoxon Signed-Rank Test indicated that both the experimental and control groups achieved statistically significant improvements from pre-test to post-test, with significance values of 0.000 and 0.001, respectively. This finding aligns with Sinurat et al. (2025), who reported a significant increase in students' speaking scores following the implementation of English songs. The within-group improvement observed in the experimental group is therefore consistent with the broader literature confirming that song-based instruction can produce measurable gains in speaking ability, a pattern also reported by Rorintulus and Wuntu (2023) at the elementary level, where songs were found to improve students' pronunciation through the repetitive and rhythmic nature of song lyrics.

Unlike the listening results, the Mann-Whitney U Test for speaking post-test scores showed no statistically significant difference between the two groups, with a significance value of 0.124. Although the mean rank for the control group, at 25.96, was slightly higher than that of the experimental group, at 19.72, this difference did not reach statistical significance. This finding indicates that the use of Islamic

English pop songs produced speaking gains that were statistically comparable to those achieved through conventional instruction, neither clearly superior nor inferior. Pino et al. (2023) offer a possible explanation, noting that rhythm perception is a strong predictor of grammatical and phonological abilities in school-aged children, which suggests that the rhythmic structure embedded in songs such as *One Big Family* and *Hold My Hand* could plausibly support speaking-related competencies such as pronunciation and fluency. However, Pino et al. (2023) also emphasized that the strength of this music-language association varies considerably across individuals and environmental factors, which may help explain why the experimental group's numerically larger advantage did not reach statistical significance.

The absence of a significant between-group difference in speaking scores may also be understood from a statistical standpoint. Abdelrasheed et al. (2026) emphasized that non-parametric tests such as the Mann-Whitney U Test generally require larger and more balanced sample sizes to achieve sufficient statistical power to detect meaningful between-group differences. In the present study, the unequal group sizes, with 30 students in the experimental group and only 12 in the control group, may have limited the statistical power of the test for the speaking comparison, meaning that a real difference between the groups could have existed in practice without reaching the threshold for statistical significance. Abdelrasheed et al. (2026) further noted that in small sample EFL studies, effect size measures should be reported alongside significance values, which is consistent with the effect size analysis reported above, where the large effect size for speaking within the experimental group, at r equal to -0.985, confirmed that the song-based treatment had a practically significant impact on speaking development even in the absence of a statistically significant between-group difference.

This result also contrasts with some prior studies. Safitri and Najwa (2024) reported that the experimental group significantly outperformed the control group across all five speaking criteria they assessed, with a 58.4% improvement compared to 26.8% in the control group, a considerably larger gap than observed in the present study. One plausible explanation for this divergence is offered by Firmansyah et al. (2024), who found that although English songs had a positive effect on students' speaking skills, students also faced difficulties understanding song lyrics and applying vocabulary learned from songs to everyday conversation. These challenges may have limited how fully the experimental group in the present study could translate song exposure into speaking gains clearly superior to conventional instruction.

Taken as a whole, the quantitative findings suggest that Islamic English pop songs, specifically “One Big Family” and “Hold My Hand” by Maher Zain, can serve as an engaging and meaningful medium for improving both listening and speaking skills

among sixth-grade students, as evidenced by the significant within-group improvements observed for both skills. However, the comparative analysis with the control group indicates that the superiority of songs over conventional instruction cannot be firmly established based on the quantitative data alone, particularly for listening skills, where the ceiling effect appears to have masked any potential advantage of the treatment, and for speaking skills, where the difference between groups did not reach statistical significance despite a numerically larger gain in the experimental group.

Students' Perceptions of Islamic English Pop Songs as Learning Media

Classroom observations were conducted across all six treatment sessions in the experimental group, using the researcher's standardized observation sheet, reproduced in Table 4.9, which rated nine aspects of classroom behavior on a scale from 1, meaning Very Poor, to 4, meaning Excellent, with each aspect accompanied by a specific behavioral indicator.

Table 4. 7 Observation Sheet

No.	Aspect	Indicator
1	Students' Attention	Students remain focused during listening and speaking activities throughout the session
2	Participation	Students actively join in singing, speaking tasks, pair work, and answering questions
3	Motivation & Enjoyment	Students show enthusiasm, interest, and positive emotion during the lesson
4	Speaking Engagement	Students attempt to produce spoken English using expressions from the song
5	Listening Engagement	Students can follow the song lyrics and identify missing or key words during listening tasks
6	Pronunciation Attempt	Students attempt to pronounce words and expressions from the song with reasonable accuracy
7	Collaboration	Students cooperate effectively with peers during pair and group work activities
8	Teacher's Role	Teacher uses the song effectively, gives clear guidance, models language appropriately, and maintains positive interaction
9	Classroom Atmosphere	The classroom atmosphere feels positive, relaxed, and conducive to

language learning

This instrument was applied across all six sessions, and the results are shown in Table below.

Table 4. 8 Summary of Classroom Observation Results Across Six Treatment Sessions

Aspect	S1	S2	S3	S4	S5	S6
1. Students' Attention	3	3	3	3	3	3
2. Participation	3	3	3	4	4	3
3. Motivation & Enjoyment	4	4	4	4	4	4
4. Speaking Engagement	3	3	3	3	3	3
5. Listening Engagement	3	3	3	3	3	3
6. Pronunciation Attempt	3	3	3	3	3	3
7. Collaboration	3	3	3	3	3	3
8. Teacher's Role	3	3	3	3	3	3
9. Classroom Atmosphere	3	3	3	3	3	3

Note. Ratings are based on a scale from 1, meaning Very Poor, to 4, meaning Excellent. Sessions 1 through 3 used the song "One Big Family," while Sessions 4 through 6 used "Hold My Hand."

As shown in Table 4.10, the observation results showed a consistently positive pattern throughout the sessions. Eight of the nine rated aspects, namely attention, speaking engagement, listening engagement, pronunciation attempt, collaboration, the teacher's role, and classroom atmosphere, remained stable at a rating of 3, meaning Good, throughout all six sessions, while motivation and enjoyment were rated 4, meaning Excellent, in every session. An interesting pattern emerged in participation, which increased from a rating of 3 in Sessions 1 through 3 to a rating of 4 in Sessions 4 and 5, before returning to 3 in Session 6. This temporary increase coincided with the introduction of the second song, Hold My Hand, suggesting that the novelty of a new song may have temporarily boosted students'

active involvement in singing, pair work, and answering questions. This pattern of strong affective engagement alongside more measured skill-specific gains mirrors Al-Smadi (2021), who found songs significantly increased young learners' motivation, participation, and attention relative to non-song instruction.

Semi-structured interviews were conducted with six students from the experimental group, labeled Student A through Student F, and the class's English teacher, following the completion of all treatment sessions. Interview transcripts were analyzed following Braun and Clarke's (2006) six-phase thematic analysis framework: familiarization with the data, initial coding, searching for themes, reviewing themes, defining and naming themes, and producing the final analysis. This process yielded five recurring themes.

1. Reduced Boredom and Increased Enjoyment. All six students reported reduced boredom and a calmer, more enjoyable learning experience compared to textbook lessons. As Student B described it, the songs made lessons less

- rigid than learning from a regular book, so it was not easy to get bored.
2. **Enhanced Vocabulary Memorization.** Several students described the songs as helping them memorize vocabulary that would otherwise be difficult to pronounce. Student A noted that it was quite fast to remember English vocabulary that is actually somewhat difficult to pronounce, and the teacher independently corroborated this observation, noting that the children became more active and creative and found it easier to memorize vocabulary.
 3. **Uneven Gains in Speaking Confidence.** Reports on speaking confidence varied in intensity across students. Student D reported that listening to the songs helped her know how to pronounce difficult English words, whereas Student A and Student C offered more moderate endorsements, describing only a slight increase in confidence. This pattern is consistent with Kim et al. (2024), who found that music listening reduces components of foreign language anxiety such as communication apprehension, but with effects that differ across individuals. This individual variability offers a qualitative explanation for why the quantitative between-group speaking difference did not reach significance, since the anxiety-reducing benefit of songs appears to be real but unevenly distributed across students.
 4. **Song Tempo as a Barrier to Comprehension and Production.** The most consistently reported difficulty, echoed across nearly all student interviews and by the teacher, was the natural singing tempo of the songs, which made it hard for some students to follow and reproduce the lyrics accurately. Student E described that when the singer says the lyrics quickly, it becomes difficult to keep up. This matches Muhamad and Rahmat (2021), who identified pace and pronunciation as primary barriers to song-based learning, and is consistent with evidence that young EFL learners comprehend better at a moderately reduced speech rate (Chiu & Chen, 2023). The teacher also noted that students could recognize and repeat what they heard in the songs more easily than they could transcribe it in writing, stating that at the basic listening level, students hear the words but remain confused about how to write them down. This gap is consistent with Nie et al.'s (2022) finding that songs strongly support immediate spoken form recognition, but that meaning and written form knowledge require more reinforcement to consolidate.
 5. **Moral and Affective Resonance of Islamic Song Content.** Students consistently connected the songs' Islamic themes of brotherhood, family, and mutual care to their own daily lives, and expressed a wish to continue learning through songs. Student C remarked that the message in the song was useful in daily life, specifically the reminder to live in harmony and help one another. This affective and moral resonance is consistent with Wati (2022), who argues that Islamic pop songs function simultaneously as linguistic input and as a vehicle for character education, and it suggests that the religious and cultural alignment of the song selection likely contributed meaningfully to student engagement in this specific school context.
- ### Limitations
- Four limitations qualify these findings. First, the unequal and modest group sizes, with 30 students in the experimental group and 12 in the control group, arising from the natural composition of two intact classes rather than

random or matched assignment, limited statistical power for between-group comparisons and may partly explain the non-significant speaking result despite a numerically favorable gain. Second, the listening instrument's ceiling effect in the control group prevented a meaningful between-group comparison for that skill, even though the pilot study item difficulty analysis had anticipated this risk. Third, the six-session, three-week intervention window captured only short-term change, and longer-term retention of vocabulary and pronunciation gains was not assessed. Fourth, the study was conducted at a single Islamic elementary school, which limits the generalizability of both the quantitative and qualitative findings to other institutional or cultural contexts

t.

CONCLUSION

This study aimed to investigate if Islamic English pop songs of Maher Zain are suitable tools to enhance the listening and speaking skill of the sixth-grade students and how the students viewed this. The within group improvements in listening and speaking were both statistically significant with large effect sizes in the experimental group suggesting that improvement in listening and speaking was worthwhile in terms of practice. The study did not, however, provide clear quantitative evidence to show that Islamic English pop songs were more effective than conventional teaching: There was a ceiling effect in the control group when the listening comparison was made and a numerical advantage for the song-based group when the speaking comparison was made, but this did not achieve statistical significance. Qualitatively, the students' responses were positive as they reported that the approach motivated them, helped them retain the vocabulary used in the songs, enhanced the understanding of the songs' moral messages,

and compromised the understanding of the song's tempo. Overall, these results confirm that Islamic English popular songs can be considered a pedagogically useful, culturally appropriate complementary language learning tool, and that its use is not proven to replace the traditional language teaching method.

REFERENCES

- Abdelrasheed, P. N. S., Saeed, M. A., Khallaf, D. N. M., & Lamri, D. R. T. A. (2026). Using statistical analysis in educational research: Methods and applications in EFL context. *Journal of World Englishes and Educational Practices*, 8(3), 1–9. doi:10.32996/jweep.2026.8.3.1
- Al-Smadi, M. H. (2021). The effect of using songs on young English learners' motivation in Jordan. *International Journal of Emerging Technologies in Learning*, 15(24), 52–63. doi:10.3991/ijet.v15i24.19311
- Chiu, C.-W., & Chen, T.-P. (2023). Speech rate and young EFL learners' listening comprehension. *English Language Teaching*, 16(7), 74–80. doi:10.5539/elt.v16n7p74
- Cox, T. L. (2022). Temporal fluency and floor/ceiling scoring of intermediate and advanced speech on the ACTFL Spanish Oral Proficiency Interview–Computer. *Language Testing*, 40(2), 325–351. doi:10.1177/02655322221114614
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Halisnawati, Nurmis, & Suriati, I. (2024). Learning methods for Islamic religious education in elementary schools. *International Journal of Education, Culture and Society*, 1(1), 9–14. doi:10.0000/IJECRE.v1.i1.a7
- Hilmy, H., Firdaus, N., & Nurizqiyanti, U. (2024). The effect of using English song

- media toward students' speaking skills. *Joey*, 3(1), 71–85. doi:10.35316/joey.2024.v3i1.71-85
- Kamil, A. T. (2024). Exploring the relationship between students' listening to English pop songs and vocabulary mastery. *Indonesian Teaching English to Speakers of Other Languages Journal*, 1(1), 39. doi:10.30587/inatesolv1i1.7367
- Kim, H.-J., Chong, H. J., & Lee, M. (2024). Music listening in foreign language learning: Perceptions, attitudes, and its impact on language anxiety. *Frontiers in Education*, 9, 1372786. doi:10.3389/feduc.2024.1372786
- Kitjaroonchai, T. (2025). Effects of English songs on EFL learners' speaking performance: Lexical complexity, accuracy, and fluency. *Arab World English Journal*, 16(3), 125–143. doi:10.24093/awej/vol16no3.7
- Lengari, Y. G. P. (2024). The role of English pop songs in listening skills learning among maritime students. *Saintara: Jurnal Ilmiah Ilmu-Ilmu Maritim*, 8(2), 74–80. doi:10.52475/saintara.v8i2.334
- Muhamad, N., & Rahmat, N. H. (2021). Investigating challenges for learning English through songs. *European Journal of English Language Teaching*, 6(1), 1–15. doi:10.46827/ejelv6i1.3270
- Mulyani, D. S., Darsih, E., Setiorini, S. H., Fauziah, Octapiani, G., & Eriani, T. (2026). English songs for enhancing university students' pronunciation: A qualitative case study. *JEIT (Journal of Educational Innovations and Technologies)*, 2(1), 70–89. doi:10.63324/jeit.2.1.2026.149
- Ndayisenga, J. d'Amour, & Laos Mbato, C. (2024). Exploring types of language learning motivation among EFL students: Insights from Rwandan high schools. *ELS Journal on Interdisciplinary Studies in Humanities*, 7(4), 617–625. doi:10.34050/els-jish.v7i4.42097
- Nie, K., Fu, J., Rehman, H., & Zaigham, G. H. K. (2022). An empirical study of the effects of incidental vocabulary learning through listening to songs. *Frontiers in Psychology*, 13, 891146. doi:10.3389/fpsyg.2022.891146
- Nurdin, M., Asmawati, A., & Najamuddin, M. (2023). Investigating the effect of using English Islamic pop songs on students' vocabulary size. *Indonesian TESOL Journal*, 5(2), 272–283.
- Pino, M. C., Giancola, M., & D'Amico, S. (2023). The association between music and language in children: A state-of-the-art review. *Children*, 10(801), 1–18. doi:10.3390/children10050801
- Rorintulus, O. A., & Wuntu, C. N. (2023). Exploring the effectiveness of songs for learning improvement: A case of elementary level students' English skills. *Edumaspul: Jurnal Pendidikan*, 7(2), 3336–3349. doi:10.33487/edumaspul.v7i2.6913
- Rumiyati, Lestari, A. P., & Juanda. (2025). The role of learning media in enhancing the effectiveness and activeness of the learning process. *Journal Didaskalia*, 8(2), 114–127. doi:10.33856/didaskalia.v8i2.614
- Safitri, I. A., & Najwa, S. (2024). A conceptual analysis of utilizing English songs in teaching speaking skills. *LinguaEducare: Journal of English and Linguistic Studies*, 1(2), 97–106. doi:10.63324/lec.1v.2i.13
- Simanjuntak, P. H., Sihalo, J. H., Limbong, N., & Sinurat, B. (2025). The effectiveness of songs as a phonology-based English pronunciation teaching tool. *Young Journal of Social Sciences and Humanities*, 1(3), 121–130. doi:10.66867/yjssh.v1i3.174
- Staus, N. L., O'Connell, K., & Storksdieck, M. (2021). Addressing the ceiling effect when assessing STEM out-of-school time

- experiences. *Frontiers in Education*, 6, 690431. doi:10.3389/feduc.2021.690431
- Supriyani, Kunah, & Fatin, R. M. (2025). Enhancing students' listening comprehension with English songs. *JELL (Journal of English Language and Literature)*, 10(1), 137–150. doi:10.37110/jell.v10i01.274
- Torras-Vila, B. (2021). Music as a tool for foreign language learning in early childhood education and primary education. *CLIL Journal of Innovation and Research in Plurilingual and Pluricultural Education*, 4(1), 35–47. doi:10.5565/rev/clil.60
- Wahyudi, M. S., & Masduqi, H. (2026). Exploring Spotify as an informal learning tool for listening comprehension: A mixed-method study. *SSRN Electronic Journal*, 1–19. doi:10.2139/ssrn.6687758
- Widyaningtyas, Y., & Bambang, W. P. (2024). Exploring the impact of English song listening habits on listening proficiency. *International Journal of Visual and Performing Arts*, 6(2), 216–225. doi:10.31763/viperarts.v6i2.1311.