E-ISSN: 3031-8629
P-ISSN: 3032-3258
Halaman 172-183
Volume 1 Nomor 4 Tahun 2024

YouTube Music Videos in EFL Listening Instruction: Song-Lyric Comprehension and Student Motivation

[1]Muhammad Amin, [2]Aprianoto, [3]Edi Firman

[123]English Language Education Department, FBMB, Universitas Pendidikan Mandalika, Indonesia
e-mail: amin06022000@gmail.com

Abstract

This quantitative study examined changes in song-lyric comprehension and self-reported learning motivation following the use of YouTube music videos in English listening instruction. A one-group pretest–posttest design was implemented with 25 eleventh-grade students at MA Malaka, North Lombok. The listening assessments consisted of ten gap-completion items and five essay questions; an eight-item questionnaire assessed motivation before and after the instructional intervention. Mean listening scores increased from 66.60 (SD = 2.38) to 76.84 (SD = 6.02), an absolute difference of 10.24 points. The source analysis reported a paired-samples significance value of $p < .001$. Motivation questionnaire totals likewise increased from a mean of 12.28 to 27.76, with post-intervention classifications of 8% very high, 72% high, and 20% moderate motivation. These findings document favorable within-group changes in the measured outcomes. However, the absence of a comparison group limits causal attribution, and the questionnaire scoring and inferential output require further verification. YouTube music videos may be a promising supplementary resource for listening practice when accompanied by structured tasks and explicit comprehension goals.

Keywords: YouTube music videos; listening comprehension; song lyrics; EFL; learning motivation

Abstrak

Penelitian kuantitatif ini mengkaji perubahan pemahaman lirik lagu dan motivasi belajar setelah penggunaan video musik YouTube dalam pembelajaran menyimak bahasa Inggris. Desain satu kelompok dengan tes awal dan tes akhir diterapkan pada 25 siswa kelas XI MA Malaka, Lombok

Utara. Instrumen meliputi tes menyimak dengan sepuluh soal melengkapi bagian rumpang dan lima soal uraian serta angket motivasi delapan butir sebelum dan sesudah perlakuan. Rata-rata nilai menyimak meningkat dari 66,60 (SD = 2,38) menjadi 76,84 (SD = 6,02), dengan selisih 10,24 poin. Analisis sumber melaporkan hasil uji berpasangan p < 0,001. Rata-rata skor angket motivasi juga berubah dari 12,28 menjadi 27,76, dengan kategori setelah pembelajaran sebesar 8% sangat tinggi, 72% tinggi, dan 20% sedang. Hasil tersebut menunjukkan perubahan positif dalam satu kelompok, tetapi tidak cukup untuk menetapkan hubungan sebab-akibat karena tidak ada kelompok pembanding. Penggunaan video musik YouTube berpotensi mendukung latihan menyimak apabila dipadukan dengan tugas terstruktur dan tujuan pemahaman yang jelas.
Kata kunci: video musik YouTube; pemahaman menyimak; lirik lagu; bahasa Inggris sebagai bahasa asing; motivasi belajar

INTRODUCTION

Listening comprehension is a central component of foreign-language learning because spoken messages require learners to recognize acoustic signals, segment words, retrieve meaning, and integrate contextual information under time constraints (Field, 2008; Rost, 2011). Unlike reading, spoken input is transient and frequently includes connected speech, reduced forms, unfamiliar accents, and background sounds. Instruction that relies exclusively on scripted textbook audio may not expose learners to the varied input they encounter in everyday digital environments.

Music videos combine auditory language, rhythm, visual context, and emotionally salient content. These characteristics may make them useful for repeated listening, vocabulary recognition, and noticing word boundaries. Multimedia learning theory suggests that verbal and visual information can complement each other when the presentation is carefully designed and unnecessary cognitive load is controlled (Mayer, 2009). Nevertheless, entertainment value alone does not guarantee comprehension. Learners need purposeful questions, manageable input, opportunities for replay, and feedback on misunderstood expressions.

YouTube offers accessible audiovisual material and playback controls that may facilitate self-paced practice. Prior work has discussed its potential as an English-learning resource (Watkins & Wilkins, 2011; Kelsen, 2009). Music and songs have also been associated with classroom engagement and opportunities for repeated language exposure (Murphey, 1992; Dörnyei, 2001). Such accounts provide pedagogical justification for exploring song-based instruction, but they do not remove the need for empirical assessment of what students understand after exposure. Research in Indonesian secondary education has also examined students' responses to English educational videos as a listening resource (Sembiring & Katemba, 2023).

Motivation is particularly relevant to listening tasks because students may disengage when they cannot immediately recognize spoken words. Self-determination theory emphasizes competence, autonomy, and relatedness as conditions supportive of engagement (Ryan & Deci, 2000), while second-language motivation research highlights the role of learners' goals and learning experiences (Dörnyei, 2005). Music videos might contribute to interest and persistence, although questionnaire responses can also reflect novelty, classroom expectations, or the appeal of familiar songs.

The local context of MA Malaka, North Lombok, provided a setting in which students' listening comprehension and motivation were investigated using a short instructional intervention. The study focused on two outcomes: performance on a song-lyric comprehension assessment and self-reported motivation toward English listening. Rather than claiming to isolate the unique effects of YouTube, the study offers a situated account of pre–post patterns and the methodological qualifications needed to interpret them.

Two questions guided the analysis: (1) How did the students' song-lyric comprehension scores differ before and after the use of YouTube music videos? (2) How did their reported motivation levels differ across the same instructional period? By reporting both learning performance and motivational classifications, the article

contributes a classroom-based example of audiovisual listening instruction in an Indonesian secondary-school EFL setting.

The study contributes a situated account of audiovisual listening instruction in an Indonesian Islamic senior secondary school. Rather than treating YouTube exposure as an outcome in itself, it considers two separate learning-related indicators: performance on a song-lyric comprehension assessment and students' reported motivation. This distinction matters because enjoyable classroom media do not necessarily produce equivalent improvements in comprehension, and performance changes cannot be interpreted solely through students' positive perceptions.

The pedagogical rationale is informed by multimedia learning, which proposes that coordinated verbal and visual information can support understanding when material is carefully selected and cognitive demands are managed (Mayer, 2009). In language education, digital tools can create additional opportunities for practice, but their value depends on the learning task rather than the technology alone. For example, Haerazi, Utama, and Hidayatullah (2020) examined mobile-assisted English writing instruction among prospective teachers; their findings concern a different skill and population but illustrate the importance of purposeful technology integration. In a more directly related context, Sulaiman et al. (2022) investigated the association between YouTube viewing frequency and listening comprehension. Their correlational design, like the present pre-experimental design, calls for careful interpretation of causal claims.

Against this background, the analytical focus was deliberately narrow: documenting the direction and size of observed score changes and considering whether the accompanying motivation responses were consistent with students' engagement in listening activities. The study does not test a theory of multimedia processing, establish a mechanism of improvement, or compare the platform with an alternative instructional method.

METHOD

The study used a quantitative pre-experimental one-group pretest–posttest design. It was conducted at MA Malaka, North Lombok. The participants were 25 Grade XI students selected from the school population. Because all participants received the intervention and no parallel comparison group was assessed, the design supported analysis of within-group change but did not control for maturation, repeated testing, or other concurrent influences.

Before the intervention, students completed a listening comprehension pretest and an eight-item motivation questionnaire. During instruction, YouTube music videos were used as listening material for recognizing and interpreting song lyrics. After the intervention, the students completed a corresponding listening posttest and motivation questionnaire. The source report did not establish a detailed intervention fidelity checklist or independent observation procedure, so these were not assumed in the present article.

The listening instrument included ten gap-completion items and five essay questions designed to assess comprehension of song lyrics. The questionnaire contained eight statements with four-point responses; the source report included item validity and internal-consistency procedures. The exact correspondence of the pretest and posttest items and the questionnaire coding rules could not be independently verified from the available summary tables. These measurement considerations were therefore retained as limitations rather than silently resolved.

The analysis summarized listening scores with means, standard deviations, score ranges, and the absolute mean difference. The original report presented paired-samples t-test outputs and interpreted a displayed significance value of .000; in this article the value is written as $p < .001$, consistent with statistical reporting conventions. Motivation results were summarized using reported mean raw scores and the distribution of the five categorical levels. The analysis distinguished statistically detectable pre–post differences from causal effects attributable specifically to the intervention.

For the listening measure, the reported mean difference was computed as the posttest mean minus the pretest mean ($76.84 - 66.60 = 10.24$). This is a descriptive

difference on the score scale used in the source study, not an adjusted treatment effect. The standard deviations and observed ranges are reported as provided; individual raw scores were not independently reanalyzed for this article.

For the motivation measure, questionnaire totals and the reported category frequencies were retained as separate forms of evidence. Because the source does not provide enough detail to reconstruct the transformation of raw scores into category cutoffs, this article does not assign a new scale interpretation, estimate internal consistency, or infer individual-level change from the aggregated categories.

FINDINGS AND DISCUSSION

The listening assessment showed a higher average score after the use of YouTube music videos. The mean increased by 10.24 points, from 66.60 to 76.84. The posttest distribution was more dispersed than the pretest distribution, suggesting that students did not benefit uniformly. The lowest posttest score was 63, below the pretest minimum of 65; therefore, the group-level increase should not be interpreted as improvement for every participant.

Table 1. Pretest and posttest score

Listening indicator	Pretest	Posttest
Participants	25	25
Mean score	66.60	76.84
Standard deviation	2.38	6.02
Minimum	65	63
Maximum	70	93
Reported paired test	—	$p < .001$

The individual-score table in the source report recorded one student whose score decreased from 65 to 63, whereas other students generally showed positive differences. This variation is educationally important: song-based audiovisual instruction may appeal

to many learners, but difficulties with speech rate, unfamiliar vocabulary, or competing visual information can remain for individual students. A more informative follow-up study would report item-level patterns and identify which lyric-comprehension processes changed most.

The paired-samples analysis was reported as statistically significant ($p < .001$). This supports a difference between the two observed test occasions under the assumptions of the analysis. However, the intervention was not randomly assigned and no control group was present. The difference may therefore reflect a combination of instruction, repeated exposure to task format, familiarity with song content, and other learning experiences rather than the isolated influence of the platform.

Motivation results followed a similar favorable direction. The source reported a mean questionnaire score of 12.28 before the intervention and 27.76 afterward, an increase of 15.48 raw-score points on an eight-item measure. The source also reported that 88% of students were classified as having low motivation and 8% very low motivation before instruction; after instruction, 72% were classified as high and 8% as very high.

Table 2. Students' motivation result

Motivation category	Before: n (%)	After: n (%)
Very high	0 (0%)	2 (8%)
High	0 (0%)	18 (72%)
Moderate	1 (4%)	5 (20%)
Low	22 (88%)	0 (0%)
Very low	2 (8%)	0 (0%)
Total	25 (100%)	25 (100%)

The categorical distributions suggest a marked change in how students endorsed the questionnaire statements. Nevertheless, interpretation requires caution. The raw totals, transformed values, and category boundaries in the source tables were not

uniformly explained; the displayed post-intervention mean and maximum should be audited against the 25 original response records. In addition, the questionnaire was administered in the same classroom as the instructional treatment, which may introduce social desirability or novelty effects. These concerns do not invalidate the descriptive pattern but limit claims about a durable motivational change.

One plausible explanation for the listening-score pattern is that repeated exposure to sung English encouraged students to focus on lexical forms and sound–word correspondences. Song lyrics can provide memorable language sequences, and repeated listening allows learners to revise earlier interpretations. This account is compatible with pedagogical discussions of songs in language teaching (Murphey, 1992) and the importance of strategic listening practice (Field, 2008). Yet singing also differs from ordinary speech in rhythm, pitch, and articulation, so comprehension gains in lyric tasks cannot automatically be generalized to academic lectures or everyday conversations. Related research has reported a positive association between English-language YouTube viewing and listening performance (Sulaiman et al., 2022), although association should not be confused with experimental evidence of effectiveness.

The motivational findings are consistent with the possibility that familiar digital media helped students approach listening tasks with greater interest. Digital audiovisual materials can provide a sense of choice and immediate contextual cues. However, interest in a particular activity and sustained motivation for language learning are different constructs. Stronger evidence would require repeated measures over time, a clear questionnaire scoring key, and possibly interviews or classroom observations that distinguish enthusiasm for the media from willingness to undertake difficult listening tasks. Studies of technology-assisted language learning at Universitas Pendidikan Mandalika likewise highlight the need to connect digital media with explicit language-learning processes rather than assuming that access alone ensures learning (Haerazi et al., 2020).

For teaching practice, YouTube music videos are best integrated into a structured sequence: activate relevant vocabulary before listening; ask learners to predict the topic; use an initial play for general meaning; replay selected segments for gap completion or interpretation; and follow with discussion of difficult expressions. Teachers should select songs with appropriate language density, comprehensible pronunciation, and suitable content, while avoiding overreliance on subtitles that may shift the task from listening to reading.

The findings also underline the importance of assessment design. Gap-filling can measure recognition of words or phrases, whereas short-answer items can probe broader meaning. Future investigations should document equivalent pretest and posttest difficulty, inter-rater agreement for essay scoring, and internal consistency for motivation items. A comparison class and delayed posttest would make it possible to distinguish immediate task familiarity from longer-term comprehension development.

This investigation was limited by a single school, a small sample of 25 students, a one-group design, incomplete reporting of measurement equivalence, and uncertainty surrounding the questionnaire transformations. The original report did not provide sufficient information to calculate a defensible independent causal effect or to confirm all reported inferential statistics from raw data. The results should thus be read as encouraging classroom evidence rather than proof that YouTube music videos alone produced the changes.

The score distributions add nuance to the overall positive result. The pretest scores clustered tightly between 65 and 70, while posttest scores ranged from 63 to 93. Thus, the higher posttest mean coincided with substantially greater dispersion. A classroom average can mask learners who need differentiated support, especially when comprehension depends on familiarity with vocabulary, speech rate, accents, or the musical arrangement. The recorded decrease for one participant is a reminder that the instructional experience was not uniform.

From an instructional standpoint, teachers should choose music videos with clear vocals, age-appropriate lyrics, manageable lexical density, and visuals that genuinely clarify meaning. Repeated playback should be accompanied by explicit listening purposes. A first viewing might target the general topic; a second could focus on missing lexical items; and a final discussion could check inferential meaning. Without these stages, learners may attend mainly to the visual narrative or entertainment value rather than to the linguistic signal.

Interpretation of motivation also requires a distinction between reported enthusiasm and durable motivation. The post-intervention category pattern suggests more favorable responses at the time of measurement, but the study did not include delayed follow-up or independent behavioral indicators such as sustained voluntary listening. Future studies could combine validated motivational scales with listening logs, delayed assessments, and learner interviews to clarify whether interest persists beyond the novelty of the platform.

A stronger subsequent evaluation would use a comparison group receiving similarly structured listening tasks through another medium, document instructional duration and song selection, and employ equivalent test forms with clear scoring rubrics. Researchers could then assess whether any difference is associated with the video format, the music, repeated exposure, or the structured listening tasks themselves. These recommendations arise from the methodological limits of the present evidence rather than from unreported observations.

CONCLUSION

Among 25 Grade XI students at MA Malaka, song-lyric comprehension scores increased from a mean of 66.60 to 76.84 after listening instruction using YouTube music videos, and the source report identified a statistically significant paired difference ($p < .001$). Reported motivation scores and categorical classifications also shifted toward more favorable levels. These results suggest that music videos may be a useful supplementary

medium for structured EFL listening activities. Nevertheless, the absence of a comparison group and incomplete measurement documentation prevent a definitive causal interpretation. Future research should employ comparable instruments, transparent questionnaire scoring, control conditions, and follow-up assessment to evaluate whether the observed changes persist and generalize to broader listening comprehension.

REFERENCES

Brown, H. D. (2007). Principles of language learning and teaching (5th ed.). Pearson Longman.

Creswell, J. W. (2012). Educational research: Planning, conducting, and evaluating quantitative and qualitative research (4th ed.). Pearson.

Dörnyei, Z. (2001). Motivational strategies in the language classroom. Cambridge University Press.

Dörnyei, Z. (2005). The psychology of the language learner: Individual differences in second language acquisition. Lawrence Erlbaum Associates.

Field, A. (2018). Discovering statistics using IBM SPSS statistics (5th ed.). SAGE.

Field, J. (2008). Listening in the language classroom. Cambridge University Press.

Kelsen, B. (2009). Teaching EFL to the iGeneration: A survey of using YouTube as supplementary material with college EFL students in Taiwan. CALL-EJ Online, 10(2).

Krashen, S. D. (1982). Principles and practice in second language acquisition. Pergamon Press.

Mayer, R. E. (2009). Multimedia learning (2nd ed.). Cambridge University Press.

Murphey, T. (1992). Music and song. Oxford University Press.

Richards, J. C. (2008). Teaching listening and speaking: From theory to practice. Cambridge University Press.

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

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Vandergrift, L., & Goh, C. C. M. (2012). Teaching and learning second language listening: Metacognition in action. Routledge.

Watkins, J., & Wilkins, M. (2011). Using YouTube in the EFL classroom. Language Education in Asia, 2(1), 113–119. https://doi.org/10.5746/LEiA/11/V2/I1/A09

Wilson, J. J. (2008). How to teach listening. Pearson Longman.

Haerazi, H., Utama, I. M. P., & Hidayatullah, H. (2020). Mobile applications to improve English writing skills viewed from critical thinking ability for pre-service teachers. International Journal of Interactive Mobile Technologies, 14(7), 58–72. https://doi.org/10.3991/ijim.v14i07.11900

Sulaiman, E., Aprianti, I., Siahaan, D. G., & Eni, G. D. (2022). A correlation study between frequency of viewing YouTube and students' listening comprehension. Jurnal Edulanguage: Jurnal Pendidikan Bahasa, 8(2), 14–27. https://doi.org/10.32938/edulanguage.8.2.2022.14-27

Sembiring, H. T. R., & Katemba, C. V. (2023). The use of YouTube English educational videos in improving listening comprehension. Journey: Journal of English Language and Pedagogy, 6(1), 161–170. https://doi.org/10.33503/journey.v6i1.622