IDEAS

Journal on Language Teaching and Learning, Linguistics and Literature

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Volume 14, Number 2, December 2026
pp. 9663 – 9670

Copyright © 2026 The Author
IDEAS is licensed under CC-BY-SA 4.0 License



Issued by English study program of UIN Palopo

Students' Perceptions of Sound-Based Media In Vocabulary Learning at Junior High School

Al Fikky Nur[1], Hastini[2], Desrin Lebagi[3], Zarkiani Hasyim[4]

[1,2]English Education Study Program, Faculty of Teacher Training and Education, Universitas Tadulako

Article Info

Received: 2026-10-23
Revised: 2026-11-11
Accepted: 2026-11-14

Keywords:
students' perceptions,
sound-based media,
spoken vocabulary,
vocabulary learning,
junior high school

DOI:
10.24256/ideas.v14i2.12989

Corresponding Author:
Al Fikky Nur,
fikijr058@gmail.com
English Education Study Program, Faculty of Teacher Training and Education, Universitas Tadulako

Abstract

This descriptive qualitative study identified the sound-based media used in vocabulary instruction and examined how Grade 8 students experienced those materials in an English class at SMP Negeri 3 Banawa. Data were collected through classroom observation and semi-structured interviews with five purposively selected students from Class VIII A and analyzed using the interactive qualitative model of Miles et al. (2014). Observation documented English songs and spoken audio, while interviews also identified textbook and pronunciation recordings, short conversations, audiovisual clips, and audio from learning applications. The materials were typically combined with teacher explanation, repetition, vocabulary identification, and sentence practice. Students generally valued the media because they provided access to pronunciation and made lessons more enjoyable, particularly when songs were familiar. These perceived benefits were conditional: fast delivery, unfamiliar vocabulary, similar-sounding words, and difficulty recognizing spoken forms of words already known in writing limited comprehension. The findings represent students' perceptions rather than measured gains in vocabulary achievement. Teachers should therefore select level-appropriate audio, pre-teach key vocabulary, repeat difficult segments, and provide written support that helps learners connect spelling, sound, and meaning.

1. Introduction

Students in Class VIII A at SMP Negeri 3 Banawa encounter a familiar vocabulary-learning problem: they may recognize English words in a textbook yet struggle to identify or pronounce the same words when they hear them. This gap matters because vocabulary supports listening, speaking, reading, and writing, and knowledge of a word includes more than its written form (Nation, 2013; Crismeilani et al., 2025). In a communicative

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

English class, learners need opportunities to connect spelling and meaning with the word's spoken form. Sound-based media can provide that input, but their value depends on whether students can process the audio and on how the teacher supports listening.

Sound-based media are defined in this study as instructional materials in which the audio component carries the target spoken vocabulary. They include audio-only materials, such as songs, textbook recordings, pronunciation models, and short conversations. Audiovisual clips are included only when their audio track provides the spoken language being examined; they are identified separately from audio-only media in the findings. This distinction avoids treating every video as sound-based simply because it contains sound.

The relevance of these materials follows a clear conceptual sequence. Sound-based media provide spoken vocabulary input; learners attend to pronunciation, stress, rhythm, and word boundaries; their experience of the input shapes whether they perceive it as accessible, enjoyable, or difficult; and teacher scaffolding helps them link sound with spelling and meaning. Receptive vocabulary includes words recognized in reading and listening, but recognition in one mode does not automatically guarantee recognition in the other (Dakhi & Fitria, 2019). Audio therefore offers a necessary opportunity to develop auditory recognition, while also creating processing demands related to speech rate, unfamiliar vocabulary, connected speech, and pronunciation clarity.

Previous research supports several parts of this relationship. Studies of songs and audiovisual materials have reported vocabulary and motivational benefits (Hartono, 2013; Rochma & Mukminatus, 2025), while research on audio-assisted and mobile-supported learning has emphasized access to varied input, pronunciation, engagement, and repetition (Bagila et al., 2019; Lhamo & Sakulwongs, 2023; Wan Ji & Abdul Aziz, 2021; Simonnet et al., 2024). Perception-focused research also shows that learners' responses to listening materials can inform the selection of classroom resources (Ardiansyah et al., 2025). Across this literature, four themes recur: audio as pronunciation input, songs as an engaging vocabulary context, learner perceptions of audio, and difficulty processing spoken words.

However, much of the previous literature has asked whether a particular song, platform, or audiovisual treatment is effective. Less attention has been given to the range of materials teachers actually use in an ordinary junior high school vocabulary class and to the way students experience those materials, including the tension between enjoyment and comprehension. A descriptive account is useful in SMP Negeri 3 Banawa because it can identify both the available forms of spoken input and the conditions that students perceive as supportive or obstructive. It does not test vocabulary achievement and therefore does not use perceptions as evidence of learning gains.

The study addressed two research questions: (1) What types of sound-based media are used in vocabulary learning in Class VIII A at SMP Negeri 3 Banawa? and (2) How do selected Grade 8 students perceive the use of those media for vocabulary learning? The study contributes a contextual account of how audio characteristics, students' existing vocabulary knowledge, familiarity with the content, and teacher scaffolding interact in students' reported experiences.

2. Method

2.1 Research design and setting

The study used a descriptive qualitative design to examine classroom practice and students' reported experiences rather than to test the effectiveness of an intervention (Creswell, 2014). It was conducted over two weeks in Class VIII A at SMP Negeri 3 Banawa. The unit of inquiry was the students' experience of sound-based media during vocabulary learning, while classroom observation provided contextual evidence about the media and the way they were used.

2.2 Participants and sampling

Five Grade 8 students from Class VIII A participated and were identified as P1-P5. Purposive sampling was used because the study required information-rich participants who had experienced vocabulary learning with sound-based media and could describe their responses. The sample was intended to support a focused qualitative account, not statistical representation. Eligibility criteria were prior exposure to the media, willingness to participate, and ability to communicate relevant experiences. The small sample and single-class setting limit the transferability of the findings.

2.3 Data collection and instruments

Data were collected through classroom observation and individual semi-structured interviews. Observation focused on (a) the audio-only and audiovisual materials used, (b) how the teacher introduced, played, repeated, and explained them, and (c) students' attention, participation, and observable difficulties. Relevant events were documented in field notes. The interviews then explored the media students remembered, how the media were used, their opinions, perceived support for understanding or remembering vocabulary, difficulties, and suggestions for instruction. Follow-up prompts were used when clarification was needed.

The interview guide contained six open-ended questions aligned with the two research questions. The first two questions addressed the types and classroom use of sound-based media; the remaining questions addressed students' evaluations, perceived benefits, challenges, and recommendations. Observation and interview evidence were kept distinct during analysis so that media witnessed by the researcher were not presented as if they had all been observed directly.

2.4 Data analysis and credibility

Analysis followed the interactive model of data condensation, data display, and conclusion drawing and verification (Miles et al., 2014). Observation notes and interview records were reviewed in relation to the two research questions. The top-level categories were deductive: types and uses of media for Research Question 1, and students' perceptions for Research Question 2. Within those categories, descriptive codes were grouped into themes concerning pronunciation access, enjoyment and familiarity, repetition, speech rate, unfamiliar vocabulary, similar-sounding words, written-versus-spoken recognition, and requested instructional support. Participant variation and contradictory evidence, particularly P5's mixed view, were retained rather than collapsed into a uniformly positive account.

Credibility was supported by comparing the interview themes with classroom observation. Agreement between the sources was treated as corroboration, while interview-only reports were labelled as reported experience rather than direct

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

observation. The analysis also distinguished students' perceptions of helpfulness from demonstrated vocabulary improvement.

3. Results

3.1 Research Question 1 Types and classroom use of sound-based media

Classroom observation documented two audio-only resources: English songs and spoken audio played through a speaker. The songs placed vocabulary in a musical context, while the spoken recordings provided pronunciation models. Interviews confirmed these resources and added media that were not observed during the researcher's classroom visit: textbook and pronunciation recordings, short English conversations, audio from learning applications, and short videos. The videos are classified as audiovisual rather than audio-only; they are included because students described their audio tracks as sources of spoken vocabulary.

Table 1. Sound-based media identified through observation and interviews

Category	Medium	Classroom observation	Interview evidence
Audio-only	English songs	Observed	P1, P2, P4, P5
Audio-only	Spoken audio / pronunciation recordings	Observed	P2; also described in different forms
Audio-only	Textbook audio recordings	Not observed	P1, P3
Audio-only	Short English conversations	Not observed	P3
Audio-only	Audio from learning applications	Not observed	P4
Audiovisual	Short videos with spoken English	Not observed	P1, P2

Across the students' accounts, the media were embedded in a recurring instructional sequence rather than used as passive listening materials. The teacher played a song, recording, or clip; students listened for words; they repeated pronunciation; and the teacher explained meanings. Students sometimes wrote new vocabulary or used it in sentences. The answer to Research Question 1 is therefore not only that several media formats were used, but that spoken input was combined with explanation, repetition, and active vocabulary practice.

3.2 Research Question 2 Students' perceptions of sound-based media

Students' responses were generally favorable but conditional. The strongest positive theme was direct access to pronunciation. P1 explained, 'I like learning vocabulary through sound because I can hear how the words are pronounced correctly,' and P3 similarly valued hearing pronunciation directly. Familiarity also shaped enjoyment: P2 and P4 reported that known songs made the lesson more interesting and less monotonous. P5's statement that the media were 'helpful, but sometimes difficult' captured the central variation in the data: a positive attitude did not mean that the spoken input was consistently comprehensible.

Students also perceived repeated listening as useful for remembering words. P1, P3, and P5 associated repeated exposure with connecting pronunciation to meaning, while P2 and P4 associated familiar songs with easier recall. These accounts describe perceived support; the study did not test retention or vocabulary gains.

Processing difficulty was reported by all five participants, although its form varied. Speech or singing that was too fast was the most common problem. P4 also mentioned similar-sounding words, and P5 emphasized unknown vocabulary. The most analytically important pattern concerned the gap between orthographic and auditory recognition. P2 said, 'Sometimes I know the written word, but I cannot recognize it when I hear it,' while P3 reported understanding vocabulary after the teacher's explanation but not recognizing it in the audio. Their difficulty was therefore not simply lack of vocabulary; some known written forms were not yet securely linked to their spoken realizations.

The students' recommendations fell into three categories. Audio modification included clear recordings, slower initial playback, repetition, and level-appropriate materials. Teacher scaffolding included pre-teaching difficult words, explaining meanings, and replaying challenging segments. Written or visual support included lyrics and displayed words that enabled students to connect spelling with sound. Table 2 summarizes the themes and participant variation that answer Research Question 2.

Table 2. Themes in students' perceptions of sound-based media

Theme	Evidence and participant variation	Participants
General orientation	Interesting, enjoyable, or useful; P5 also emphasized difficulty	P1-P5
Pronunciation access	Hearing correct pronunciation directly	P1, P3, P5
Enjoyment and familiarity	Familiar songs made lessons more enjoyable and less monotonous	P2, P4
Perceived support for recall	Repeated listening and linking a word's sound with its meaning	P1, P3, P5
Processing barriers	Fast delivery, unfamiliar vocabulary, and similar-sounding words	P1-P5
Written-spoken recognition gap	Known written words were not always recognized in the audio	P2, P3
Requested audio modification	Clearer recordings, slower first playback, repetition, suitable difficulty	P1, P3, P5
Requested teacher scaffolding	Pre-teaching and explaining difficult vocabulary; focused replay	P1, P3, P5
Requested written support	Lyrics or displayed words to connect spelling, sound, and meaning	P2, P4

4. Discussion

4.1 The perceived value of audio is conditional

The central finding is not simply that the students liked sound-based media. They valued the media when spoken input was accessible and connected to support. Pronunciation models, familiar songs, repetition, and teacher explanation made the

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

materials attractive, whereas fast delivery, unfamiliar words, and weak sound-spelling connections made the same type of input difficult. The pedagogical value perceived by these students therefore emerged from the interaction among audio characteristics, learner knowledge, familiarity, and teacher scaffolding. This account extends research that associates audio and technology with engagement by identifying conditions under which students in this class experienced that engagement (Bagila et al., 2019; Simonnet et al., 2024).

The findings also reveal a tension between enjoyment and comprehension. Familiar songs were enjoyable and were perceived as memorable, which is consistent with reports of motivational value in song- and audio-supported learning (Lhamo & Sakulwongs, 2023; Rochma & Mukminatus, 2025). Yet a student could enjoy a song while missing words because singing was fast or pronunciation was difficult to segment. Familiarity may reduce apprehension and increase attention, but the present evidence does not establish that it produces vocabulary acquisition. Enjoyment and comprehensibility should therefore be evaluated separately when teachers select materials.

4.2 Written vocabulary knowledge does not ensure auditory recognition

The contrast between written-word recognition and spoken-word recognition is the study's strongest analytical contribution. P2 and P3 described words that were known in writing or after explanation but were not identified in continuous speech. This pattern suggests that their receptive knowledge was uneven across modes. Vocabulary knowledge includes both orthographic and phonological representations, and listening requires learners to locate word boundaries and match a transient sound sequence with stored meaning (Nation, 2013; Dakhi & Fitria, 2019). A learner may therefore possess partial knowledge of a word without recognizing its natural spoken form.

This distinction helps explain why repetition was perceived as useful. Replaying an item can give learners another opportunity to notice the sound pattern and relate it to a known spelling and meaning. However, repetition alone may be insufficient if the recording remains too fast or contains many unknown words. Its value depends on focused attention, explanation, and an appropriate level of difficulty. The study supports using repetition as scaffolding, not claiming it as a proven cause of retention.

4.3 Balancing authentic input and comprehensibility

Natural recordings can expose learners to realistic pronunciation, rhythm, and connected speech, but authenticity can increase processing demands for students with limited listening proficiency. The participants' reports point to four variables that teachers should inspect before using a recording: speech rate, pronunciation clarity, vocabulary familiarity, and linguistic complexity. A material need not be slowed permanently; teachers can first make the input comprehensible, then move toward a more natural rate as students become familiar with the words. This staged approach preserves exposure to spoken English while responding to the accessibility problems found in this class.

The findings support a concrete classroom sequence. Teachers can pre-teach a small number of essential words, conduct a first listening for general meaning, replay the material for target-word identification, display lyrics or key words, repeat difficult segments, and finish with pronunciation or sentence practice. Such sequencing connects

sound, spelling, meaning, and use. It also reflects the existing classroom pattern reported by the students, in which audio was combined with explanation and active practice rather than delivered as passive listening (Nguyen, 2022).

4.4 Contribution and limitations

The study contributes a contextual explanation of how Grade 8 EFL learners experience spoken vocabulary input. It shows that media format alone does not determine perceived value; accessibility results from the relationship among the recording, the learner, and classroom support. This interpretation moves beyond the broad claim that audio is useful and identifies why the same material can be enjoyable yet difficult.

The findings remain specific to five students in one class at one school. The study did not measure vocabulary achievement, compare proficiency levels, or include the teacher's perspective. Participants may also have given socially acceptable responses because they were speaking to a researcher about classroom practice. Future studies could include teachers, compare different audio difficulty levels, examine more classes, and use vocabulary or listening measures when the objective is to test learning outcomes.

5. Conclusion

This study examined the sound-based media used in one Grade 8 English class at SMP Negeri 3 Banawa and the perceptions of five purposively selected students. English songs and spoken recordings were observed directly. Students additionally reported textbook and pronunciation recordings, short conversations, learning-application audio, and short audiovisual clips. These materials were generally integrated with teacher explanation, repetition, vocabulary identification, writing, and sentence practice.

Students valued the media primarily for access to pronunciation and the enjoyment associated with familiar songs. Their positive orientation was conditional rather than uniform. Fast delivery, unfamiliar vocabulary, similar-sounding words, and difficulty recognizing spoken forms of known written words reduced comprehensibility. The findings therefore support a contextual conclusion: for these participants, the perceived usefulness of sound-based media depended on the interaction among audio difficulty, existing vocabulary knowledge, familiarity, and teacher scaffolding. They do not demonstrate improvement in vocabulary mastery.

Teachers can respond by selecting level-appropriate recordings, pre-teaching essential vocabulary, using a first listening for general meaning and a second for target words, replaying difficult segments, providing lyrics or written key words, and concluding with pronunciation and sentence practice. These recommendations are grounded in the reported experiences of the five participants and should not be generalized to all junior high school learners.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

6. References

Ardiansyah, M. R., Sodiq, J., & Rossana. (2025). The students' perspective in studying report text through British Council listening audio. ELite Journal: International Journal of Education, Language, and Literature, 4(2), 71–80. https://doi.org/10.26740/elitejournal.v4n2.p71-80

Bagila, S., Kok, A., Zhumabaeva, A., Suleimenova, Z., Riskulbekova, A., & Elmira, U. (2019). Teaching primary school pupils through audio-visual means. International Journal of Emerging Technologies in Learning, 14(22), 122–140. https://doi.org/10.3991/ijet.v14i22.11760

Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). SAGE Publications.

Crismeilani, K., Melani, B. Z., Saputra, A., & Amin, M. (2025). Strategies in teaching vocabulary for the 8th-grade students: A case study at SMP Dwijendra Mataram in academic year 2024/2025. Journal of Classroom Action Research, 7(3). https://doi.org/10.29303/jcar.v7i3.11502

Dakhi, S., & Fitria, T. N. (2019). The principles and the teaching of English vocabulary: A review. JET (Journal of English Teaching), 5(1), 15. https://doi.org/10.33541/jet.v5i1.956

Hartono, R. (2013). The use of audiovisual media to increase the students' vocabulary: A case of the tenth grade students of SMA N 1 Cepiring Kendal. ETERNAL (English Teaching Journal), 4(1), 91–102.

Lhamo, J., & Sakulwongs, N. (2023). The effectiveness of audio-assisted reading to enhance English reading comprehension skills for Bhutanese students. THAITESOL Journal, 36(2), 20–38.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). Qualitative data analysis: A methods sourcebook (3rd ed.). SAGE Publications.

Nation, I. S. P. (2013). Learning vocabulary in another language (2nd ed.). Cambridge University Press.

Nguyen, T. C. D. (2022). The impact of context on EFL learners' vocabulary retention. European Journal of Foreign Language Teaching, 6(2), 23–60. https://doi.org/10.46827/ejfl.v6i2.4295

Rochma, A., & Mukminatus, Z. (2025). English kids' songs: Effective media for teaching vocabulary. Discovery: Jurnal Ilmu Pengetahuan, 10(1), 77–88. https://doi.org/10.33752/discovery.v10i1.8767

Simonnet, E., Loiseau, M., & Lavoué, É. (2024). A systematic literature review of technology-assisted vocabulary learning. Journal of Computer Assisted Learning. https://doi.org/10.1111/jcal.13096

Wan Ji, P., & Abdul Aziz, A. (2021). A systematic review of vocabulary learning with mobile-assisted learning platforms. International Journal of Academic Research in Business and Social Sciences, 11(11). https://doi.org/10.6007/IJARBSS/v11-i11/11383.