Volume 4 Number 2 (2026)
July – December 2026
Page: 1351-1361

EDUJAVARE: International Journal of Educational Research
ISSN: 3031-2884
https://edujavare.com/index.php/edujavare
DOI: 10.70610/edujavare.v4i02.1815

Students' Perceptions of Clash A Rama Videos for Listening and Vocabulary Learning

Farid Hidayatul Arofi[1], Syifa' Khuriyatuz Zahro[2], Buyun Khulel[3]

[1] Universitas Islam Darul 'ulum Lamongan, Indonesia; farid.2022@mhs.unisda.ac.id
[2] Universitas Islam Darul 'ulum Lamongan, Indonesia; syifazahro@unisda.ac.id
[3] Universitas Islam Darul 'ulum Lamongan, Indonesia; buyunkhulel@unisda.ac.id

* 085815031161

Article history Submitted: 2026/06/29; Revised: 2026/07/07; Accepted: 2026/08/19

Abstract

Listening and vocabulary are mutually dependent in English as a Foreign Language (EFL) learning, yet students often struggle with rapid speech and unfamiliar words when their exposure to English is limited. Animated YouTube videos can combine spoken language, images, actions, captions, and narrative context, but entertainment-oriented content may also overload learners. The purpose of this study was to examine students' perceptions of Clash-A-Rama animated videos in listening and vocabulary learning and to identify the benefits, barriers, and instructional support they considered necessary. A qualitative descriptive design supported by descriptive questionnaire data was used with 26 Grade X students at an Indonesian senior high school. Data were collected through a 15-item Likert questionnaire and ten open-ended interview questions. Questionnaire data were analysed using frequencies, percentages, means, and an internal-consistency check, while interview data were analysed through Braun and Clarke's thematic analysis. The findings showed conditionally positive listening perceptions: 61.54% reported easier conversational comprehension, whereas 73.08% struggled with speech rate and 76.92% needed subtitles. Vocabulary perceptions were stronger: 88.46% reported learning new words and 92.31% found pictures and situations helpful, although 53.85% still encountered difficult vocabulary. Interview themes showed that multimodal cues, narrative context, and enjoyment supported learning, while fast dialogue, uneven proficiency, fragile recall, and pronunciation difficulties limited outcomes. Clash-A-Rama is therefore most useful as supplementary material when teachers segment clips, sequence captions strategically, pre-teach essential vocabulary, and provide repetition, pronunciation, and retrieval practice.

Keywords Animated Videos; EFL Learning; Listening; Students' Perceptions; Vocabulary

INTRODUCTION

Listening and vocabulary are interdependent components of English as a Foreign Language (EFL) learning. To understand spoken language, learners must segment a continuous

speech stream, recognise spoken forms, access word meanings, and integrate them with context in real time. Limited lexical knowledge can interrupt this process, while repeated listening encounters can expand vocabulary. Context-supported listening materials are therefore especially important in EFL settings, where exposure to English outside the classroom is often limited (Field, 2008; Nation, 2013; Vandergrift & Goh, 2012) In Indonesia, carefully developed YouTube-based materials have also been evaluated as acceptable and applicable for language instruction (Zahro & Tasaufy, 2019).

However, the literature also shows that audiovisual attractiveness does not automatically ensure linguistic accessibility. Fast or lexically dense dialogue may overload working memory when learners must listen, read captions, follow visual action, and interpret a story simultaneously(Sweller, 1988). Captions can connect sound, spelling, and meaning, but strong dependence on them may indicate limited independent auditory decoding. Likewise, visual context may support an initial meaning guess without guaranteeing accurate pronunciation, durable retention, or productive vocabulary use(Peters et al., 2016; Webb & Nation, 2017).

Clash-A-Rama is an entertainment-oriented YouTube animated series whose humorous narratives, expressive characters, action-based scenes, and rapid dialogue may create different experiences across listening and vocabulary learning. Previous studies have mainly examined achievement outcomes or general attitudes toward YouTube and animation. They have rarely investigated how learners distinguish the benefits and barriers of one specific popular series across these two related language domains. This is the explicit research gap addressed in the present study.

The study therefore examined two questions: (1) How do students perceive the use of Clash-A-Rama YouTube animated videos in listening lessons? and (2) What are students' perceptions of the videos in vocabulary learning? Its purpose was to identify perceived benefits, barriers, and required instructional support rather than merely determine whether students liked the medium. The study's novelty lies in integrating item-level questionnaire distributions with interview themes to explain why the same animated material may be experienced differently in listening and vocabulary learning.

METHODS

The study used a qualitative descriptive design supported by quantitative descriptive questionnaire data. The operational constructs were (a) perceived listening support and barriers, (b) perceived vocabulary support and barriers, and (c) supporting perceptions involving visual assistance, language suitability, enjoyment, participation, and willingness to reuse the medium. This design was selected to describe students' experiences and interpretations rather than to test causal effects or relationships between variables.

The participants were 26 Grade X SMA Wahid Hasyim Model. Purposive sampling was applied because all participants had taken part in English learning activities using a selected Clash A Rama video. The same 26 students completed the questionnaire and interview

questions. To protect confidentiality, interview responses were anonymised as R1 to R26. The available dataset did not include measured proficiency or detailed demographic variables; therefore, references to proficiency differences were interpreted cautiously and based on students' reported experiences.

Two instruments were used. The questionnaire contained 15 statements rated on a five-point Likert scale from Strongly Disagree to Strongly Agree: six listening items, three vocabulary items, and six supporting-perception items. Four barrier-oriented statements were reverse-coded for the internal-consistency check. The overall Cronbach's alpha was .69, indicating adequate exploratory consistency for descriptive use. The interview protocol contained ten open-ended questions addressing comprehension, helpful features, listening barriers, visual and narrative support, recalled vocabulary, engagement, challenges, and suggestions. Before data collection, both instruments were reviewed for relevance, clarity, and alignment with the research questions.

The learning procedure included an introduction to the topic, viewing the animated video, and short follow up activities. After the lesson, students completed the questionnaire and responded to the interview questions. Questionnaire data were analysed using frequencies, percentages, and mean scores. Negatively worded statements—such as difficulty with speech rate, subtitle dependence, and unfamiliar vocabulary—were interpreted as barriers rather than positive perceptions.

Interview data were analysed using the six phases of thematic analysis: familiarisation, initial coding, construction of candidate themes, theme review, definition and naming of themes, and report production (Braun, 2022). Listening related and vocabulary related codes were developed separately, while enjoyment and engagement were treated as cross cutting themes. Because one response could receive several codes, theme frequencies were overlapping and descriptive rather than mutually exclusive.

Finally, questionnaire and interview findings were integrated through convergence, complementarity, and the retention of negative or mixed cases. The conceptual framework in Figure 1 shows how multimodal input was expected to shape listening and vocabulary perceptions, with engagement, proficiency, cognitive load, and teacher scaffolding influencing the resulting learning experience.

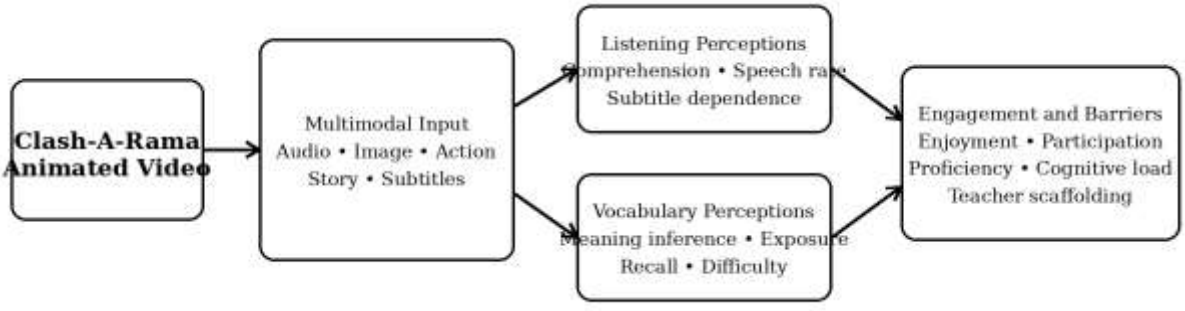

Figure 1. Conceptual framework of the study

RESULTS AND DISCUSSION

Results

All 26 students completed every questionnaire item and the interview questions; therefore, no missing value treatment was required. The quantitative and qualitative datasets referred to the same participant group, allowing direct cross analysis of response distributions and explanatory themes.

Table 1. Research data and analytical procedures

Data source	Scope	Analysis	Purpose
Questionnaire	15 Likert items; n = 26	Frequency, percentage, and mean	Show the distribution and strength of perceptions
Interview	10 open ended questions; n = 26	Inductive semantic thematic analysis	Explain mechanisms, barriers, and suggested support
Integrated analysis	Questionnaire–interview comparison	Convergence, complementarity, and mixed cases	Produce a balanced answer to both research questions

The listening results were positive but conditional. A total of 61.54% of students agreed that the video helped them understand English conversations more easily, and the same proportion perceived improvement in listening ability. In addition, 69.23% considered the audio clear. However, 73.08% reported difficulty following the speed of the conversations, 76.92% needed subtitles, and 57.69% experienced difficulty understanding some parts. These results indicate that students often accessed the general meaning but did not consistently decode every utterance.

Vocabulary perceptions were stronger and more consistent. Most students (88.46%) reported that the video helped them learn new vocabulary, and 92.31% stated that pictures and situations helped them understand word meanings. At the same time, 53.85% still found some vocabulary difficult. The video therefore functioned effectively as a source of contextualised lexical exposure, but it did not remove the need for explanation, repetition, pronunciation support, and review.

The animated format also produced substantial affective value. Visual elements supported comprehension for 80.77% of the students, 84.62% found the lesson enjoyable, 69.23% felt more active, and 73.08% wanted animated videos to be used in future lessons. Nevertheless, only 38.46% clearly agreed that the language matched their ability level, while 61.54% were neutral, suggesting that entertainment value and linguistic suitability were not identical.

Table 2. Key questionnaire findings

Domain	Indicator	Response	Interpretation
Listening	Easier conversation comprehension	61.54% positive	Generally helpful, but benefit was not universal
Listening	Difficulty with speech rate	73.08% agreement	Fast dialogue was the clearest listening barrier
Listening	Need for subtitles	76.92% agreement	Captions served as essential linguistic scaffolding
Vocabulary	Learning new vocabulary	88.46% positive	Strong perceived value for lexical exposure
Vocabulary	Pictures and situations clarified meaning	92.31% positive	Visual narrative context was the strongest benefit
Vocabulary	Some vocabulary remained difficult	53.85% agreement	Contextual inference did not ensure full mastery
Engagement	Enjoyable learning experience	84.62% positive	Animation supported emotional engagement
Engagement	Willingness to use videos again	73.08% positive	Most students accepted continued use

The thematic analysis produced four listening themes, four vocabulary themes, and two cross cutting engagement themes. Listening was supported by scenes, gestures, subtitles, clear audio, and pronunciation models, but the gains were often partial and dependent on learner proficiency. Fast speech and unfamiliar language created the greatest barriers, leading students to request slower playback, repetition, translation, shorter clips, and clearer presentation. Vocabulary learning was supported by visual and narrative inference and by exposure to memorable words or phrases. However, recall was selective, pronunciation remained difficult for some students, and explicit lexical support was still needed. Enjoyment increased attention and willingness to participate, but it did not guarantee comprehension or retention.

Table 3. Final themes from the interview analysis

Domain	Theme	Central meaning	Integrated interpretation
Listening	Multimodal scaffolding	Scenes, movement, subtitles, audio, and context supported meaning	Positive when several modes worked together
Listening	Partial and proficiency dependent gains	Students often understood only some words, scenes, or general meaning	Benefit varied across learners and support conditions
Listening	Speech rate and linguistic load	Fast dialogue and unfamiliar English interrupted real time processing	Attractiveness did not guarantee accessibility
Listening	Adaptive support required	Students requested replay, slower speed, translation, and shorter segments	Teacher mediation was necessary
Vocabulary	Visual and narrative inference	Pictures, action, situations, and captions supported meaning guesses	Strong contextual benefit
Vocabulary	Exposure and selective recall	Students encountered and recalled several words and expressions	Evidence of perceived learning, not comprehensive mastery
Vocabulary	Uneven and fragile gains	Some students showed weak recall or pronunciation difficulty	Positive perception coexisted with shallow knowledge
Vocabulary	Explicit lexical support	Glosses, meanings, pronunciation, and practice were requested	Incidental exposure should be followed by deliberate learning
Engagement	Enjoyment supports participation	Fun and reduced boredom encouraged attention and willingness	Affective value supported, but did not replace, instruction

Discussion

The listening findings are best interpreted as an effect of multimodal scaffolding rather than as a simple effect of video exposure. Students combined speech with character movement, facial expression, objects, scene changes, story events, and captions. In multimedia and dual-coding terms, these coordinated cues supplied alternative routes to meaning when the auditory signal alone was insufficient(Mayer, 2021; Paivio, 1990). The animation therefore functioned as a comprehension scaffold, not merely as visual decoration.

This scaffold mainly supported gist and situational understanding, not necessarily detailed decoding of spoken forms. Spoken input is transient, so learners must identify word boundaries and retrieve meanings before the signal disappears (Field, 2008). A student may correctly infer what is happening from an action sequence while still failing to recognise the exact words. This explains why positive perceptions of the storyline coexisted with only partial listening gains and why perceived access to a scene should not be equated with complete listening competence.

Speech rate was the clearest barrier because technical audibility and linguistic comprehensibility are different. Even clear audio remains difficult when learners must segment rapid speech, retrieve unfamiliar vocabulary, read captions, and follow visual action at the same time. Cognitive load theory suggests that these competing demands can exceed limited working-memory capacity (Sweller, 1988). The result extends previous reports of listening benefits from YouTube and animation by showing that such benefits depend on pacing, linguistic load, and opportunities to monitor comprehension (Al Hakim, 2024; Chien et al., 2020; Khulel, 2021)

Subtitles operated both as scaffolding and as a diagnostic sign. They helped students connect sound, spelling, and meaning, consistent with caption research (Montero Perez et al., 2013; Winke et al., 2010). At the same time, strong subtitle dependence indicated that independent auditory decoding was still developing. A staged sequence is therefore preferable: initial prediction with no or minimal captions, verification with English captions, focused replay of difficult segments, and a final viewing with reduced caption support. Such sequencing preserves the listening objective while gradually lowering cognitive assistance.

Discussion of Vocabulary Perceptions

Vocabulary was the strongest perceived contribution because visual events made word-meaning relationships salient. Objects supported concrete nouns, actions supported verbs, and facial expressions or narrative consequences supported descriptive and affective meanings. This process is consistent with multimedia learning and research showing that coordinated spoken forms, written forms, and relevant images increase opportunities for noticing and form-meaning mapping (Kabooha & Elyas, 2018; Mayer, 2021; Peters et al., 2016). This interpretation is consistent with Rindi et al., (2025) who found that structured use of popular audio media supported listening performance, vocabulary enrichment, motivation, and teacher-supported learning.

Nevertheless, noticing a word is not equivalent to knowing it. The selective recall found in the interviews suggests that visually prominent or narratively important items attracted more attention than less salient language. Word knowledge also includes pronunciation, spelling, meaning, collocation, grammatical behaviour, and receptive and productive use (Nation, 2013; Webb & Nation, 2017). Students could therefore infer a general meaning from a scene while remaining unable to pronounce the word accurately or use it independently.

The coexistence of positive vocabulary perceptions and reports of difficult vocabulary is therefore not contradictory. Encountering unfamiliar words creates an opportunity for lexical growth, but incidental exposure usually produces uneven and fragile knowledge. The instructional response should be selective rather than translation-heavy: teachers can allow an initial contextual inference, confirm meaning with a concise gloss, model pronunciation and stress, and then use retrieval, matching, sentence construction, and spaced review. This sequence retains the contextual advantages of the video while strengthening accuracy and retention.

Engagement, Integration, and Pedagogical Implications

Enjoyment and participation supported both learning domains by sustaining attention and willingness to continue. This affective value is consistent with research on engagement and digital storytelling (Akdamar & Sütçü, 2021; Fredricks et al., 2004; Kabooha & Elyas, 2018) However, enjoyment did not guarantee comprehension. Neutral responses concerning language suitability show that a visually attractive lesson can remain linguistically demanding, so motivation should be treated as a facilitating condition rather than evidence of learning by itself. A comparable perception pattern was reported by Hapsari, (2023), whose participants evaluated digital learning media in terms of usefulness, ease, and suitability.

The integrated analysis explains why listening and vocabulary perceptions diverged. Visual context can support an individual word through a relatively direct link to an object or action, whereas listening requires rapid processing across a continuous sequence of sounds, words, and sentences. Interviews also clarified apparently mixed questionnaire patterns: clear audio could coexist with poor comprehension, and successful contextual guessing could coexist with weak recall. These findings argue against an uncritical endorsement of technology; students valued the medium, but its effectiveness depended on how cognitive and linguistic demands were managed.

The practical implication is that entertainment content should be pedagogically transformed rather than played as an uninterrupted video. Teachers should select short and age appropriate segments, activate background knowledge, pre teach only essential words, adjust playback temporarily, use captions strategically, pause for prediction and verification, and add post viewing retrieval and production tasks. The language goal must determine whether English captions, selective Indonesian glosses, or temporary translation support is appropriate. This emphasis on authentic, interest aligned, and technology supported listening

resources is consistent with (Irmayani et al., 2023), who highlighted their role in supporting self directed extensive listening.

Table 4. Recommended instructional sequence

Stage	Listening focus	Vocabulary focus	Recommended support
Before viewing	Activate topic and predict content	Introduce only essential target words	Brief context, key images, limited glosses
First viewing	Listen for general meaning	Notice words linked to salient scenes	Short segment; no captions or minimal support
Second viewing	Verify details and identify word boundaries	Connect sound, spelling, and meaning	English captions, pausing, focused replay
After viewing	Reflect on breakdowns and retell	Practise pronunciation, retrieval, and use	Teacher explanation, matching, sentences, spaced review
Final viewing	Listen more independently at normal speed	Recognise target words in context	Reduce captions and translation gradually

Several limitations define the scope of interpretation. The study involved only 26 students from one school and did not include measured proficiency or detailed demographic data. It examined perceptions rather than achievement, and some interview responses were brief. Findings may also vary across Clash A Rama episodes, subtitle conditions, and teacher procedures. Future research should combine perception data with listening tests, vocabulary recognition and recall measures, delayed assessment, and comparisons of captioning or playback sequences.

CONCLUSION

Students perceived Clash-A-Rama animated videos positively in both listening and vocabulary learning, but the benefits were conditional. Visual scenes, movement, audio, captions, and narrative context supported general listening comprehension, while rapid dialogue, unfamiliar language, partial decoding, and subtitle dependence limited independent understanding. Vocabulary received the strongest positive response because pictures and situations helped students infer meanings; however, uneven recall and pronunciation difficulties showed that contextual exposure alone did not produce complete word knowledge.

Clash-A-Rama should therefore be used as supplementary EFL material within structured instruction: short segmented viewing, strategic caption sequencing, controlled replay, selective glossing, pronunciation modelling, retrieval, and productive practice. Because the study examined perceptions in one small group rather than measured learning outcomes, future research should test these procedures through listening and vocabulary assessments, including pre-test, post-test, and delayed-test designs.

REFERENCES

Akdamar, N. S., & Sütçü, S. S. (2021). Effects of digital stories on the development of EFL learners' listening skill. *Education Quarterly Reviews*, *4*(4), 271–279. https://doi.org/10.31014/aior.1993.04.04.391

Al Hakim, M. I. I. (2024). Using animation video to enhance EFL students' listening comprehension. *Retain: Journal of Research in English Language Teaching*, *12*(1), 24–29. https://doi.org/10.26740/rt.v12i01.59071

Braun, V. (2022). Thematic analysis: a practical guide/Virginia Braun and Victoria Clarke. *Cham: SAGE Publications Ltd*.

Chien, C.-C., Huang, Y., & Huang, P. (2020). YouTube videos on EFL college students' listening comprehension. *English Language Teaching*, *13*(6), 96–103. https://doi.org/10.5539/elt.v13n6p96

Field, J. (2008). *Listening in the language classroom*. Cambridge University Press. https://doi.org/10.1017/CBO9780511575945

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Hapsari, A. D. (2023). Students' perception toward using Canva in EFL business correspondence class. *Edulitics (Education, Literature, and Linguistics) Journal*, *8*(2), 47–56. https://doi.org/10.52166/edulitics.v8i2.5268

Irmayani, I., Rozak, R. R., Amrullah, A. Z., & Maslakhatin, M. (2023). Extensive listening: Indonesian teacher educators' and student teachers' perspectives and experiences in initial teacher education context. *Edulitics (Education, Literature, and Linguistics) Journal*, *8*(2), 38–46. https://doi.org/10.52166/edulitics.v8i2.5414

Kabooha, R., & Elyas, T. (2018). The effects of YouTube in multimedia instruction for vocabulary learning: Perceptions of EFL students and teachers. *English Language Teaching*, *11*(2), 72–81. https://doi.org/10.5539/elt.v11n2p72

Khulel, B. (2021). Metacognitive strategies instruction and its relationship with listening anxiety and listening comprehension. *Edulitics (Education, Literature, and Linguistics) Journal*, *6*(1), 40–46. https://doi.org/10.52166/edulitics.v6i1.1914

Mayer, R. E. (2021). *Multimedia learning*. Cambridge University Press.

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*(3), 720–739. https://doi.org/10.1016/j.system.2013.07.013

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

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, *63*, 134–148.

https://doi.org/10.1016/j.system.2016.10.002

Rindi, R. S., Sofeny, D., & Huda, K. (2025). The effect of Taylor Swift songs on students' listening skills in vocabulary learning. *Bliss Journal: Broadening Linguistics, Literature, Education, and Study of Learning Media*, *1*(2), 57–71. https://e-jurnal.unisda.ac.id/index.php/BlissJournal/article/view/10903

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge. https://doi.org/10.4324/9780203843376

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*(1), 65–86. https://doi.org/10.64152/10125/44203

Zahro, S. K., & Tasaufy, F. S. (2019). The development of educational YouTube videos-based instructional material for speaking for beginner course. *IJET (Indonesian Journal of English Teaching)*, *8*(2), 48–57. https://doi.org/10.15642/ijet2.2019.8.2.48-57