

TUTORS' PERCEPTIONS OF USING YOUTUBE CLIPS FOR VOCABULARY LEARNING

Yana Setia Maulana

English Education, FKIP, Universitas Galuh, Ciamis, Indonesia

Email: yana_setia@student.unigal.ac.id

Lilies Youlia Friatin

English Education, FKIP, Universitas Galuh, Ciamis, Indonesia

Email: liliesyouliafriatin@unigal.ac.id

Leni Irianti

English Education, FKIP, Universitas Galuh, Ciamis, Indonesia

Email: leni-englishedu@unigal.ac.id

APA Citation: Maulana, Y. S., Friatin, L. Y., & Irianti, L. (2026). Tutors' perceptions of using YouTube clips for vocabulary learning. *Journal of English Education Program (JEEP)*, 9(2), 1-9. doi: -xx.xxxxx/jeep.xxx.xxxx.

Received: xx-6-2026

Accepted: xx-7-2026

Published: xx-7-2026

Abstract: This study examined how English tutors used YouTube clips to support vocabulary learning among young learners at Bumi Neng Mom Community and how the tutors evaluated the medium in their teaching practice. The study responded to the growing use of digital video in English instruction and the limited attention given to tutor practices in non-formal learning environments. A qualitative case study was conducted with two English tutors who regularly taught children and incorporated YouTube English clips into their lessons. Data were gathered through classroom observations and semi-structured interviews, then organised through descriptive and thematic analysis. The findings indicate that the tutors integrated YouTube through purposeful clip selection, pre-viewing preparation, guided viewing, post-viewing reinforcement, and classroom management. Narrative-based and song-based clips were used in different ways, with the former supporting contextual interpretation and the latter encouraging repetition and recall. The tutors perceived the clips as useful for pronunciation, vocabulary memorisation, participation, and classroom engagement. Nevertheless, learners could become more attentive to animation, music, rhythm, or lyrics than to vocabulary meaning. The study concludes that YouTube is most useful as supplementary media when tutors provide explicit guidance, repetition, explanation, contextual practice, and opportunities for later vocabulary recycling.

Keywords: *YouTube clips; tutors' perceptions; vocabulary learning; young learners; digital media*

INTRODUCTION

The continuing expansion of digital technology has changed the resources available for English language instruction. Digital platforms now allow learners to encounter language through combinations of sound, images, movement, written language, and contextual situations. For young learners, these forms of input can be particularly useful because language is often easier to recognise when it is connected to something visible, audible, or familiar. YouTube is one of the platforms that can provide such input through songs, stories, animations, and short educational videos. Its accessibility also makes it practical for tutors working in settings where instructional materials may be flexible rather than fixed to a particular textbook. Vocabulary occupies a central

position in the development of English proficiency because learners need words to understand messages and to express meaning in listening, speaking, reading, and writing. Limited vocabulary can restrict learners' ability to participate in classroom interaction and to understand English input. Vocabulary knowledge also involves more than recognising a word. Learners gradually need to understand its form, meaning, and use in context. Nation (2017) emphasises the importance of developing vocabulary knowledge through repeated encounters and meaningful use. This principle is especially relevant to young learners, whose vocabulary development is still progressing from recognition towards more productive use.

For children, introducing vocabulary through isolated lists may not provide sufficient support for understanding and remembering new words. Young learners often respond more readily to concrete examples, repetition, visual cues, sounds, gestures, and opportunities to participate. YouTube English clips can combine several of these features in a single learning activity. A song, for example, may expose learners to the same expression several times through rhythm and melody, whereas a story can place unfamiliar words within actions and events. Such contexts can help learners associate a word with an observable situation rather than encountering the word without a meaningful reference.

The potential of digital media in language learning is also related to multimodal learning. Mayer (2014) explains that learning can benefit when relevant information is organised across verbal and visual forms and when learners are guided towards the information that matters. In vocabulary instruction, audiovisual material can therefore provide more than entertainment. Images may clarify meaning, spoken language may provide a pronunciation model, and repeated expressions may create opportunities for recall. However, these benefits depend on the relevance of the material and on the instructional support provided while learners interact with it. Previous research has reported positive possibilities for YouTube and video-based vocabulary instruction. Al-Ameri and Rababah (2020) reported that YouTube-supported instruction could contribute to vocabulary development among elementary learners. Hariyono (2020) also described the use of YouTube video in teaching vocabulary to young learners, while Rini et al. (2024) reported benefits from YouTube-assisted vocabulary learning in EFL classrooms. More recent work has continued to discuss video as a useful source of language exposure, including systematic reviews concerning YouTube and vocabulary acquisition (Ardianti & Chabibah, 2025; Tohamba, 2025). Together, these studies indicate that audiovisual media can provide useful conditions for vocabulary learning, particularly when learners receive repeated and comprehensible exposure.

At the same time, using YouTube in an English lesson does not automatically produce vocabulary learning. Videos contain visual and auditory features that can attract children's attention, but those same features may compete with the instructional objective. Learners may remember a melody, animation, character, or lyric without developing a clear understanding of the vocabulary presented in the material. This issue makes tutor mediation important. Tutors need to decide which clip is appropriate, identify the vocabulary that deserves attention, prepare learners before viewing, intervene during viewing when necessary, and provide activities that allow learners to revisit the target words afterwards.

The distinction between viewing and learning becomes particularly important in English for Young Learners (EYL). Children in this age group generally benefit from activities that are concrete, interactive, and suitable for their developmental level. Younas and Dong (2024) discuss the relevance of animated video and multimedia learning for young learners, while Hasanah et al. (2024) highlight the use of video-based vocabulary instruction for young EFL learners. These perspectives suggest that the value of video lies partly in how its visual and auditory resources are connected to purposeful classroom interaction. Therefore, the tutor remains responsible for turning exposure into an instructional experience. In Indonesia, opportunities for English exposure can differ substantially across learning environments. Non-formal learning communities can provide additional opportunities for children to encounter English outside regular school lessons. Bumi Neng Mom Community in Padaherang, West Java, is one such setting. English activities are

provided in a flexible community-based environment, and tutors have incorporated YouTube English clips as supplementary material. The learners are primarily children between six and twelve years old. Within this setting, YouTube offers tutors access to songs, stories, animations, and other materials that can be selected according to the vocabulary focus of a lesson.

Although research has examined the relationship between YouTube and vocabulary learning, much of the discussion has concentrated on learner outcomes, motivation, or vocabulary achievement. Less attention has been given to the practical decisions tutors make when using clips in a real non-formal learning environment. It is important to examine what tutors actually do before, during, and after a video is shown because the same clip may produce different learning experiences depending on how it is introduced and followed up. The tutor's perception also matters because it reflects how the medium is evaluated in relation to classroom realities.

The present study therefore focuses on tutors' practices and perceptions rather than attempting to measure vocabulary achievement statistically. It examines how two English tutors at Bumi Neng Mom Community select and use YouTube English clips when introducing vocabulary to young learners. Attention is given to preparation before viewing, guidance during viewing, reinforcement after viewing, and classroom management. The study also explores the tutors' views of the benefits, limitations, learner responses, motivational value, and overall effectiveness of YouTube English clips.

The study addresses two research questions: first, how do the tutors use English clips in introducing vocabulary to young learners; and second, what are the tutors' perceptions towards the use of YouTube English clips? In accordance with these questions, the study aims to describe the tutors' instructional practices and to examine how they interpret the contribution and limitations of YouTube clips in vocabulary learning. The focus on a small non-formal learning community provides a specific contextual perspective that may not be visible in studies based only on tests or general measures of learner achievement. The significance of the study is both practical and academic. Practically, the findings can help tutors and teachers understand how YouTube clips may be integrated into structured vocabulary instruction rather than used simply as a reward or entertainment activity. Academically, the study contributes a qualitative account of tutor mediation in a non-formal EYL setting. By examining the connection between digital material, tutor decisions, and learner responses, the study reinforces the idea that technology is one component of a wider teaching process. Its contribution is therefore located not only in the use of YouTube itself, but in the instructional practices surrounding that use.

METHOD

This research employed a qualitative case study design. The design was selected because the study sought to understand a specific instructional phenomenon within its natural setting, namely the use of YouTube English clips by tutors when introducing vocabulary to young learners. A qualitative approach was appropriate because the study focused on descriptions of classroom practice and tutors' interpretations rather than numerical measurement of vocabulary achievement. Creswell and Poth (2018) describe qualitative inquiry as suitable for exploring experiences and meanings within particular contexts. The case study perspective consequently allowed the researcher to examine the phenomenon in relation to the actual learning environment at Bumi Neng Mom Community. The research took place at Bumi Neng Mom Community in Padaherang, West Java. The community functions as a non-formal learning centre and provides English learning activities for children. The learners involved in the observed setting were primarily six to twelve years old, corresponding to the EYL context used in this study. The site was selected because YouTube English clips had already been incorporated into the tutors' instructional practices. This made the community a relevant setting for examining how digital video was selected, presented, discussed, and reinforced during vocabulary lessons.

Two English tutors participated in the study. Both regularly taught young learners and had direct experience using YouTube English clips in their lessons. Purposive selection was used

because the participants needed to possess experience relevant to the research questions. The number of eligible tutors in the community was limited, and both tutors were included so that the study could represent the practices available within this particular setting. The study therefore does not claim statistical generalisation. Instead, it provides an in-depth description of the practices and perceptions found among the participating tutors.

Two forms of data were collected: classroom observations and semi-structured interviews. The use of two sources allowed the researcher to examine both what tutors did and how they explained their decisions. Classroom observations provided direct evidence of the instructional sequence, tutor intervention, learner responses, and technical or classroom management. Semi-structured interviews provided opportunities for the tutors to explain their views about video selection, instructional strategies, advantages, challenges, learner engagement, motivation, and the effectiveness of YouTube clips. Using both sources also made it possible to compare reported perceptions with observed classroom practice.

The observation process focused on several stages of YouTube-supported vocabulary instruction. These stages included planning and clip selection, pre-viewing preparation, while-viewing guidance, post-viewing reinforcement, and classroom and technical management. An observation sheet and field notes were used to document the tutors' actions, interaction patterns, vocabulary reinforcement, and learner responses. The researcher acted as a non-participant observer so that the regular classroom activities could continue with minimal interference. Permission was obtained before the observations were conducted.

Semi-structured interviews were conducted face-to-face after the classroom observation process. An interview guide was prepared to maintain consistency across the two participants while still allowing follow-up questions when clarification was needed. The questions covered the reasons for selecting YouTube clips, strategies used before and during viewing, reinforcement after viewing, perceived benefits, challenges, learner responses, motivational effects, and overall effectiveness. The participants gave consent for the interviews, and the responses were transcribed for analysis. This procedure allowed the researcher to connect individual statements with the practices recorded during observation.

The data were analysed descriptively and thematically. Observation notes were first organised according to the instructional stages and recurring practices were identified. Interview transcripts were read repeatedly to develop familiarity with the responses. Significant statements were then coded and grouped into broader themes related to the research questions. The themes were compared with the observation findings to identify patterns of agreement and difference. This comparison was important because the study was concerned not only with tutors' stated opinions but also with whether those opinions were reflected in their actual classroom practices.

Ethical considerations were maintained throughout the research. Permission was obtained from the relevant community and participants before data collection. The tutors were informed about the purpose of the research and the use of the data for academic purposes. Personal information that could unnecessarily identify the participants was not included in the research report. This precaution was particularly important because the study involved a small community in which participants could be readily recognised. The researcher treated the observation and interview data as research material and maintained confidentiality when presenting the findings.

RESULT

The findings were derived from observations of two English tutors and from their semi-structured interview responses. The observation data showed how YouTube English clips were incorporated into lessons, while the interviews clarified the tutors' reasoning and perceptions. Across the data, five broad areas were apparent: clip planning and selection, pre-viewing preparation, while-viewing guidance, post-viewing reinforcement, and classroom or technical management. The tutors did not treat YouTube as an independent teaching method. Instead, the clips were embedded within a sequence of tutor-led activities that included questioning, repetition,

visual support, writing, singing, and recall. The first observation involved a narrative-based fable clip. The tutor used the story as a context for introducing unfamiliar vocabulary and encouraged learners to infer meaning through questions connected to the events in the video. Vocabulary was therefore presented as part of a situation rather than as an isolated list. The tutor asked learners whether they recognised particular words and used the storyline to support interpretation. This practice encouraged learners to pay attention to meaning while following the narrative. The second observation used a song about shapes. In this lesson, the tutor connected the spoken and sung vocabulary with board drawings, gestures, and repeated oral production.

Tabel 1. Summary of Classroom Observation Findings

Observed Dimension	Key Observation	Interpretive Note
Clip planning and selection	The tutors chose YouTube clips according to the lesson topic, vocabulary focus, and learners' needs.	The story clip supported contextual interpretation, whereas the song clip emphasised repeated exposure and recall.
Pre-viewing preparation	Tutors introduced the topic, activated learners' prior knowledge, and used questions, drawings, or examples before viewing.	Preparation helped learners establish meaning before encountering the target vocabulary in the clip.
While-viewing guidance	Tutors directed attention through pauses, replay, repetition, questions, singing, gestures, and visual cues.	Pausing supported processing and clarification, while continuous song activities maintained rhythm and participation.
Post-viewing reinforcement	Learners revisited target words through oral repetition, writing, drawing, questioning, recall, and story retelling.	Immediate reinforcement supported recall, although systematic recycling across later meetings was limited.
Classroom and technical management	Both lessons progressed without major technical disruption, and tutors organised transitions between viewing and follow-up tasks.	Clear instructions and controlled transitions helped learners maintain attention and understand the lesson sequence.

Before viewing, both tutors prepared learners for the topic, although they did so in different ways. The tutor using the narrative clip activated prior knowledge through questions related to the expected content of the story. This encouraged learners to make predictions and connect the video with familiar ideas. The tutor using the song introduced the topic more directly by drawing shapes on the board and providing examples. These practices show that the pre-viewing stage was used to establish a reference point before the digital material was presented. Visuals and questions helped learners approach unfamiliar vocabulary with some contextual support rather than encountering the words without preparation. During viewing, tutor intervention remained visible. The narrative-based lesson used pauses and replay to give learners additional time to process vocabulary, answer questions, and confirm understanding. The tutor did not allow the video to determine the entire pace of the lesson. In contrast, the song-based lesson maintained the continuity and rhythm of the song and used singing and repetition to reinforce target words. The difference suggests that while-viewing strategies were adjusted to the characteristics of the clip. Pausing was useful when learners needed explanation or processing time, while continuous repetition was useful when rhythm was central to participation.

After viewing, both tutors provided reinforcement rather than ending the lesson when the clip finished. Learners were asked to repeat vocabulary and respond to questions. In the narrative lesson, follow-up activities included writing words, copying illustrations, recalling vocabulary, and retelling parts of the story. In the song lesson, repetition and singing were used to strengthen recognition and recall. These activities provided opportunities for learners to interact with the target vocabulary after the audiovisual input had ended. However, the observations showed limited evidence of systematic vocabulary recycling across later meetings, meaning that most reinforcement occurred within the same lesson. Classroom and technical management also

supported the implementation. The observed lessons proceeded without major technical problems involving the presentation of the clips. The tutors managed transitions between watching, questioning, repetition, board work, and writing. Clear directions helped learners understand what they were expected to do at different points in the lesson. This was important because young learners could lose focus when an activity changed without clear guidance. The observations therefore indicate that technical readiness alone was not sufficient. The tutors also needed to manage the instructional flow around the video.

Table 2. Summary of Interview Findings

Participant	Paraphrased Interview Point	Emerging Theme
Tutor 1	The tutor explained that direct listening through the clips gave learners a pronunciation model that they could imitate.	Audiovisual exposure supports pronunciation development.
Tutor 2	The tutor reported that songs made it easier for learners to remember vocabulary because important words occurred repeatedly.	Songs and repetition support vocabulary recall.
Tutor 1	The tutor observed that learners could pay greater attention to images and animation than to the vocabulary that the lesson was intended to teach.	Entertainment features can compete with vocabulary learning.
Tutor 2	The tutor noted that learners might remember the words in a song while still lacking a complete understanding of what those words meant.	Meaning requires explanation and follow-up practice.
Tutor 1	The tutor described learners as more cheerful and more willing to participate when video materials were incorporated.	Video can support engagement and classroom participation.
Tutor 2	The tutor considered YouTube useful but stressed that it produced better results when integrated with other instructional strategies.	YouTube is supplementary rather than a standalone method.

The interview data were broadly consistent with the observation findings. Both tutors considered YouTube English clips useful, particularly for pronunciation, vocabulary memorisation, and learner engagement. One tutor emphasised that learners could listen directly to pronunciation and imitate it, while the other highlighted the role of songs in helping learners remember words. Both tutors also described positive classroom responses, noting that learners appeared more cheerful, active, and willing to participate when videos or songs were incorporated. These perceptions corresponded with the observed use of repetition, singing, and interactive questioning. The tutors also identified clear limitations. They observed that learners could become more interested in animation, rhythm, lyrics, or other entertaining features than in vocabulary meaning. One tutor noted that learners might focus on the visual aspects of a clip, while the other explained that learners could remember lyrics without fully understanding their meaning. Both therefore considered additional explanation and reinforcement necessary. Their responses indicate that memorisation and pronunciation may occur through video exposure, but deeper understanding requires the tutor to connect the words with explanations, questions, visual support, and contextual practice.

The tutors differed somewhat in their views of motivation and confidence. One tutor considered the motivational effect of YouTube to be situational because entertainment could sometimes replace attention to the learning objective. The other perceived the videos as helping

create a more relaxed atmosphere in which learners were more comfortable participating. Despite this difference, both tutors agreed that the medium worked best when combined with other strategies. Their overall perception was that YouTube was useful as supplementary instructional media rather than as a complete method for vocabulary teaching.

A comparison of the two data sources showed a consistent relationship among the video, tutor mediation, and learner response. The tutors' statements about repetition and explanation were reflected in the observed lessons through repeated pronunciation, board illustrations, questioning, writing, and post-viewing activities. Similarly, the tutors' positive perceptions of engagement were supported by observations of active participation during songs and guided video activities. This convergence strengthens the interpretation that the educational value of the clips depended substantially on how the tutors structured and mediated the learning experience.

DISCUSSION

The findings indicate that YouTube English clips can contribute to vocabulary learning when they are incorporated into a planned instructional sequence. The clips provided audiovisual exposure, but the tutors did not rely on exposure alone. They selected materials according to lesson themes and vocabulary objectives, prepared learners before viewing, guided attention during viewing, and followed the clips with reinforcement. This pattern is compatible with research describing video and YouTube as potentially useful resources for vocabulary learning (Al-Ameri & Rababah, 2020; Hariyono, 2020; Rini et al., 2024). The present study adds a contextual explanation of how that usefulness was produced through tutor-led classroom activity. The planning and selection stage demonstrates that the choice of video is an instructional decision rather than a purely technical one. The tutors selected clips that corresponded to the topic and vocabulary focus of the lesson. The narrative clip and the song about shapes served different pedagogical purposes. The story created opportunities for learners to encounter vocabulary in context, whereas the song provided a structure for immediate repetition and recall. These differences suggest that the most appropriate clip depends on the intended learning process. A tutor seeking contextual interpretation may favour a narrative, while a tutor introducing basic vocabulary may benefit from a repetitive song.

Material suitability is particularly important for young learners. A clip can contain useful vocabulary but still be difficult if the language is too fast, the duration is excessive, or the visual information is not connected to the target words. The observations showed that tutors adjusted their use of clips when learners required additional support. Pausing and replaying the narrative video allowed learners more time to process information. The use of drawings and gestures in the song lesson provided an additional representation of meaning. These practices support the view that digital material should be adapted to learners' needs rather than treated as a fixed instructional package.

The multimodal nature of YouTube clips can explain part of their perceived usefulness. Learners encountered spoken forms through listening, visual representations through the video, and repeated language through songs or recurring expressions. Mayer (2014) argues that multimedia learning can support understanding when relevant verbal and visual information is coordinated. In this study, the coordination became more evident when tutors linked video content to gestures, board drawings, questions, and oral repetition. The finding suggests that multimodality was not valuable simply because several modes were present. Its value increased when the tutor helped learners establish meaningful connections among those modes. The pre-viewing stage functioned as a form of instructional preparation. By asking questions or presenting drawings before the clip, tutors gave learners an initial framework for interpreting the incoming language. Activating prior knowledge also created an opportunity for learners to make predictions about what they might see or hear. This is important for children because unfamiliar vocabulary can be difficult to process when it appears without contextual support. The findings therefore indicate

that pre-viewing should not be treated as a short introduction that can be skipped. It can prepare learners to notice the target vocabulary once the video begins.

The two tutors used different forms of pre-viewing support, and this difference illustrates the flexibility of video-based instruction. The narrative tutor encouraged contextual discovery, whereas the other tutor provided more direct visual explanation. Neither practice appeared to function as a universal model. Instead, each approach corresponded to the nature of the material and the tutor's teaching style. This supports a broader principle in technology-assisted language learning: the platform does not prescribe one pedagogy. Tutors still need to make decisions about how much explanation, prediction, modelling, and visual support learners require.

The while-viewing stage further demonstrates the importance of active mediation. The tutors paused, replayed, asked questions, repeated words, and used visual cues instead of allowing learners to watch without direction. This transformed the video from a presentation resource into part of an interactive lesson. The finding is consistent with the emphasis in multimedia learning on directing learners towards relevant information (Mayer, 2014). It also corresponds with studies reporting that YouTube and video-based activities can support vocabulary learning when accompanied by appropriate classroom strategies (Kaboocha & Elyas, 2018; Rini et al., 2024). Pausing was particularly relevant to the narrative clip because learners needed time to interpret words from the story. A pause created space for questioning, clarification, repetition, and checking comprehension. In the song lesson, however, maintaining the rhythm supported participation and repeated exposure. This contrast indicates that pause-and-play should not be applied mechanically. The tutor needs to consider the purpose of the clip and the point at which intervention will help rather than interrupt the learning activity. Effective mediation therefore involves both intervention and restraint.

Repetition emerged as a common mechanism across the two lessons. Repetition through songs provided learners with recurring encounters with the same words, while repetition through replay, oral production, and questioning provided additional opportunities to recall vocabulary. Nation (2017) identifies repeated encounters as important for vocabulary development, and the findings here show how YouTube can create such encounters within ordinary classroom activities. Nevertheless, repetition should not be equated with complete vocabulary mastery. A learner may pronounce or recall a word without knowing how to interpret or use it in a new context.

The distinction between memorisation and understanding was one of the clearest findings from the interviews. The tutors recognised that songs could help learners remember vocabulary, yet they also observed that remembering lyrics did not necessarily mean understanding the words. This distinction is pedagogically important. Vocabulary learning requires attention to meaning and use, not only sound or sequence. Consequently, song-based instruction can be strengthened when the tutor stops at appropriate points, explains key words, asks learners to connect words with pictures or actions, and provides a short activity that requires them to retrieve meaning rather than reproduce a lyric. Post-viewing reinforcement provided the link between exposure and active practice. Writing, retelling, questioning, drawing, singing, and recall required learners to do something with the vocabulary after watching. The narrative lesson was particularly useful in showing how vocabulary could move from recognition towards more productive use through retelling. The song lesson provided stronger emphasis on oral recall and pronunciation. These approaches can be complementary. A lesson may introduce words through a song and then extend them through pictures, simple questions, writing, or short contextual sentences. Such sequencing would broaden the kinds of vocabulary knowledge learners practise.

The limited evidence of vocabulary recycling across later meetings is an important limitation of the observed practice. Both tutors repeated words during the same lesson, but there was little evidence that the same vocabulary was deliberately revisited in subsequent sessions. Immediate recall does not guarantee durable retention. Nation (2017) stresses the importance of repeated encounters with vocabulary, and this principle suggests that the target words from YouTube activities should be reused over time. Short review games, picture matching, oral recall,

simple writing tasks, or later lessons using the same words could provide additional encounters without requiring the original video to be replayed in full.

The findings also have implications for classroom management. Young learners can shift attention quickly, particularly when a video contains highly attractive entertainment features. The tutors' management of transitions helped maintain the lesson sequence. Clear instructions about when to watch, when to answer, when to repeat, and when to write reduced uncertainty and kept the video connected to the lesson objective. Therefore, classroom management is not separate from digital pedagogy. It is part of the process through which a tutor controls attention and maintains the relationship between the digital resource and the intended vocabulary learning. Technical conditions were generally supportive in the observed setting, but this should not obscure the need for preparation. A video-based lesson depends on access to the selected clip, adequate sound, visible presentation, and a workable internet connection. When these conditions are stable, the tutor can concentrate on interaction rather than solving technical problems. When they are unstable, instructional time can be lost and learners can become distracted. For non-formal learning communities, this suggests that tutors should preview clips and prepare alternatives when possible. Technical readiness should be treated as part of lesson planning rather than as an issue considered only after the lesson begins.

The positive classroom atmosphere reported by the tutors also needs to be interpreted carefully. Engagement was visible through participation, enthusiasm, and willingness to sing or respond. However, engagement alone does not demonstrate vocabulary mastery. The tutors themselves recognised this distinction when they described learners who remembered lyrics but did not always understand their meanings. This finding supports a balanced interpretation of motivation. YouTube can make participation more attractive, but the instructional objective remains vocabulary development. The tutor must use the learners' interest as an entry point for attention, practice, explanation, and retrieval.

The differences between the tutors further show that YouTube-supported vocabulary teaching does not follow a single fixed procedure. One tutor emphasised contextual interpretation through narrative, while the other emphasised repetition through song and visual demonstration. These approaches reflect different ways of helping children encounter new vocabulary. A contextual approach may encourage learners to infer meaning from events and actions, whereas a repetitive approach may make recognition and pronunciation easier. The findings therefore support flexibility, provided that the selected strategy remains connected to the vocabulary objective and the learners' developmental needs. From the perspective of EYL pedagogy, the results reinforce the importance of concrete and interactive learning experiences. Children can benefit when new words are connected to things they can see, hear, imitate, draw, or act out. YouTube clips can provide a useful starting point for these experiences, but the tutor's actions determine whether the different resources are used meaningfully. This is consistent with literature on video-based vocabulary learning among young learners (Hasanah et al., 2024; Younas & Dong, 2024). The present study extends that perspective by showing the practical sequence through which tutors mediated the clips in a non-formal community.

The findings also align with the understanding of YouTube as supplementary rather than standalone media. Kabooha and Elyas (2018) and later studies have discussed the value of YouTube within multimedia instruction, while Ardianti and Chabibah (2025) highlight its potential in relation to motivation and vocabulary acquisition. In the present study, the tutors' own perceptions were conditional rather than absolute. They did not describe YouTube as sufficient by itself. Instead, they valued it when it was combined with questioning, explanation, repetition, writing, visual support, and other teaching strategies. This conditional view is important because it avoids treating the platform as the source of learning independent of pedagogy.

The study contributes a specific perspective by examining tutor practice in a non-formal learning community. The small number of participants limits the extent to which the findings can be transferred to other settings, but it also allows detailed attention to the instructional process.

The strongest pattern across the data is the interaction between media, tutor mediation, and learner response. The clips attracted attention and provided language input, while tutor intervention directed that attention towards vocabulary. This relationship indicates that digital media should be evaluated as part of an instructional system rather than as an isolated teaching technique.

For tutors working in similar contexts, the findings suggest a practical sequence. First, identify a clear vocabulary objective and select a clip that matches the learners' age and language level. Second, preview the clip and determine which words require emphasis. Third, prepare learners with questions, pictures, drawings, or simple explanations. Fourth, guide viewing through appropriate pausing, repetition, and comprehension checks. Fifth, provide a post-viewing task that requires learners to recall or use the target words. Finally, recycle those words in later lessons. This sequence maintains the strengths of YouTube while addressing the risk that children may remember the entertainment features more readily than the vocabulary meaning.

CONCLUSION

This study examined how two English tutors at Bumi Neng Mom Community used YouTube English clips to introduce vocabulary to young learners and how they perceived the value of the medium. The findings show that YouTube was integrated as supplementary instructional media rather than used as an independent teaching method. The tutors selected clips according to lesson themes and vocabulary objectives, prepared learners before viewing, guided their attention during the clips, and provided reinforcement afterwards. Although their strategies differed, both tutors connected video exposure with additional classroom activities.

The tutors viewed YouTube English clips positively because the medium provided useful pronunciation models, repeated exposure, opportunities for vocabulary memorisation, and an engaging classroom atmosphere. Songs were particularly useful for repetition and recall, while narrative material provided opportunities for learners to interpret vocabulary in context. At the same time, the tutors identified a recurring limitation: learners could become more interested in animation, rhythm, lyrics, or other entertainment features than in vocabulary meaning. The findings therefore distinguish between engagement and learning. A learner can participate enthusiastically without necessarily developing a complete understanding of the target words.

Overall, the study concludes that the effectiveness of YouTube English clips depends on tutor mediation and instructional structure. Video selection, pre-viewing preparation, active guidance, repetition, explanation, and post-viewing reinforcement all contribute to making audiovisual exposure meaningful. The findings also indicate a need for vocabulary recycling across subsequent lessons because repetition within a single session may mainly support immediate recall. For educators, YouTube is best understood as an accessible supplementary resource whose educational value increases when it is embedded within purposeful EYL pedagogy. Future research can extend this study by involving more tutors and learners across different formal and non-formal educational contexts. Longer observation periods could examine whether vocabulary introduced through YouTube is retained over time. Further studies could also compare different types of clips, such as songs, animated stories, and short instructional videos, in relation to particular vocabulary outcomes. Research using additional forms of evidence, including learner work or vocabulary tasks, could complement tutors' perceptions and provide a broader account of how video-based instruction affects vocabulary development.

ACKNOWLEDGEMENT

The researcher gratefully acknowledges the English Education Program, Faculty of Teacher Training and Education, Universitas Galuh, for its academic support. Appreciation is also extended to the supervisors, the tutors, and the founders of Bumi Neng Mom Community for

granting access to the research setting and providing cooperation throughout the observation and interview process.

REFERENCES

- Al-Ameri, A. M., & Rababah, L. (2020). The effect of using YouTube on developing elementary students vocabulary. *Education and Linguistics Research*, 6(1), 129. <https://doi.org/10.5296/elr.v6i1.16796>
- Alakrash, H. M., & Razak, N. A. (2021). Technology-based language learning: Investigation of digital technology and digital literacy. *Sustainability*, 13(21). <https://doi.org/10.3390/su132112304>
- Ardianti, A. P., & Chabibah, F. M. (2025). YouTube for enhancing EFL students' motivation in vocabulary acquisition: A systematic review. *English Language Teaching Journal*, 5(2), 162-187. <https://doi.org/10.35897/eltj.v5i2.1863>
- Cohen, L., Manion, L., & Morrison, K. (2017). *Research methods in education*. Routledge. <https://doi.org/10.4324/9781315456539>
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Hariyono, T. C. (2020). Teaching vocabulary to young learner using video on YouTube at English course. *Language Research in Society Journal*, 1(1), 41-46.
- Hasanah, N., Sari, F. M., & Prasetyo, D. (2024). Video-based vocabulary instruction for young EFL learners. *Journal of English Language Teaching and Learning*, 5(1), 22-34.
- Kaboocha, R., & Elyas, T. (2018). The effects of YouTube in multimedia instruction for vocabulary learning. *English Language Teaching*, 11(2), 72-81. <https://doi.org/10.5539/elt.v11n2p72>
- Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge University Press.
- Nation, I. S. P. (2017). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.
- Rini, J. E., Wijaya, K. F., & Aprilia, R. (2024). YouTube-assisted vocabulary learning in EFL classrooms. *Indonesian Journal of English Language Teaching*, 19(1), 55-69.
- Tohamba, C. P. P. (2025). The use of YouTube in improving English vocabulary mastery among early childhood learners: A systematic literature review. *Journal of English Education and Development*, 8(1), 1-15.
- Younas, M., & Dong, Y. (2024). Animated video and vocabulary learning for young learners: A multimedia learning perspective. *Language Teaching Research Quarterly*, 39, 110-127.