

UNRAVELING THE POWER OF VIDEO CONTENT AND UNLEASHING ENGAGEMENT AS MULTIMODAL TEXTS IN INDONESIA

Retno Purwani Sari

Universitas Komputer Indonesia

retno.purwani.sari@email.unikom.ac.id

Tatan Tawami

Universitas Komputer Indonesia

Zaenab Zahratunnisa

Universitas Komputer Indonesia

Rizki Taufik Mustika

Universitas Komputer Indonesia

ABSTRACT

This study intends to investigate the level of viewer engagement in co-produced video. The focus is primarily on the performance of the video content. How video engagement promotes video's effectiveness is an issue of practices in the measurement and improvement of persuasive technology. The video content, as a multimodal text, has the power to change viewer's attitudes and behaviors. In addition, studies noted that viewer engagement is necessary to determine the success of video content creation for specific purposes, for example, creating brand equity. Using the *Video Dokumenter Pengabdian kepada Masyarakat Program Insentif Kemendikbud 2022* as a case study, Kalodikis' model was applied to measure audience engagement. Questionnaires collected at a single point in time were used to illustrate participant's assumptions. A cross-sectional qualitative study with interaction analysis was applied to describe video engagement metrics and how they correlated to participants' evaluations of the video. The results highlight multiple images displayed in sequential order and show that audio reinforces the message conveyed in the visual and verbal elements of the scheme. Message framing is created through frame per second, image resolution, bit depth, and time. The obtained results indicate that there is still room for improvement.

Keywords: Video content, Multimodal text, Video engagement, Viewer engagement

INTRODUCTION

In today's digital age, video content has exploded in popularity, with platforms like YouTube attracting millions of viewers worldwide. By utilizing a variety of communication modes such as images, sound, and text, video content has the power to create a multi-layered experience that can evoke emotion, convey information, and encourage interaction. Benson (2019: 10) notes that video content's ability to facilitate social interaction through text-based communication within the video

itself is a key feature that can enhance the viewing experience. Analyzing the various modes of video content, including visual imagery and sound, is crucial for identifying the specific elements that contribute to viewer engagement and enhancing the overall experience. For instance, analyzing how visual imagery contributes to conveying information and evoking emotion, or how sound enhances the overall viewing experience, can help identify specific areas that can be improved. By optimizing the qualities that contribute to viewer engagement we can create more effective and impactful video content (Dimitrova et al., 2002: 42). In this paper, we explore the various elements of multimodal video content and analyze how they can be optimized to improve engagement and enhance the viewer experience. Specifically, we examine the English Department at UNIKOM's video "*Video Dokumenter PkM Program Insentif Kemendikbud 2020*," which documents and promotes a community service program in the education sector. Through this case study, we gain a deeper understanding of the potential and limitation of video content as a multimodal text in this particular context.

In recent studies, multimodal analysis has been applied to all types of communication, though it primarily focuses on texts that interact with and integrate two or more semiotic resources, or "modes," to carry out the text's communicative functions (O' Halloran and Smith, 2012). Understanding how invested and interactive your video or content's viewers are is critical to increasing viewer engagement (MayeCreate Design, n.d.). One way to increase viewer engagement is through the use of branding. A brand can be defined as the idea or image that people have in mind when they think about a company's specific products, services, and activities, both practically (e.g., "the shoe is light-weight") and emotionally (e.g., "the shoe makes me feel powerful"). As a result, it is not only the physical characteristics of a company or its product that create a brand, but also the feelings that consumers develop towards the company or its product. When exposed to the name, logo, visual identity, or even the message communicated, this combination of physical and emotional cues is triggered (Marion, 2023). By incorporating branding elements such as a company logo or tagline into video content, viewers can easily associate the content with the brand, which can strengthen the emotional connection between the viewer and the brand. This emotional connection can contribute to increased engagement with the video content, which can lead to increased brand awareness and loyalty. Therefore, video content is strongly associated with multimodal text and viewer engagement, and incorporating branding elements is an effective way to promote engagement and create a strong connection between the viewer and the brand.

According to a previous study on engagement with the source text, accessing and annotating dynamic audio-visual media presented a challenge to bringing transcriptions (O' Halloran and Smith, 2012). To address the challenge, the video-content analysis and retrieval model provides keyframes or key sequences as index entries for stories extracted from video to allow for quick and accurate content access (Dimitrova et al., 2002: 42). The keyframe's reflections on the image create a unique interaction with viewers (Macken-Horarik, 2004: 6). This is how viewer engagement works. The objective of this study is to investigate the level of viewer engagement in co-produced video. A cross-sectional qualitative study with interaction analysis was implemented. As this study focuses on a

single video produced by a specific department at a single university, it is important to highlight that the findings may not be generalizable to other contexts.

DATA SOURCE

This research utilized a documentary video titled “*Video Dokumenter PkM Program Insentif Kemendikbud 2022*,” which was published on the Sastra Inggris Unikom YouTube channel, <https://www.youtube.com/watch?v=g6ADXcBTjU>. The video showcases a series of activities in the latest community service program of the English Department at Universitas Komputer Indonesia. Specifically, the video highlights the chronological events that took place during a collaborative project between the English Department team and Emaki Al Ma’soem. The project involved the development of an information system to meet the needs of Emaki Al Ma’soem and the optimization of their transactional management system.

Considering the growing significance of video content as a marketing tool in both industrial industries and educational institutions, this research aims to not only highlight the collaborative project between the English Department and Emaki Al Ma’soem but also evaluate the effectiveness of the accompanying documentary video in engaging audiences and achieving marketing objectives.

METHODS

The main focus of our study was to advance the investigation of message framing and viewer engagement in co-produced YouTube videos. This was aimed at deriving valuable guidance for tailored persuasion based on engagement. Our research methodology was primarily based on the prior work of Dimitrova et.al (2002), in which the researchers conducted a needs assessment to examine meaningful composition structure and to retrieve and depict content attributes from any video sources. In order to meet research objectives, this study applied Benson’s ideas on structures, processes, and contents in YouTube discourse (2019).

We considered video content as the biggest marketing tool available today. According to the most recent benchmark report from (Wyzowl, 2023), 2023 represents an all-time high since 2016, with 96% of marketers claiming that video content provides them with persuasive enforcement and a good ROI (return on investment). Thus, 91% of businesses, let alone educational institutions, use video on a regular basis. Adjusting for branding, it provides emotionally engagement. By being an engaging and emotion-eliciting channel, it is powerful in creating a relational bond between an institution and its viewers in ways beyond a mono-modal text.

Current approaches to foster the level of video engagement have explored the creation of video-effective content based on a two-types analysis, classifying a video version using direct and implicit methods (Wang and Ji, 2015: 412, 418). Since emotion was involved in video’s multimodal interaction, aesthetic emotion and emotion induction concepts, as well as the various represented-emotions, were also introduced to promote the multidisciplinary cutting-edge technology for emotive

movie content analysis which focuses on video emotion recognition and violence detection (Baveye et al., 2017: 396).

This study explores how video engagement improves the effectiveness of the marketing video. Using some performance indicators, video content was measured for viewer retention. Various metrics and statistics of how the video has been viewed and interacted with can be used to develop video strategy and improve the quality and performance of future videos. Among the topic dealt with marketing videos, the described investigation focuses on viewer engagement, while message framing is being created to promote a specific brand of English Department, Universitas Komputer Indonesia.

Design, Participants, and Hypothesis

We followed a cross-sectional study in which we collected participants' various emotions and ideas at a single point in time (Phakiti et al., 2018; Thomas, 2022). Participants watched the video, *Video Dokumenter PkM Program Insentif Kemendikbud 2022* (Sastra Inggris Unikom, 2022), about a sequence of activities in the latest community service program of English Department Universitas Komputer Indonesia. More specifically, the message pertained chronological events that occurred when English Department, an educational institution, and Emaki Al Ma'soem, industry, collaborated: developing information system to meet the needs of Emaki Al Ma'soem and optimizing the transactional management system.

This study included 31 participants (16 male, 15 female), who were chosen at random from across Indonesia. They were senior high school students, college students, undergraduates, and practitioners who volunteered. The age group ranged from 18 to 55. The purpose of this study, they were all told, was to evaluate and measure video engagement. They watched the video first, then filled out a questionnaire to express their thoughts and feelings about it.

We hypothesized that video content has the power to change viewer's attitudes and behaviors based on previous surveyed literatures on video engagement, which has shown that a higher level of engagement indicates a more effective video. Since YouTube counts views at 30 seconds or more, viewer's attitudes and behaviors were assessed based on the emotions that drew them to act. Closed observation of participants' responses to questions and actions may help to determine the hypothesis.

Measures

We used the following metrics to assess the level of video engagement: frame per second (FPS), image resolution, bit depth, and time (Kalodikis, 2020). Video content, as a complex multimodal text, utilizes all modalities to convey messages in frames per second (Benson, 2019: 40), including moving text and visuals, music, and voiceover.

Furthermore, the parameters influencing video engagement are quality, length/time, audio, and viewer relevance (Henken, 2022). Aside from that, video metrics reveal how viewers interact with and respond to video content (Segal, 2018). To measure engagement, data describing play rate, shares, likes, comments, and reactions, click-through rate, and watch time were collected.

Finally, video has become a more powerful marketing tool than ever. Businesses can measure the success of their videos and make improvements to better connect with their audience by monitoring the four key indicators of play rate, engagement, CTR, and watch time.

THEORETICAL REVIEW

Multimodal Text Analysis

Multimodal text communicates through at least two different modes. In digital media such as cinema, animation, slideshows, and web pages, various combinations of written and spoken language, visual (still and moving images), aural gestural, and spatial modalities are utilized to convey content to the viewers. Each mode generates meaning by utilizing distinct semiotic resources (Department of Education of Victoria State Government, n.d.).

Multimodal text analysis places a special emphasis on considering the communicative meaning included within the text for linguists. However, problems arise when considering semiotic resources other than language, and their interaction with one another and with language, such as visual and aural art, image-text relationships and page layout, cinematography and sound design. These factors have recently emerged as significant challenges to conveying meaning to the viewers. Multimodal analysis examines all forms of communication, but it particularly focuses on the communicative purpose of the texts. This type of analysis creates analytical challenges due to the characteristics of the medium, the depth and breadth of the study, and the integration of semiotic resources across media (O' Halloran and Smith, 2012: 2).

A common framework for analyzing multimodal texts alerts the analyst to two different things: 1) the 'lack of it' when one mode's categories are applied to another; and 2) the deconstructive power of this type of analysis. This power can reveal voids, silence, and, surprisingly, providential riches in trans modal analysis (Macken-Horarik, 2004).

Video-Content Analysis

Content-based video analysis involves identifying and categorizing semantic content in videos, such as news or brands. Online services like YouTube have made videos a popular media of communication and entertainment. Researchers use both visual and aural elements to capture viewer' subjective assessments of affective content through emotional descriptions. Descriptors can be categorical or dimensional, and nonverbal response of the viewers while watching videos can reveal physiological and behavioral reactions (Wang and Ji, 2015: 411).

According to Dimitrova et.al (2002: 43), there are four major processes in content-based video analysis: feature extraction, structural analysis, abstraction, and indexing. Content-based video indexing efficiency depends on the effectiveness of attributes in content representation. The next stage of video content analysis is video structure parsing, which involves extracting structural temporal information from video programs or sequences. This entails identifying temporal boundaries and locating important video chunks.

The next step is to create a shorter version of the original video that presents visual data about a scene or its structure. Similar to the extraction of keywords or summaries in text document processing, this abstraction process involves extracting keyframes or highlights from the source video to use as entry points for shots, scenes, or narrative. Finally, indexing involves creating metadata based on structural and content properties extracted during the feature's extraction, video parsing, and abstraction procedures, as well as manually entered attributes. This metadata is used to generate video indices and table content by employing techniques like clustering, which divides sequences or frames into several visual categories, or an indexing structure.

In conclusion, content-based video analysis is a critical field with significant research challenges. The four major processes of feature extraction, structural analysis, abstraction, and indexing provide a standard methodology for analyzing and indexing video content. Understanding these processes can help researchers more effectively categorize semantic content in videos, providing important insights into viewers' preferences and behaviors.

Video Marketing

There are many different types of videos, including screen recordings, live action, and animation. In 2023, 91% of businesses will use video as a marketing tactic, up from 2016 when tracking of this data began (Wyzowl, n.d.). Marketers also consider video more important than ever, with 96% believing it is a crucial part of their marketing strategy (Wyzowl, n.d.). To gauge the success of a video, there are four key indicators to monitor:

Screen recordings, live action, and animation are all examples of video formats. In 2023, 91% of businesses will use video as a marketing tactic, an increase from 2016 when data collection began (Wyzowl, n.d.). Marketers regard video as more important than ever, with 96% believing it is an essential component of their marketing strategy (Wyzowl, n.d.). There are four key indicators to monitor when determining the success of a video: the play rate; engagement; clickthrough rate (CTR); and watch time (Segal, 2018).

The play rate of a video is the ratio of the total number of views to the number of times it has been viewed. A low play rate may indicate that the video does not appeal to the interests of the target viewers, or that there are issues with the video's copy, subtitles, aesthetic quality, or network on which it was posted. Meanwhile, comments, shares, likes, and responses are all forms of engagement. These metrics can reveal the emotional impact the video had on viewers. Sharing a video is similar to spreading online word of mouth about the institution and can significantly increase viewership. Adding to the indicators, clickthrough rate (CTR) measures the percentage of viewers who were persuaded to take a desired action after watching the video, such as clicking a link. A low CTR may indicate that the call-to-action in the video is unclear or unrelated to the topic at hand. To encourage viewers to take action, the video should be designed brief, concise, and clear. Finally, watch time indicates how much time viewers spent watching the video overall. It provides a clear indication of what information viewers saw and missed based on when they drifted off. When tracking view time, future video length is also considered.

RESULTS AND DISCUSSION

As previously stated, video metrics are used to measure viewer engagement, which include play rate, shares, likes, comments, and reactions, click-through rate, and watch time. This data shows how viewers communicate with and react to video content.

Results

YouTube is a platform that has evolved to accommodate longer viewing sessions. Thus, it counts views at 30 seconds, as opposed to Facebook's 3 seconds. English Department Unikom uses YouTube to increase brand equity for the most widely used video marketing platform, as claimed by 90% of marketers, and 78% of marketers said it is effective for marketing media (Wang and Ji, 2015: 410). A stronger personal connection to the brand is created and improved through the use of video's visual storytelling. For example, *Video Dokumenter PkM Program Insentif Kemendikbud 2022* is successful in establishing a bond and initiating action (see Figure 1).

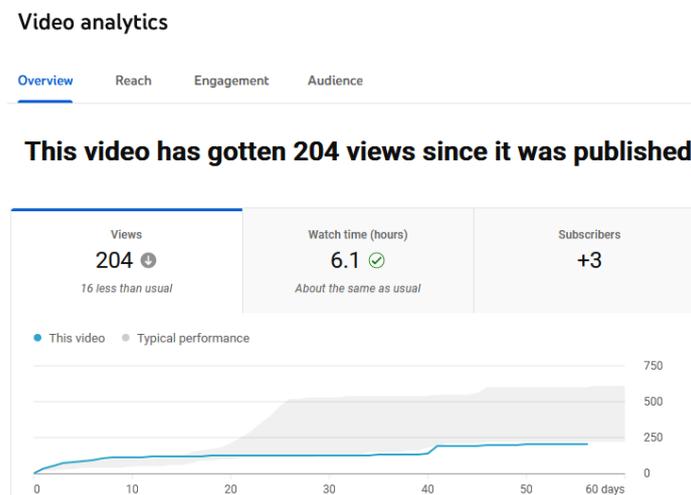


Figure 1 Viewer's Interaction with the Video

Since its release, the video has gained three subscribers. With 6.1 watch time, it indicates that it was successful in retaining and increasing interest. Figure 2 depicts viewer retention in terms of engagement.

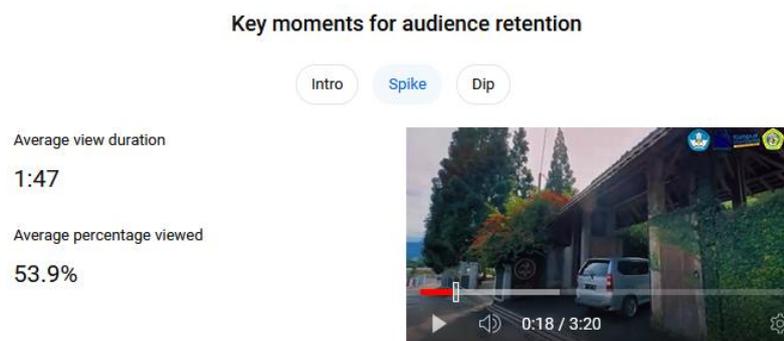


Figure 2 Viewer Retention

The average viewing time is 1:47, with a viewing percentage of 53.9%. It explains why 70% of viewers are still watching at around the 0.30. The metrics report shows a high intro percentage, indicating that the content in the first 30 seconds met the viewer’s expectation based on the video’s thumbnail and title; the content kept the viewers interested (see Figure 3).

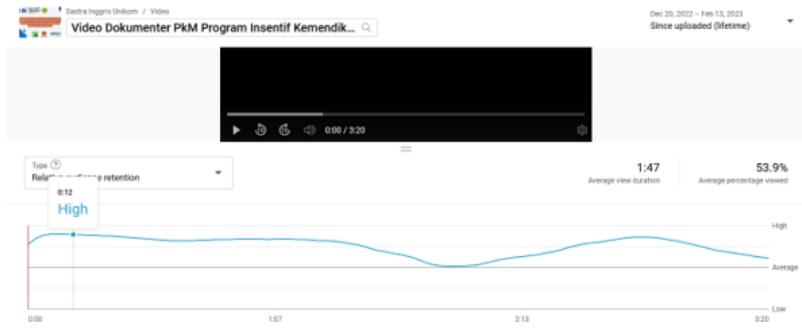


Figure 3 Intro Moment for Viewer Retention Report

Meanwhile, spikes laying in 0:22 provides reasons for improvement. The viewers watched that segment more than previous segments to understand the message.

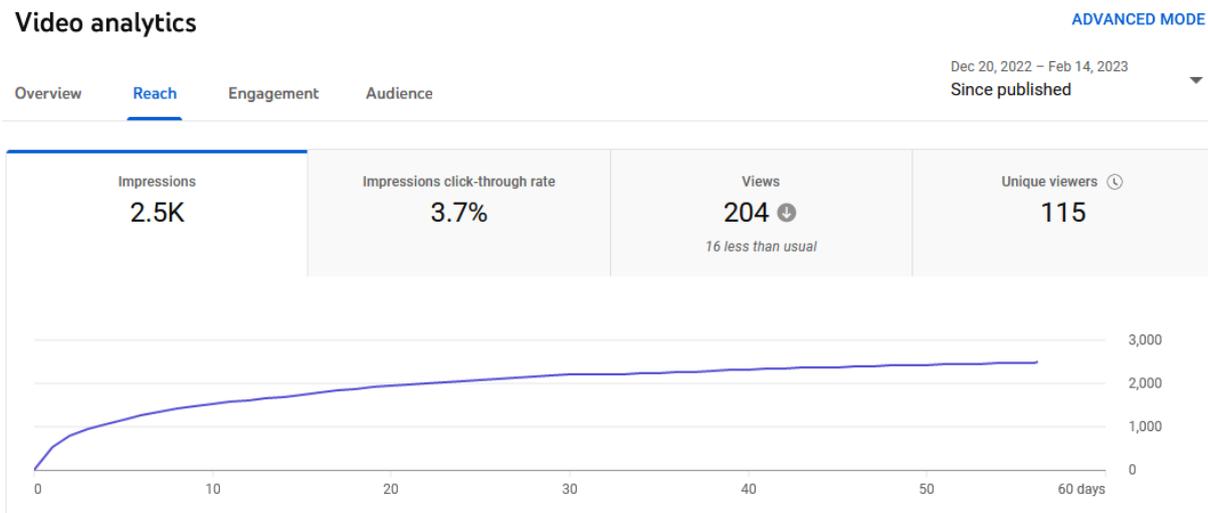


Figure 4 Click-through Rate

Click-through rate (CTR) of the video is 3.7%, meaning that the video had 92,5 clicks and 2500 impressions. Showing the number of people who click the video after viewing it, 3.7% of CTR indicates the video’s quality score reach the average click-through rate.

Discussion

The findings of our study on viewer retention and engagement strongly support our hypothesis that video content has the potential to effectively influence viewer attitudes and behaviors. The results from the questionnaire response indicate that the video’s title successfully met the participants’ expectations, setting a positive initial impression. This finding aligns with previous research

highlighting the importance of captivating titles in capturing viewer interest and engagement (Segal, 2018; Henken, 2022).

The visual storytelling employed in the video was found to be highly effective in establishing a stronger personal connection between the viewers and the brand. This results is consistent with the literature on the power of visual narratives in enhancing brand engagement (Segal, 2018). The ability of the video to evoke emotional response and create a compelling narrative likely contrivuted to its potential as a powerful multi-channel advertising strategy.

As a multimodal text, the observed video-content utilizes a range of resources, including moving text, visuals, and music, to convey its message. The clarity of the storytelling was evident, with 84% of participants stating that they were able to follow the story clearly. This finding emphasizes the importance of visual coherence and narrative flow in engaging viewers and facilitating their understanding of the content.

The keyframe scenes employed in the video were particularly effective in persuading viewers to evaluate the sequence of events an immerse themselves in every moment of the activities. This results aligns with previous studies highlighting the impact of keyframes in enhancing viewer engagement and immersion (Benson, 2019). Additionally, a significant portion of participants (77%) reported sharing the video with others, indicating a high level of engagement and the potential for viral reach.

The successful transmission of the message can be attributed to the effective utilization of sequential images. The manipulation of frames per second, in combination with moving text, visuals, and audio, contrivuted to the video's readability and impact. These findings highlight the importance of careful attention to technical aspects, such as image resolution, bit depth, and timing, in enhancing the viewer's experience and knowledge acquisition (Kalodikis, 2022).

Overall, these findings underscore the significance of video content as a persuasive and engaging medium. The successful combination of visual and textual elemts, along with the careful structuring of the content, has the potential to captivate viewers, elicit emotional responses, and drive desired behaviors. By harnessing the power of video as a multimodal text, brands and advertisers can effectively convey their messages and establish meaningful connections with their target audience.

In future research, it would be benefecial to explore additional metrics such as viewer emotions and cognitive processing during video consumption. Understanding the emotional and cognitive impact of video content can provide valuable insights into its persuasive potential. Additionally, condungting comparative studies between different types of video content and their impact on viewer engagement could further enhance our understanding of effective video marketing strategies.

CONCLUSION

The study used play rate, shares, likes, comments, and reactions, click-through rate, and watch time to investigate the level of viewer engagement in co-produced video. Video content effectively and persuasively engages viewers through the use of multimodal text. Message framing is affected

by frame per second (FPS), image resolution, bit depth, and time. As a result, viewers' attitudes and behaviors are influenced to take the necessary actions. In addition, measuring video engagement provides some justification for specific improvements to digital marketing media.

REFERENCES

- Baveye, Y., Chamaret, C., Dellandréa, E., & Chen, L. (2017). Affective video content analysis: A multidisciplinary insight. *IEEE Transactions on Affective Computing*, 9(4), 396-409.
- Benson, P. 2019. *The Discourse of YouTube: Multimodal Text in a Global Context*. 1st Edition. UK: Taylor & Francis Group.
- Department of Education of Victoria State Government. (n.d.). *Multimodal Literacy*. <https://www.education.vic.gov.au/school/teachers/teachingresources/discipline/english/literacy/readingviewing/Pages/litfocusmultimodal.aspx>
- Dimitrova, N., Zhang, H. J., Shahraray, B., Sezan, I., Huang, T., & Zakhor, A. (2002). Applications of video-content analysis and retrieval. *IEEE multimedia*, 9(3), 42-55.
- Henken, P. (2022, April 17). *Video Engagement: How to Measure, Improve, and Succeed*. Kultura. <https://corp.kultura.com/blog/video-engagement/>
- Kalodikis, C. (2020, March 20). *Digital Representation of Data: Video*. <https://www.youtube.com/watch?v=UtfZZrUk2s4>
- Macken-Horarik, M. (2004). Interacting with the multimodal text: reflections on image and verbiage in Art Express. *Visual communication*, 3(1), 5-26.
- Marion. (2023, May 24). *What is Branding?* The Branding Journal. <https://www.thebrandingjournal.com/2015/10/what-is-branding-definition/>
- MayeCreate Design. (n.d.). *The Ins and Outs of Viewer Engagement*. <https://mayecreate.com/blog/the-ins-and-outs-of-viewer-engagement/>
- O'Halloran, K. L., & Smith, B. A. (2012). Multimodal text analysis. *The encyclopedia of applied linguistics*, 1-5.
- Phakiti, A., De Costa, P., Plonsky, L., & Starfield, S. (2018). Applied linguistics research: Current issues, methods, and trends. *The Palgrave handbook of applied linguistics research methodology*, 5-29.
- Sastra Inggris Unikom. (2022, December 20). *Video Dokumenter PkM Program Insentif Kemendikbud 2022*. <https://www.youtube.com/watch?v=g6ADXcBTtjU>
- Segal, S. (2018, December 16). *How to Measure Video Marketing: The Metrics that Matter*. Oktopost. <https://www.oktopost.com/blog/measure-video-marketing/>
- Thomas, L. (2022, July 21). *Cross-Sectional Study: Definition, Uses & Examples*. Scribbr. <https://www.scribbr.com/methodology/cross-sectional-study/>
- Wang, S., & Ji, Q. (2015). Video affective content analysis: a survey of state-of-the-art methods. *IEEE Transactions on Affective Computing*, 6(4), 410-430.
- Wyzowl. (n.d.). *Video Marketing Statistics 2023*. <https://www.wyzowl.com/video-marketing-statistics/>