Contents list avaliable at Directory of Open Access Journals (DOAJ)
Aulad : Journal on Early Childhood
Volume 8 Issue 3 2025, Page 1483-1492
ISSN: 2655-4798 (Printed); 2655-433X (Online)
Journal Homepage: https://aulad.org/index.php/aulad

Early Childhood Vocabulary Acquisition through Multimodal Strategies: A Semiotic Study of Miss Rachel

Thalita Shafira Azhari[1]✉, **Rohmani Nur Indah**[2]
Sastra Inggris, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia(1)(2)
DOI: 10.31004/aulad.v8i3.1195

✉ Corresponding author:
[thalitashafira111@gmail.com]

Article Info

Keywords:
Multimodality;
Miss Rachel Channel;
Early Childhood Vocabulary Acquisition;
Semiotic Study

Abstract

In this digital era, YouTube Kids has become one of the applications most frequently used by children with a variety of content, some aimed at education and others at entertainment. This research analyzes educational videos specifically for children to examine the multimodal strategies in Miss Rachel's YouTube educational videos that support early childhood vocabulary acquisition. It uses Kress and Leeuwen (2006) multimodality theory, which includes seven semiotic modes: linguistic, visual, gestural, audio, spatial, temporal, and three-dimensional. The method used is descriptive qualitative, with data consisting of video footage of Miss Rachel introducing vocabulary through various multimodal strategies. The results show that each mode contributes meaningfully to the vocabulary learning process. All modes support each other and create a fun, interactive, and meaningful learning experience for children. This research concludes that multimodal strategies in Miss Rachel videos effectively support vocabulary acquisition in early childhood. It also recommends that educators, parents, and digital content developers consider using multimodal approaches in children's language learning activities.

1. INTRODUCTION

Vocabulary acquisition is a process through which individuals, particularly children, learn and internalize new words, both receptively and productively. It is an essential component of early language development because vocabulary growth serves as the foundation for communication, comprehension, and literacy (Yule, 2020). Without sufficient vocabulary, children may face difficulties in expressing ideas, understanding messages, and engaging in meaningful social interaction. The development of vocabulary in children has a significant impact on their cognitive and academic growth. A rich vocabulary helps children perform better in reading comprehension, problem-solving, and overall school readiness (Gleason & Ratner, 2024). Early vocabulary acquisition also contributes to socio-emotional skills, as children are more capable of articulating their feelings and needs. Thus, fostering vocabulary acquisition during the early years is crucial for children's holistic development.

In today's digital era, this process can be greatly supported by the abundance of children's content on YouTube, where engaging visuals, sounds, and gestures are used to introduce and reinforce new vocabulary in enjoyable and meaningful ways. YouTube has introduced the platform for children called YouTube Kids that can be educational medium for children, not only for entertaining. This idea is also supported by Kilag et al. (2023) that state the educational videos on YouTube offers an insightful and interactive way of learning for children. It is no wonder many parents choose YouTube Kids as a daily watch for their children. However, despite its myriad benefits, YouTube can also bring negative effects if the content watched is inappropriate for the audience's age. Some sensitive content deliberately targets children as its audience (Papadamou et al., 2020). So, in this case,

Received 03 Juni 2025; Received in revised form 18 November 2025; Accepted 16 December 2025
Available online 31 December 2025

parents play an important role in supervising children when watching videos on the YouTube platform. It is hoped that all parents will always choose and sort out good educational shows for children. By discussing who they watch with, what they watch, where they watch, why they watch, and when they watch, parents and teachers can help children use YouTube more wisely (Neumann & Herodotou, 2020).

One of the reasons of how YouTube videos could affect vocabulary acquisition is related to toddler learning patterns that require audiovisual stimulation. This is related to visual spatial intelligence in the theory of Gardner (2011) who identified visual spatial intelligence as one of the many types of intelligence that humans possess. Children with high spatial intelligence tend to be good at drawing, designing and understanding the relationships between objects in space. In the context of early childhood education, stimulation through audio-visual media plays a significant role in developing children's visual spatial intelligence. Audio-visual media, such as educational videos, can help children understand spatial concepts in a more interesting and fun way. The use of audio-visual media can also strengthen children's visual memory that supports learning (Fathonah et al., 2020; Kilag et al., 2023). Accordingly, it needs further investigation on how audio-visual as part of multimodal stimulation can have impact to children learning particularly on vocabulary acquisition through children's educational videos on the YouTube platform. Moreover, examining multimodal strategies is more significant since it can bring various impacts based on the diverse need and contexts (Amatullah et al., 2019).

There are lots of specialized content on YouTube for children, but some channels are well-known and have a large audience. One of them is a YouTube channel called “Miss Rachel,” which has 17.2 million subscribers worldwide and an average of millions of views on each video. The content created by Miss Rachel is made specifically for children as an educational video. There are many types of content created by Miss Rachel, such as children's songs, storytelling and imagination, learning colours, numbers and letters, teaching first words, and others. Miss Rachel's YouTube channel became famous thanks to the many viral videos on the TikTok platform that showed the reactions of babies and young children when watching Miss Rachel's videos. Most of these videos are of babies smiling and laughing within the first second of Miss Rachel's opening video, while Miss Rachel says “Hi, Hello!” in the opening.

Then, some other videos show that young children successfully say the first words or words taught in Miss Rachel's videos. For example, an account username @ruby.red30 shows a toddler watching Miss Rachel's vocabulary teaching video. In the short video, she mimics and says the words taught by Miss Rachel, including “apple, cookie, pottie, bath, shoes”. The video caught the public's attention and garnered 4.9 million views and thousands of positive comments. Of the thousands of comments, many parents have shared their experiences of watching Miss Rachel's content with their children and stated that their children could also say their first words thanks to watching Miss Rachel's videos. This phenomenon is noteworthy because Miss Rachel's videos have positively impacted many babies and young children.

This enthusiastic reaction highlights how effective the specialized strategies, such as friendly facial expressions, clear articulation, repetitive language, bright visuals, and engaging music, can support young children’s attention, enjoyment, and language development. The use of signs, images, and symbols can be analyzed through semiotics lense, specifically the multimodal theory. This theory was introduced by Gunther Kress and Theo Van Leeuwen in 1996. This theory discusses using semiotic modes such as images, language, sound, movement, and others to convey certain meanings. Some semiotic modes that exist in multimodal include visual modes that include images, colours, and important layouts in the media. The second is the verbal mode, which involves text and spoken or written language. The third is the auditory mode, which includes sound, music, and intonation that enriches the message's meaning. Fourth, the gestural mode relates to body movements and facial expressions in non-verbal communication. Last, the spatial mode relates to spatial arrangements and media elements that affect how messages are understood (Septiani & Sari, 2022). Thus, the elements involved such as text, image, sound, and movement can shaped meaning in the video. This is as noted by Basyid and Firmansyah (2023) that the use of several elements (text, image, sound, and movement) is called multimodality. To help clarify the multimodal theory, Figure 1 is an example of the use of multimodal in the Miss Rachel video:

Figure 1. The Example Of Multimodal Strategy

The picture shows Miss Rachel teaching the vocabulary of 'pig' using several semiotic modes at one time. Miss Rachel uses an animated picture of a pig (visual), the words 'pig' (text), and the additional backsound of a

typical farm piano (sound). The use of several semiotic modes at the same time is called multimodality. This multimodal theory can be used in semiotic research to understand how multimodal strategies are used in YouTube videos.

Research on multimodal mostly chooses advertisements as the object of research (Amatullah et al., 2019; Chen & Cheung, 2022; Herman et al., 2022; Ni Made Sila Ulati, 2021; Pricilia & Suyudi, 2023; Raharjo et al., 2020; Rizvi, 2020; Sari & Noverino, 2021) which shows that the use of multimodal in advertisements aims to add a certain impression of meaning according to the advertised product, besides that several studies also chose posters (Belgrimet & Rabab'ah, 2021; Dallyono & Sukyadi, 2019; Linh, 2021; Nurudeen et al., 2021) which shows that the use of multimodal in posters aims to attract readers' attention and convey messages more effectively, and also video games (Hawreliak & Lemieux, 2020; Setiawati et al., 2024) whose findings show that the use of multimodal in video games can make it easier for players to understand how to play.

Many similarities in the objects of this research indicate that further research is still needed to explore other research objects, namely special children's content on the YouTube platform. The use of multimodal will certainly differ in each field in terms of its use and purpose (Amatullah et al., 2019). In addition, few studies on multimodal examine how the audience reacts or the impact caused by the use of multimodal itself, so, research that focuses on other objects by studying the multimodal used can be interesting. With this research, there will be several benefits. First, an understanding of how multimodal is used to attract audiences will be obtained. Second, a knowledge of which the content creators often use multimodal elements. Moreover, finally, how the use of multimodal strategies impacts the audience and achieves certain goals. Then, an understanding of whether the use of multimodal is useful can be obtained.

This study aims to obtain an in-depth description of the multimodal strategies used by Miss Rachel in her videos on the YouTube platform. Specifically, it describes the multimodal elements used in the videos. In addition, this research also produces the detail explanation concerning the context of the multimodal strategies used to helping language acquisition. Based on the above explanation, this study aims to answer the following two questions which are what are the multimodal elements of strategies used in the Miss Rachel's videos? and also how are these multimodal elements used to help deliver vocabulary learning to the target audience namely early childhood?

2. METHOD

This study employed a descriptive qualitative content and multimodal discourse analysis to explore how various semiotic modes combine to form multimodality in Miss Rachel's YouTube videos. The descriptive qualitative content was chosen because it allows the researcher to provide detailed, systematic, and factual descriptions of the multimodal elements without statistical generalization. The multimodal discourse analysis is suitable for studies aiming to interpret meaning based on observation and description of visual, linguistic, and auditory features, where the researcher acts as the primary instrument for data collection and interpretation. Through this approach, the study seeks to portray how multimodal strategies support children's vocabulary acquisition naturally and contextually.

The data of this research consist of linguistic, visual, auditory, gestural, spatial, temporal, and three-dimensional elements found in Miss Rachel's videos. These elements include spoken words, background music, gestures, visual layouts, images, and physical objects that appear on screen. The data sources are selected videos from two playlists on Miss Rachel's YouTube channel—"First Words" and "Videos for Toddlers with Miss Rachel." The selection criteria were based on high viewer numbers, thematic relevance to vocabulary learning, and a publication year up to 2022. These criteria ensure that the analyzed materials are representative of her most influential educational content.

The data collection process involved several stages. First, the researcher selected the videos that met the inclusion criteria. Then, each video was repeatedly watched to identify segments related to vocabulary teaching. During this observation, screenshots were taken to capture visual, gestural, and spatial modes, while transcripts were made for the linguistic mode and audio recordings for the auditory mode. These collected materials were organized into a coding sheet, which contained video identity, timecode, description of each multimodal mode, communicative function, and initial interpretation. The researcher used observation, documentation, and transcription as the main techniques, with herself as the main instrument. To guide observation and coding, an analytical instrument was developed based on indicators for each mode—for instance, word repetition and clarity for the linguistic mode; gaze, facial expression, and gesture direction for the gestural mode; sound emphasis and music synchronization for the auditory mode; and object interaction or color contrast for the visual and three-dimensional modes.

For data analysis, a systematic approach was employed to facilitate an in-depth understanding. Once data collection was complete, the data were categorized according to the multimodal modes outlined by Kress and van Leeuwen (2006): linguistic, visual, gestural, audio, three-dimensional, spatial, and temporal modes. This categorization addressed the first research question concerning the types of multimodal elements present in Miss Rachel's videos. Subsequent analysis of each multimodal element utilized analytical tools developed by Kress and

van Leeuwen. These tools include guiding questions such as: What is the communicative purpose of this representation? Who is the intended audience? How do visual and textual elements interact to convey meaning? These questions supported the examination of how multimodal strategies contribute to vocabulary acquisition in children, thereby addressing the second research question. The analysis also considered how combinations of these modes work together synergistically to enhance the learning experience, with particular attention to how repeated exposure, synchronization of audio-visual cues, and child-friendly design features support word recognition and retention. The findings from this analysis were synthesized to draw conclusions regarding the role of multimodal strategies in supporting vocabulary development in early childhood education. The flowchart is as presented in Figure 2.

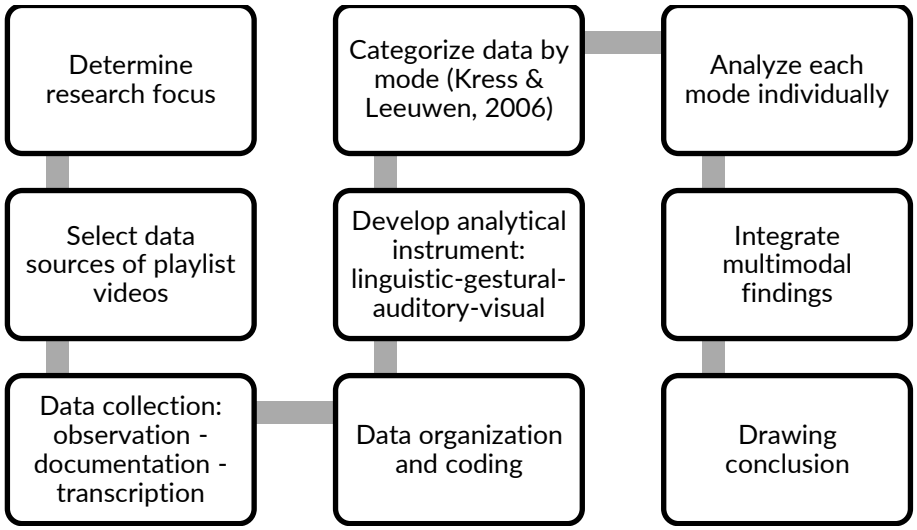

Figure 2. Flowchart of research procedure

3. RESULT AND DISCUSSION

The Multimodal Features Supporting Children Vocabulary Acquisition

This research aims to analyze the multimodal elements or features identified in Miss Rachel's videos on YouTube platform. Her videos are specialized for children to help their vocabulary acquisition. It involved 11 data containing multimodal elements by Kress & Leeuwen (2006). The analysis reveals several multimodal elements namely visual, linguistic, gestural, three-dimensional, spatial, audio, and temporal modes on Miss Rachel's videos. Figure 3 demonstrates three samples from the data. The findings show that the video created by Miss Rachel to help vocabulary acquisition in children uses the multimodal elements to support audience understanding. Each mode contributes uniquely to meaning-making, fostering early childhood vocabulary acquisition.

The Linguistics Mode: Implementation of Bruner's Scaffolding

Kress and Leeuwen (2006) see linguistic mode as one of the main carriers of meaning that reinforces other meanings through narrative and direct instruction. This mode contains of both written and spoken language shown in the videos. Miss Rachel employs simple sentences, repetition, and an interactive style to introduce and reinforce new vocabulary. Using questions, invitations to repeat, and emphasis on keywords helps children recognize and remember the words taught. For instance, in introducing the word keys, the linguistics mode is reinforced by the following utterance:

You say keys, and the keys will come out.
Keys… keys…
Say it again louder! Keys…
The keys came out, good job!

In her videos. Miss Rachel effectively implements Bruner's concept of scaffolding through the use of simple sentences, repetition, and an interactive style to break down the complex task of new vocabulary acquisition into manageable, supported steps. The exposure to the vocabularies is progressively made as she begins with initial support by modelling the word ("You say keys..."), moves to prompting and cueing when she engages the children by asking for repetition ("Keys... keys... Say it again louder!"). Afterwards, she concludes with confirmation and reward to validate the child's effort ("The keys came out, good job!"). Through this sequence, it allows the child to successfully perform the task of articulating the word, a skill they might not yet be able to master independently, thereby facilitating language learning. These steps represent constructive learning practicing Bruner's view of assisting learning through the process of mastering new vocabulary that is divided into small, guided increments.

It becomes undeniable that in the context of teaching vocabulary to young learners, Bruner's view is applicable not only in the classical learning environment (Mudawamah et al., 2025; Putri et al., 2025) but also in online platform such as Miss Rachel's videos.

Visual	Gestural	Three-dimentional	Spatial
			
			
			

Figure 3. Sample of multimodal elements

The Visual Mode: Iconic Representation Supporting Children Attention

Miss Rachel employs object visualization, vibrant colours, and interesting graphics support children's focus and attention. The use of clean backgrounds and clearly displayed objects also support the visual meaning of vocabulary. It is in line with Kress and Leeuwen's (2006) visual grammar framework that salience aspects such as primary colours and high contrast help focus the audience's (children's) attention on the main focus (the vocabulary being taught). For example, look at Figure 4 where the text labels 'Keys' are superimposed near the relevant item to connect spoken and written forms, promoting literacy through visual reinforcement.

Figure 4. Visual Mode Showing A Combination Of Bright Key Colours And Attractive Facial Expression

The visual mode in Miss Rachel's videos is not only relevant to the visual grammar framework (Kress & van Leeuwen, 2006) but also Bruner's theory of iconic representation to optimize toddler vocabulary learning. The visual mode takes the role as the iconic bridge, connecting the spoken word to the tangible object as the targeted vocabulary. It becomes vital for toddlers' fast mapping since they learn primarily through sight and sound. Miss Rachel videos employ vibrant colours, captivating graphics, and focused object visualization (like colourful keys). This technique aligns with Kress and van Leeuwen's (2006) principle of salience, where high-contrast visuals effectively mandate a child's attention toward the key vocabulary item written in yellow colour. By presenting concepts clearly and accessibly in this visual, iconic form, the videos successfully prepare toddlers for the later mastery of the abstract symbolic form of the language, supporting Bruner's theory of representation (Myers, 2021).

The iconic representation as the key for engaging children to pay more attention and keep their focus to the video content supports the process of vocabulary learning. It is in line with the finding of Minalla (2024) that mentions the use of visual representation in the videos helps young learners associate the new words with images, making it easier for them to understand and remember. Moreover, Bobkina et al. (2025) also affirms that the visual representation in the videos attract deeper engagement and better vocabulary retention of the children. While in Miss Rachel videos, the use of bright colors and attractive facial expression become the elements of iconic representation supporting children attention.

The Gestural Mode Supporting Enactive Learning

In line with Kress and Leeuwen's (2006) theory, gestures are part of meaningful bodily modes, Miss Rachel is using expressive and appropriate body movements and gestures to the learning context. Hand gestures, pointing gestures and action demonstrations such as driving a tractor (look at Figure 5) or brushing a cat's fur (look at Figure 5) visualize the meaning of vocabulary and enhance children's understanding. Fourth, the three-dimensional mode, Miss Rachel uses 3D objects, such as cake toy (Figure 5), and airplane toy (Figure 5), appear real and interactive in the video. This representation helps children make the connection between the physical form of the object and the vocabulary being taught. This is in line with Kress and Leeuwen's (2006) view that three-dimensional representations provide a more complete multisensory experience.

Figure 5. Iconic Gestures Engaging Child's Imitation

The figures demonstrate the use of iconic gestures for pointing and showing the object that directly engages the child's body and movement, linking the conceptual word to a physical motor action that they can imitate. This dynamic process supports Bruner's enactive mode, where knowledge is initially represented through action and manipulation. It means the toddler can internalize the word through physical imitation before they can fully master verbal articulation. Bruner's concept of enactive mode is put into action when the modelling reinforces gestures, showing objects and facial expression that can attract attention even to children with short attention span (Rakshit, 2024).

The Spatial Mode: Visual Scaffolding for Joint Attention

The spatial mode, the placement of characters and objects in the video frame is strategically arranged by Miss Rachel. Miss Rachel's position in the centre or slightly to the side creates a comfortable visual space for the child, and the movement of objects from inside to outside the frame (such as the circles that roll out of the frame at Figure 6) adds an element of surprise and interactivity. The movement of the circle rolling out of the frame demonstrates the application of framing and spatial composition principles. According to Kress and Leeuwen's (2006) multimodal theory, space serves not just as a background, but as a medium for conveying meaning. When the circle shifts from the centre of the screen (the focal point) to beyond the frame, it indicates a transition or change in visual focus.

Figure 6. Spatial Mode Showing The Movement Of Circle Object

As shown in Figure 6, Miss Rachel's videos are dominated by the background minimization for reduced cognitive load. The background is typically bright, monochromatic, and uncluttered. This deliberate simplification of the visual field reduces cognitive distraction, ensuring that the child's focus remains exclusively on Miss Rachel, her facial expressions, her gestures, and the objects being presented. When objects are introduced, they are placed clearly in the foreground or held centrally. This use of framing ensures the object is the visual nucleus of the composition, clearly signalling its importance and supporting the learning of the associated word or concept. The organization aids in introducing basic visual literacy and scanning skills of toddlers. In the concept of Vygotsky, this elements of spatial mode can scaffold substantive development even without the presence of teacher direction offline in the context of distant learning though videos (Taber & Li, 2021). Miss Rachel's videos provide toddlers with immersive digital environment in the form of 2D with simple design as a way of scaffolding as a realization of Vygotsky learning concept. This type of spatial elements can be a basic before the children are expose to a more complex spatial mode in other immersive digital environment containing patial contiguity, natural embedding, and directing into joint attention (Bacca-Atosta et al., 2022).

The Audio Mode to Maximize Learning and Engagement

Concerning the audio mode, Miss Rachel uses cheerful background music, sound effects to dramatize the action, and clear and expressive voice intonation to attract children's attention and reinforce the meaning of the words. This aligns with Kress and Leeuwen's (2006) perspective that auditory elements such as intonation, volume, and rhythm are crucial for creating an engaging experience and enhancing understanding. Last, the temporal mode, she structures her videos coherently, with an easy-to-follow sequence of activities and smooth transitions between segments. Repetition is inserted at appropriate times to reinforce learning, and each section has a duration tailored to young children's attention spans. This aligns with Kress and Leeuwen's (2006) perspective that inter-mode synchronization, such as the coordination between motion, sound, and words, constitutes a fundamental aspect of multimodal grammar. This grammar integrates simultaneous and sequential experiences of meaning, enabling richer and more cohesive communication across different modes.

In Miss Rachel's videos, the rhythmic and prosodic audio cues (the sing-song voice, the rhythmic musical underscore, the specific repetition of a word) function as contingent supports. Because they are naturally embedded within the pace of the song or conversation, they prevent the child from encountering a learning difficulty. The audio mode as a form of assistance to enable the toddlers engange in the normal flow of the interaction, support Bruner's view. Beside maximizing learning, the audio mode such as as intonation, volume, and rhythm also become the key elements that engange the child to participate in a task before they can fully master it on their own. More importantly, the audio mode becomes a central modelling technique for the targeted pronunciation (Putri et al., 2025).

Each of the modes mentioned earlier plays a vital role in creating a comprehensive learning experience. The interplay between these modes ensures that vocabulary development incorporates not only verbal components but also visual, physical, and emotional dimensions. Miss Rachel's deliberate and consistent application of multimodal strategies reflects the framework outlined by Kress and Van Leeuwen, demonstrating how digital media can help enhance early childhood language development through diverse communication methods.

The findings of this study support Soriano and Ruiz (2024) that Miss Rachel's educational videos have a positive impact on early childhood. They noted that Miss Rachel's interactive content focuses on the pronunciation of words and phrases and incorporates visual aids, hand gestures, and sign language, as well as real-life examples that are essential for verbal and non-verbal communication. Moreover, this was a case study of a 2-year-old child, so it can be seen that Miss Rachel's educational videos positively impact early childhood vocabulary acquisition. In addition, Yelland (2018) showed that children learn more effectively when they are exposed to multimodal-based learning, such as a combination of images, sound, text, movement, and direct interaction through digital media. This is in line with the current research findings which show that Miss Rachel's video, which uses seven multimodal elements, made specifically to supports children's vocabulary acquisition. Yelland (2018) also disproves the belief that the use of technology is just for entertainment, and instead shows that the conscious and structured use of digital media can build children's literacy. This reinforces the conclusion that the proportional use of videos like Miss Rachel should not be considered as "negative screen time", but rather as a modern and effective form of digital pedagogy, provided it is properly directed and selected. While digital media such as Miss Rachel's videos are effective in supporting language acquisition, it is important to regulate the duration and supervise children to ensure that the use of technology remains aligned with age-appropriate developmental principles.

Another study from Crescenzi-Lanna (2020) also confirms that multimodal learning environments are especially beneficial for young children because they integrate various sensory inputs that mirror children's natural learning processes. This supports the multimodal strategies examined in this study, such as using real objects (e.g., a toy cake or a comb), coordinated gestures, and expressive speech to teach vocabulary. Crescenzi-Lanna (2020) study calls for multimodal frameworks to be implemented in early childhood learning environments. While the findings of the current study provide a real-world example of how such theoretical frameworks are already being effectively employed in an accessible, child-friendly digital content. These results suggest that future research and

educational practice should consider integrating similar multimodal strategies not only in digital media but also in traditional classroom settings to maximize vocabulary development. Furthermore, the findings emphasize the need for collaboration between educators, content creators, and parents in curating and utilizing digital content that leverages multimodal elements to support early language learning in meaningful ways.

4. CONCLUSION

This finding of this study confirms that semiotically the videos of Miss Rachel contain five among seven multimodal elements namely linguistic, visual, gestural, spatial, and audio. Each of the modes mentioned plays an important role in vocabulary acquisition as its representation is in line with the learning principles of Bruner and Vygotsky. The linguistic mode shown through the use of simple sentences, repetition, and interactive questions encourage children's participation following Bruner's scaffolding. The visual elements such as bright colours, recognizable objects, and text placed alongside images to strengthen the association between words and objects implement Bruner's theory of representation. The gestural mode supports understanding of meaning through facial expressions and gestures that are expressive and representative of the word's meaning also connect to Bruner's enactive mode. The spatial and audio modes also interpret Vygotsky concept to support joint attention. Overall, the multimodal strategies in Miss Rachel's video create a fun, contextualized and interactive learning experience for young children. Each mode does not work in isolation, but complements each other to build meaning simultaneously. This research confirms that the multimodal approach is indeed a promising way to support vocabulary acquisition in video-based educational media.

Based on the research findings, some suggestions are given for readers or future research. First, as this study found that Miss Rachel used the seven multimodal elements completely, this study suggests that other teachers can use and develop the use of multimodal elements to help children acquire vocabulary not only with English but also with other languages, first languages, or others. Second, from this study, it can be concluded that educational videos by Miss Rachel on the YouTube channel, which utilize multimodal strategies, can support the vocabulary acquisition process in children. Thus, appropriate screen time for children can positively impact them, of course, with the assistance of parents or other assistants to regulate appropriate screen time. Lastly, further studies can expand the scope of the study by analyzing other platforms that are growing rapidly as a medium for children's learning, such as TikTok Edu, YouTube Shorts, or educational reels on Instagram. These platforms have short-duration formats, fast-paced rhythms, and a highly visual-auditory approach, opening up new possibilities in how children receive and absorb information.

5. ACKNOWLEDGEMENTS

All praise and gratitude are due to Allah SWT for His blessings and guidance throughout the completion of this research. Sincere thanks are also extended to all parties who have supported and contributed to this study, directly or indirectly.

6. REFERENCES

Amatullah, F., Rosa, R. N., & Fitrawati, F. (2019). An Analysis of Multimodal in Beauty Product Advertisements. *English Language and Literature*, *8*(1). https://doi.org/10.24036/ell.v8i1.103122

Bacca-Acosta, J., Tejada, J., Fabregat, R., Kinshuk, & Guevara, J. (2022). Scaffolding in immersive virtual reality environments for learning English: an eye tracking study. *Educational technology research and development*, *70*(1), 339-362. https://doi.org/10.1007/s11423-021-10068-7

Belgrimet, S., & Rabab'ah, G. (2021). A Multimodal Discourse Analysis of English Posters in Violence Awareness Campaigns against Women. *Theory and Practice in Language Studies*, *11*(2), 191. https://doi.org/10.17507/tpls.1102.12

Bobkina, J., Baluyan, S., & Dominguez Romero, E. (2025). Tech-Enhanced Vocabulary Acquisition: Exploring the Use of Student-Created Video Learning Materials in the Tertiary-Level EFL (English as a Foreign Language) Flipped Classroom. *Education Sciences*, *15*(4). https://doi.org/10.3390/educsci15040450

Chen, Z. T., & Cheung, M. (2022). Consumption as extended carnival on Tmall in contemporary China: A social semiotic multimodal analysis of interactive banner ads. *Social Semiotics*, *32*(2), 163–183. https://doi.org/10.1080/10350330.2020.1720992

Crescenzi-Lanna, L. (2020). Multimodal Learning Analytics research with young children: A systematic review. *British Journal of Educational Technology*, *51*(5), 1485–1504. https://doi.org/10.1111/bjet.12959

Dallyono, R., & Sukyadi, D. (2019). An analysis of multimodal resources in environmental protection posters. *Indonesian Journal of Applied Linguistics*, *9*(2). https://doi.org/10.17509/ijal.v9i2.20245

Doyle, L., McCabe, C., Keogh, B., Brady, A., & McCann, M. (2020). An overview of the qualitative descriptive design within nursing research. *Journal of Research in Nursing*, *25*(5), 443–455. https://doi.org/10.1177/1744987119880234

Fathonah, M. F., Wahyuningsih, S., & Syamsuddin, M. M. (2020). Efektivitas Media Audio Visual Terhadap Kecerdasan Visual Spasial Anak Usia 5-6 Tahun. *Kumara Cendekia*, *8*(2), 142. https://doi.org/10.20961/kc.v8i2.39789

Gardner, H. (2011). *Frames of mind: The theory of multiple intelligences* (10th anniversary ed). Basic Books.

Gleason, J. B., & Ratner, N. B. (Eds.). (2024). *The development of language* (Tenth edition). Plural Publishing.

Hasyim, M., & Arafah, B. (2023). Semiotic Multimodality Communication in The Age of New Media. *Studies in Media and Communication*, *11*(1), 96. https://doi.org/10.11114/smc.v11i1.5865

Hawreliak, J., & Lemieux, A. (2020). The semiotics of social justice: A multimodal approach to examining social justice issues in videogames. *Discourse: Studies in the Cultural Politics of Education*, *41*(5), 723–739. https://doi.org/10.1080/01596306.2020.1769936

Herman, H., Sulistyani, S., Ngongo, M., Fatmawati, E., & Saputra, N. (2022). The Structures of Visual Components on a Print Advertisement: A Case on Multimodal Analysis. *Studies in Media and Communication*, *10*(2), 145. https://doi.org/10.11114/smc.v10i2.5717

Kilag, O. K. T., Malbas, M. H., Arcillo, M. T., & Barcena, M. C. (2023). The Role of YouTube Children's Educational Videos in Enhancing Early Childhood English Language Proficiency: An Investigation of Parental Perceptions. *International Journal of Scientific Multidisciplinary Research*, *1*(7), 833–846. https://doi.org/10.55927/ijsmr.v1i7.3545

Kress, G., & Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2e éd). Routledge.

Linh, N. T. T. (2021). A Multimodal Discourse Analysis of Romantic Comedy Movie Posters. *VNU Journal of Foreign Studies*, *37*(3). https://doi.org/10.25073/2525-2445/vnufs.4647

Manoli, P., & Papadopoulou, M. (2015, December 18). *Young Children Accessing Multimodal Texts: A Case Study*. New Semiotics. Between Tradition and Innovation. https://doi.org/10.24308/iass-2014-110

Minalla, A. A. (2024). Enhancing Young EFL Learners' Vocabulary Learning Through Contextualizing Animated Videos. *Theory & Practice in Language Studies (TPLS)*, *14*(2). https://doi.org/10.17507/tpls.1402.31

Mudawamah, R. Y., Latifah, E., Lestari, S. N., & Hildayani, S. (2025). From ABC To English: Strategies to boost children's speaking confidence early. *Jurnal Intelek Insan Cendikia*, *2*(6), 12628-12639. https://jicnusantara.com/index.php/jiic/article/view/3937

Myers, S. (2021). Jerome Bruner: An overview of key ideas. *Jerome Bruner, Meaning Making and Education for Conflict Resolution: Why How We Think Matters*, 1-14. https://doi.org/10.1108/978-1-80071-074-020211001

Neumann, M. M., & Herodotou, C. (2020). Young Children and YouTube: A global phenomenon. *Childhood Education*, *96*(4), 72–77. https://doi.org/10.1080/00094056.2020.1796459

Ni Made Sila Ulati. (2021). Multimodal Critical Discourse Analysis on advertisement of Colgate. *International Journal of Systemic Functional Linguistics*, *4*(1), 10–13. https://doi.org/10.55637/ijsfl.4.1.4092.10-13

Nurudeen, M. A., Ogungbe, E. O., & Zakariyah, M. (2021). A Socio Semiotic Approach to Multimodal Discourse of Selected Nollywood Film Advertisement Posters. *Studies in Pragmatics and Discourse Analysis*, *2*(2), 31–45. https://doi.org/10.48185/spda.v2i2.364

Papadamou, K., Papasavva, A., Zannettou, S., Blackburn, J., Kourtellis, N., Leontiadis, I., Stringhini, G., & Sirivianos, M. (2020). Disturbed YouTube for Kids: Characterizing and Detecting Inappropriate Videos Targeting Young Children. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*, 522–533. https://doi.org/10.1609/icwsm.v14i1.7320

Pricilia, M., & Suyudi, I. (2023). A Multimodal Critical Discourse Analysis of "Garnier Sakura White" Advertisement. *International Journal of English and Applied Linguistics (IJEAL)*, *3*(1), 28–38. https://doi.org/10.47709/ijeal.v3i1.2095

Putri, A., Prastiwi, A. E., & Hidayani, S. (2025). The effectiveness of Jerome Bruner's scaffolding method in improving speaking skills in primary education. *Jurnal Pendidikan dan Keguruan*, *3*(9), 1167-1177. https://jutepe-joln.net/index.php/JURPERU/article/view/938

Raharjo, A. R., Hidayat, D. N., Alek, A., & Jalil, N. (2020). Multimodal Analysis of Wardah Lipstick Advertisement. *Elite English and Literature Journal*, *7*(2), 149. https://doi.org/10.24252/10.24252/elite.v7i2a4

Rakshit, S. S. (2024). Toy-Based Pedagogical Approaches to Foster Social Interaction and Communication in Children with ASD: A Constructivist Perspective. *Diversity, Equity & Inclusion*, 46.

Rizvi, S. Z. (2020). *A Multimodal Discourse Analysis of E-Advertisement Visuals*. *24*(10).

Sari, V. W., & Noverino, D. R. (2021). *A Multimodal Discourse Analysis in Pantene Advertisement*. *4*(10). https://doi.org/10.32996/ijllt.2021.4.10.4

Search Engine Journal. (2024, September 18). *The Top 10 Social Media Sites & Platforms*. Search Engine Journal. https://www.searchenginejournal.com/social-media/social-media-platforms/

Septiani, E., & Sari, N. I. (2022). Analisis Multimodal Video Promosi Aplikasi Kasir Pintar. *Pujangga*, *8*(1), 100. https://doi.org/10.47313/pujangga.v8i1.1634

Setiawati, E., Purba, R., Suwondo, T., Judijanto, L., Kencanawati, D., Herman, H., & Fatmawati, E. (2024). *An investigation on discourse structure of video games tutorials: A case in multimodal discourse analysis*. *5*(3). https://doi.org/10.58256/c4h2qk76

Siregar, I. (2022). Semiotics Analysis in The Betawi Traditional Wedding "Palang Pintu": The Study of Semiotics Roland Barthes. *International Journal of Linguistics Studies*, *2*(1), 01–07. https://doi.org/10.32996/ijls.2022.2.1.1

Soriano, M. A. L., & Ruiz, M. J. G. (2024). *The Impact of the "Ms. Rachel" Educational Video on the Communication Skills Development of a Child: A Case Study*.

Taber, K. S., & Li, X. (2021). The vicarious and the virtual: A Vygotskian perspective on digital learning resources as tools for scaffolding conceptual development. *Advances in psychology research*, *143*, 1-72.

Yelland, N. J. (2018). A pedagogy of multiliteracies: Young children and multimodal learning with tablets. *British Journal of Educational Technology*, *49*(5), 847–858. https://doi.org/10.1111/bjet.12635

Yule, G. (2023). *The Study of Language*. Cambridge University Press.