

# Enhancing Vocabulary Learning: Multimodal Engagement in English Teaching Social Media Video

Zihan Bu<sup>1</sup>, Sunyun Pan<sup>2</sup>  
and Ziyi Wang<sup>3,\*</sup>

<sup>1</sup>Department of School of Foreign Languages, XiangNan University, Chenzhou, 423000, China

<sup>2</sup>Department of German, University of Shanghai for Science and Technology, Shanghai, 200093, China

<sup>3</sup>Department of Teaching English as A Second Language, Universiti Putra Malaysia, Serdang, 43400, Malaysia

\*Corresponding author: 213833@student.upm.edu.my

## Abstract:

The growing digitalization and multimodality in the field of English acquisition have aroused an inclination to research English teaching videos on social media. Previous research primarily focused on multimodal engagement in school-based English learning, while this research was dedicated to the survey of multimodality in English vocabulary teaching through social media videos. Grammar and composition are also the major subjects mentioned in other studies. In this case, we chose to select vocabulary learning. This study conducted two distinct case studies that discussed characteristics and merits in this category of videos. Data from a questionnaire developed is also analyzed through SPSS. The findings indicate that multimodality is a great strategy for content creators of English vocabulary instructing videos, and the majority of the content creators have a heightened awareness of multimodal engagement. Additionally, both English teachers and students can use the degree of multimodality as a criterion for selecting supplementary vocabulary learning video resources.

**Keywords:** Multimodality; English Teaching Social Media Videos; Vocabulary Learning.

## 1. Introduction

Given the development of both information technology and learning analytics, English teaching social media videos have become one of the major online study sources for language learners regarding vocabulary enrichment. These videos generally incorporate multiple modes. Images, texts, and audio grant learners access to interactive learning experiences. This multimodal learning environment is recognized for enhancing cognitive engagement and improving

learning outcomes.

However, as the representation of multimodal education, the potential of English vocabulary teaching videos has not been fully learned and explored. Hence, immediate research on multimodal engagement in English vocabulary teaching social media videos is necessary. Aghaei and Gouglani demonstrated the critical role of multimodality in L2 vocabulary retention [1]. It is undeniable that this study highlighted the students' positive perspectives on multimodality. However, the specific context discussed in the arti-

cle imposed certain limitations on the research strategies employed. Since pedagogies of online pictures and slides contain fewer multimodal elements compared to videos under current digitalization.

Smith and Reed examined Task-Based Language Teaching (TBLT) and its effectiveness in enhancing oral language proficiency among English learners [2]. Smith and Reed argued that Task-Based Language Teaching (TBLT) provides a structured, engaging approach, concluding that TBLT positively impacts learners' oral proficiency in terms of fluency, accuracy, and complexity. Lastly, the authors call for further exploration of the integration of form-focused instruction within the framework of TBLT and an investigation of the long-term effects of TBLT on speaking proficiency.

Hu's study on teaching through multimodal online videos emphasizes the importance and effectiveness of using both linguistic and nonlinguistic resources in language learning [3]. The study offered a valuable framework for analyzing multimodal interactions. However, the generalizability and accuracy of the findings were constrained by the specific context of the research and the reliance on self-reported data.

Further research with a broader scope and more objective measures is needed to strengthen the persuasiveness of the findings.

In summary, the articles above only addressed the utilization of multimodal English learning under specific circumstances and lacked concrete strategies for implementing multimodality in the field of extracurricular learning materials. Consequently, practical strategies, perspectives from both teachers and video content creators, integration of social media, and multimodal learning are all considered in our investigation. The questions this research aims to solve include: What characteristics and merits do English vocabulary instructing videos with multimodality attain? What multimodal-strategy content creators should utilize to elevate the quality of English vocabulary tutoring videos? How could students and teachers select tutoring videos? Case studies and questionnaire surveys are implemented to obtain the solutions to the proposed questions.

## 2. Methodology

### 2.1 Research Approach

This study employs a mixed-methods approach, combining case analysis and a questionnaire to explore the effectiveness of multimodal teaching methods in English language education. Aghaei and Gouglani emphasize the significance of multimodal metaphors in English vocabu-

lary learning [1]. However, the study's small sample size limits the generalizability of its findings and its broader applicability. Smith and Reed, in their TESOL Journal article, discuss the effectiveness of Task-Based Language Teaching (TBLT) in improving oral proficiency, underscoring the need for carefully designed tasks and adequately prepared teachers [2]. Hu examines the use of multimodal online video instruction to teach pragmatic competence, but the study's specific context and methodology restrict the generalizability of its findings [3]. Additionally, the questionnaire used in this study was adapted from research by Xu Lei, who investigated factors influencing user engagement with short English educational videos [4]. Xu's work provided a valuable tool for gathering quantitative data on student interaction with multimodal learning resources. These references collectively informed the study's approach to analyzing how various multimodal elements enhance vocabulary learning, providing both theoretical and practical frameworks for the research.

## 2.2 Case Study

### 2.2.1 Overview

To probe the multimodal characteristics of online English vocabulary teaching videos, 60 videos are collected. They originate from mainstream social media platforms (RED, Bilibili, TikTok, Instagram, YouTube, and Kuaishou). These videos encompass multiple teaching modalities, including text, audio, visual presentation, and host interaction.

Among 60 samples, all of them used audio with the purpose of both pronouncing instructions and memory enhancement. 45 (60%) videos involved visual demonstrations. This suggests that the majority of English video producers are aware of incorporating multimodality into their teaching. Additionally, over 50% (32) of the videos featured hosts to increase interactivity.

### 2.2.2 Analysis of an English teaching video: vocabulary related to times of the day

Video teaching can be used to teach vocabulary engagingly and memorably. The video titled "Can You Name These Times of a Day?" provides an excellent example of how visual and auditory elements can be combined to enhance language learning.

The first advantage of this video is that it provides clear content and a focus on vocabulary. Geoff Bull and Michèle Anstey point out that the text producer selects from the semiotic systems and technologies to design a text with a particular purpose, context, and audience in mind [5]. In this video, the audience is beginner English

speakers who need information about the time of day, so the vocabulary and presentation should be clear and easy to understand. The main content of the video revolves around seven distinct times of the day: 6 am (dawn), 8 am (morning), 12 pm (noon), 3 pm (afternoon), 5 pm (dusk), 6 pm (evening), and 9 pm (night). This selection covers fundamental concepts used in daily conversations, making teaching and learning practical in life.

Moreover, visual aids also play a crucial role in this video. Jill Bourne, Anton Franks, and his colleagues point out that multimodal teaching encompasses not only the spoken and written language of classrooms but also other modes of representation and communication that are important in English teaching [6]. For example, images and spatial organization. In the video, each time of the day is paired with an image, aiding in vocabulary acquisition by helping learners associate words with visual contexts. For example, a sunrise image helps illustrate “dawn,” while a sunset represents “dusk.” These visual aids are very effective for language learners.

In addition, the role of the facilitator is important. Querol-Julián, M. emphasizes the importance of the facilitator in the English classroom [7]. His results showed the complexity of classroom interaction in this digital environment, the importance of lecturer waiting time, the high modal density and functional diversity of the follow-up/feedback stage, and the most frequent discourse functions expressed during the interaction. The facilitator’s tone, pace, and enthusiasm will make the knowledge delivery more engaging and more precise. Through this process, the audience is more likely to participate and enjoy the learning process.

In conclusion, the video “Can You Name These Times of a Day?” is a valuable resource for teaching vocabulary related to the times of day. It uses clear content and explanations, visual aids, and an enthusiastic facilitation process to make it an effective tool for beginning English language learners. This multimodal video instruction has been shown to be superior to traditional methods and can potentially increase motivation and enjoyment of language learning.

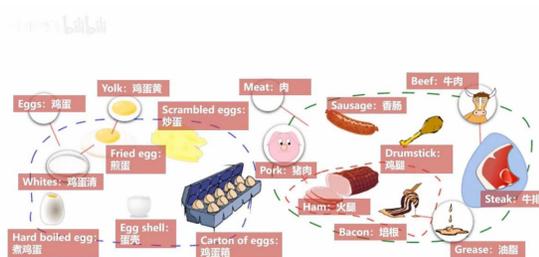
### 2.2.3 Analysis of English teaching videos: vocabulary related to food and beverage

Learning English has become the key to effective communication in today’s globalised world. This analysis explores how English teaching videos focusing on food and beverage vocabulary can help language learners master basic vocabulary. This study aims to highlight the advantages and challenges of using video-based teaching methods for language learning by assessing the integration of various instructional approaches and considering the

potential limitations posed by different cultural contexts. According to Johnson, “using video-based instruction can enhance English vocabulary learning” [8]. The potential of video as a teaching tool. Video can provide authentic language input, expose learners to different accents and speaking styles, and provide visual backgrounds that aid in understanding.

Teaching methods in English videos

The analyzed video uses memory technology to associate vocabulary with vivid images or captivating stories to enhance memory retention and recall. In the middle of the video, the host connects various foods through pictures, which helps improve learners’ memory. For example, as seen in Figure 1, when learning the word ‘pork,’ the real park helps you quickly establish connections between words and objects.



**Fig. 1 Bilibili 《English food, beverage, and diet-related vocabulary sharing》: Learn about food vocabulary through some vivid pictures**

Benefits of Native Speaker Pronunciation

Another advantage of this video is that it typically features English spoken by native speakers or professional voice actors, providing learners with standard English pronunciation for learning vocabulary. Learners can hear standard English pronunciation to correct learner pronunciation errors. Learners can improve their oral expression ability by repeatedly listening to and imitating the pronunciation in the video. For example, when faced with words that describe food with similar pronunciations, such as “honey” and “hollow,” learners are prone to pronunciation errors, as detailed descriptions are provided in the video.

Scene introduction

This video can create various contexts, allowing you to learn words in specific situations and better understand their usage. For example, in the context of a restaurant, the learners can learn words such as “menu,” “waiter,” and “order” and understand how these words are used in practical situations.

In short, on dietary vocabulary provide valuable tools for language learners. Through effective teaching methods,

visual aids, and learner engagement strategies, this video can significantly enhance vocabulary acquisition. Video can provide engaging learning experiences but may distract or challenge learners with lower language proficiency. Brown stated that “the impact of English teaching videos on vocabulary learning strategies is significant” [9]. With the continuous advancement of technology, there is hope for producing more attractive and effective English teaching videos in this field.

### 2.3 Questionnaire Survey

A questionnaire was employed to serve as an instrument to learn the utilization status of English vocabulary tutoring videos on social media among the public. The

questionnaire covered seven statistical questions to collect fundamental information and used a five-level scale, which included 25 questions to gain a deeper insight. The questionnaire from Research on the factors affecting the stickiness of users in English knowledge short videos was revised to suit the current study’s demand.

#### 2.3.1 Reliability Test

Reliability analysis was conducted for one variable (reliability) in the questionnaire. As shown in the table below, Cronbach’s alpha coefficients for the variables were all greater than 0.9, indicating that the scale used in this study had good reliability and that the data collected were reliable.

**Table 1. Reliability Analysis**

Variable	Number of Items	Cronbach’s Alpha:
Reliability	25	0.932.

#### 2.3.2 Analysis of Validity

The validity of the questionnaire was evaluated using several tests, including the results of the KMO (Kaiser-Meyer-Olkin) sampling adequacy measure and the Bartlett

sphericity test. As seen from Table 1 and Table 2, the KMO value of the scale is 0.908, and the Bartlett sphere test is less than 0.01, which means that the scale is suitable for factor analysis.

**Table 2. Analysis of validity**

Variable Meyer Olkin Sampling Suitability Scale		908
Bartlett’s sphericity test	Approximate chi square test	4328.030
	Freedom	300
	Significance	000

Additionally, as seen in Table 1, the factor analysis results show that a total of 4 factors are extracted, and the cumulative explanatory variation of the factors is 77.380%, greater than the recommended cumulative variation standard. In social science, more than 60% is the best, so the explanation level is good.

#### 2.3.3 Descriptive Statistics

In the present research, we compiled 294 questionnaires. 31 of the respondents had not watched English vocabulary learning videos previously. The SPSS 27.0 was used to run descriptive statistics, reliability and correlation index. Among the remaining 263 respondents, the majority of them were aged between 18-25, constituting 63.1% of the sample, reflecting a heightened inclination towards embracing online English learning videos among the younger groups.

Furthermore, 40.7% of the respondents are students, and 32.4% are enrolled in undergraduate or graduate programs. Since 95.8% of them dedicate at least 10 minutes

per day to watching online videos for the purpose of studying English vocabulary. This medium is a favored tool for vocabulary learning within the student population, which aligns with the phenomenon mentioned in an article from Noetel, which states that large numbers of students in higher education use video learning at present [10].

#### 2.3.4 Correlation Analysis

As demonstrated in Table 3, content value, social value, information source, social influence, satisfaction, and user viscosity were correlationally analyzed. Each dimensionality includes three to five scale items. Among 49 correlation indexes, over 70% of the correlation indexes were greater than 0.4, while every index was over 0.2. This outcome indicated that each factor was positively correlated. The pronounced positive correlation of 0.516 between sentimental value and user viscosity implies that the more our samples endorse the efficiency of learning English vocabulary through videos, the more willingness they possess to integrate this learning pattern into their daily

English study. The correlation between content value and user viscosity equally infers that the individuals who are more willing to adopt this format of learning are more likely to appreciate the professionalism and quality of videos.

**Table 3. Correlation Index**

		Content Value	Sentimental Value	Social Value	Source	Social Influence	Satisfaction	User Viscosity
Content Value	Pearson	1	461**	407**	395**	418**	408**	452**
	Significance		000	000	000	000	000	000
	N	263	263	263	263	263	263	263
Sentimental Value	Pearson	461**	1	453**	522**	514**	410**	516**
	Significance	000		000	000	000	000	000
	N	263	263	263	263	263	263	263
Social Value	Pearson	407**	453**	1	409**	394**	411**	267**
	Significance	000	000		000	000	000	000
	N	263	263	263	263	263	263	263
Source	Pearson	395**	522**	409**	1	456**	369**	365**
	Significance	000	000	000		000	000	000
	N	263	263	263	263	263	263	263
Social Influence	Pearson	418**	514**	394**	456**	1	302**	420**
	Significance	000	000	000	000		000	000
	N	263	263	263	263	263	263	263
Satisfaction	Pearson	408**	410**	411**	369**	302**	1	314**
	Significance	000	000	000	000	000		000
	N	263	263	263	263	263	263	263
User Viscosity	Pearson	452**	516**	267**	365**	420**	314**	1
	Significance	000	000	000	000	000	000	
	N	263	263	263	263	263	263	263

### 3. Discussion

This study confirms that multimodal engagement significantly enhances English vocabulary learning through social media videos, supporting the use of visual, auditory, and interactive elements to improve learner engagement and retention. The correlation analysis between variables such as Content Value, Sentimental Value, and User Viscosity further highlights these findings. For example, the Pearson correlation of .452 between Content Value and User Viscosity indicates a moderate positive relationship, meaning that the more users value the content, the more likely they are to engage with the videos consistently. This suggests that the professionalism and usefulness of the

video content are important factors in attracting continued user engagement. This suggests that video producers need to focus on improving the quality of their content to ensure that videos have high learning value. Similarly, the high correlation of .516 between Sentimental Value and User Viscosity suggests that emotional attachment to the videos enhances engagement. This shows that in videos, the tone of the host, the style of the video, and the way it is interacted with all can increase user attachment to the video. This suggests to video creators that videos need to have not only pedagogical value in terms of content but also elements that can trigger emotional resonance in learners. All correlations were statistically significant ( $p <$

.05), underscoring the reliability of these relationships and exploring the integration of AI-driven, personalized video content to further enhance learning outcomes.

## 4. Conclusion

A key finding of this study is that multimodal engagement significantly enhances English vocabulary learning through social media videos. For content creators, proper integration of multimodality significantly elevates the quality of short online videos designed for English vocabulary learning. From the perspectives of the audience, who are mostly students in the current context, this strategy not only fortifies the efficiency of learning but also promotes the overall studying experience.

In accordance with two distinct comprehensive case studies of typical online short-term English vocabulary instruction videos and one questionnaire survey with the help of SPSS, an enhanced consciousness of integrating multimodality in after-curricular online learning is dictated. The findings illustrated above provide additional suggestions concerning the content innovation and selection of online English tutoring videos. Hence, the intended demographic is content creators of the videos and teachers who select the videos as extra learning resources for students. Firstly, more attention should be paid to the synergistic effects between multiple modalities so that the sophistication of English vocabulary learning videos would be promoted. Another point to mention is that viewers, especially students, should attain a mature multimodal awareness to cultivate a sharp capability of resolution. Considering the ongoing trend toward educational digitalization and the continuous development of multimodality theory, this study offers valuable insights into post-curricular English vocabulary education on social media platforms.

In essence, this article hopes the findings and suggestions of current research illustrated above could provide content creators and viewers of English instruction online videos with valuable concepts and practices. However, the contemporary survey only focused on vocabulary instruction videos and could not study various formats of online videos such as grammar, oral communication, writing, etc. A

further study could also assess the long-term effects of online multimodal English tutoring videos in all categories to upgrade and profound the understanding of the utilization of multimodality in the context of English instruction via online short-term videos.

### Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

## References

- [1] Aghaei, K., & Gouglani, F. Multimodal pedagogy and L2 vocabulary retention. *International Journal of English Language & Translation Studies*, 2016, 4(3), 142-153.
- [2] Smith, J., & Reed, B. The effectiveness of task-based language teaching (TBLT) in improving English language learners' oral proficiency. *TESOL Journal*, 2020, 6(2), 105-124.
- [3] Hu Yanyan. Research on Teaching Second Language Interactional Pragmatic Ability from a Multimodal Perspective: A Case Study of Online Video Teaching by Native English-Speaking Teachers for English Majors. *Taste Classic*, 2023, (19), 42-44.
- [4] Xu Lei. Research on Factors Influencing User Stickiness of English Knowledge Short Videos (Master's thesis, Liaoning University). Master's Thesis, 2022.
- [5] Bull, G., & Anstey, M. Elaborating multiliteracies through multimodal texts: Changing classroom practices and developing teacher pedagogies. Routledge, 2018.
- [6] Bourne, J., Franks, A., Hardcastle, J., Jewitt, C., Jones, K., Kress, G., & Reid, E. English in urban classrooms: A multimodal perspective on teaching and learning. Routledge, 2004.
- [7] Querol-Julián, M. Multimodal interaction in English-medium instruction: How does a lecturer promote and enhance students' participation in a live online lecture?. *Journal of English for Academic Purposes*, 2023, 61, 101207.
- [8] Johnson, A. Using video-based instruction to enhance English vocabulary learning. *ELT Journal*, 2021, (45), 321-328.
- [9] Brown, C. The impact of English teaching videos on vocabulary learning strategies. *Language Learning & Technology*, 2020, (24), 123-135.
- [10] Noetel, M., Griffith, S., Delaney, O., Sanders, T., Parker, P., del Pozo Cruz, B., & Lonsdale, C. Video improves learning in higher education: A systematic review. *Review of educational research*, 2021, 91(2), 204-236.