

# TEACHER'S PERSPECTIVES ON USING TAILORED YOUTUBE SHORTS TO TEACH ENGLISH VOCABULARY MEANINGS TO FIFTH GRADERS: AN ACTION RESEARCH STUDY AT A PRIMARY SCHOOL IN HAI PHONG

QUAN ĐIỂM CỦA GIÁO VIÊN VỀ VIỆC SỬ DỤNG YOUTUBE SHORTS ĐƯỢC THIẾT KẾ RIÊNG TRONG DẠY NGHĨA TỪ VỰNG TIẾNG ANH CHO HỌC SINH LỚP 5: MỘT NGHIÊN CỨU HÀNH ĐỘNG TẠI MỘT TRƯỜNG TIỂU HỌC Ở HẢI PHÒNG

Nguyen Thi Huong<sup>1</sup>, Bui Nguyen Khanh<sup>1</sup>,  
Pham Xuan Quynh<sup>2,\*</sup>

DOI: <https://doi.org/10.57001/huivh5804.2026.202>

## ABSTRACT

Many primary schools in Vietnam are still using translation and rote learning methods in teaching vocabulary, which limit students' engagement with word meaning in context. The present study examined the viewpoints of one teacher with regard to the utilization of a series of 10 to 15-second-long video clips designed around a single target word (tailored YouTube Shorts) for teaching English vocabulary meanings to 44 fifth-grade learners in a public primary school in Hai Phong, Vietnam. The two-cycle action research design (planning, action, observation and reflection) was used in this 10-week intervention. Classroom observations, video recording and reflective field notes were gathered by the teacher-researcher and analysed using six-phase thematic analysis. The teacher noted that the videos were multimodal inputs which allowed students to infer word meanings before translating into Vietnamese, limiting translation dependence. The videos were, however, used only at the "Meaning" stage of a Meaning-Use-Form sequence: teacher mediation (questioning, modelling, scaffolding and feedback) was required to move students from observation to comprehension and was enhanced across the two cycles. The study proposes that when embedded in a structured sequence and supported by active teacher mediation, video clips created by the author in the form of YouTube Shorts can be used as a supplement in teaching students vocabulary.

**Keywords:** Tailored YouTube Shorts; teacher's perspectives; vocabulary teaching; teacher mediation; action research; primary school learners.

## TÓM TẮT

Dạy từ vựng tiếng Anh ở nhiều trường tiểu học Việt Nam vẫn dựa nhiều vào dịch nghĩa và ghi nhớ máy móc. Nghiên cứu tìm hiểu quan điểm một giáo viên về dùng YouTube Shorts thiết kế riêng (video 10 - 15 giây, một từ mục tiêu) để dạy nghĩa từ vựng cho 44 học sinh lớp 5 tại một trường tiểu học ở Hải Phòng. Qua nghiên cứu hành động hai chu kỳ (10 tuần), giáo viên thu thập dữ liệu bằng quan sát, ghi hình và nhật ký phản tư, phân tích theo quy trình chủ đề sáu giai đoạn. Giáo viên xem video là đầu vào đa phương thức giúp học sinh suy luận nghĩa từ trước khi có bản dịch. Video chỉ hỗ trợ giai đoạn Nghĩa trong trình tự Nghĩa - Sử dụng - Hình thức; sự trung gian có chủ đích của giáo viên là điều kiện cần để học sinh chuyển từ quan sát sang hiểu nghĩa, và được tinh chỉnh dần. YouTube Shorts thiết kế riêng có thể là học liệu bổ trợ hữu ích khi lồng trong trình tự dạy học có cấu trúc, cùng trung gian tích cực của giáo viên.

**Từ khóa:** YouTube Shorts; nhận thức giáo viên; dạy từ vựng; trung gian của giáo viên; nghiên cứu hành động; học sinh tiểu học.

<sup>1</sup>School of Languages and Tourism, Hanoi University of Industry, Vietnam

<sup>2</sup>Lai Cach Primary School, Hai Phong city, Vietnam

\*Email: [xuanquynhhd2013@gmail.com](mailto:xuanquynhhd2013@gmail.com)

Received: 30/7/2026

Revised: 28/8/2026

Accepted: 30/8/2026

## 1. INTRODUCTION

Vocabulary is a foundation of English proficiency, especially for primary students who are still developing their basic language skills [1, 2]. For fifth graders, vocabulary also supports their reading, listening, and speaking skills. Many EFL classrooms in Vietnam are still dominated by the translation method in teaching vocabulary, which provides little chance for learners to explore meaning in context and retain new vocabulary [3].

Short-form video platforms have become prominent among young people, particularly TikTok, which has gained a massive young user base and high engagement rates in Vietnam [4]. Research in this area has so far focused on full-length YouTube videos used as supplementary classroom material [5]. YouTube Shorts, by contrast, communicate meaning in terms of image, sound and action within 10-15 seconds and with little or no explanation, offering a more compressed form of multimodal input. They are appropriate for young learners who need concrete, engaging input rather than abstract explanations [6] and who learn through observable action and classroom interaction [7].

Limited exposure to English outside the classroom is a particular problem in Vietnamese primary schools, where students have little opportunity to encounter real English. According to international research, video including YouTube has a positive impact on vocabulary learning and attitudes [5, 8], and learners generally hold positive perceptions of YouTube use in the language classroom [9]. However, three important limitations remain in the existing body of research. First, the vast majority of research is on full-length video, and the short duration of a Short raises concerns about processing; studies using repeated viewing of dual-subtitled videos indicate that repeated viewing improves vocabulary learning more than single viewing [10]. Second, many existing studies do not connect their design to an explicit theoretical framework. Third, for teachers the pressing question may be less whether Shorts are effective than whether they are manageable given the documented barriers to English language instruction in Vietnamese primary education, including limited lesson time [11].

The research gap, therefore, is not whether short-form video is “effective”, but how teachers, who organise and mediate its use, perceive its role, potential, and limitations in teaching word meaning. Few studies have linked ultra-short video to the teacher’s mediating role within a sociocultural framework in a mainstream public

primary EFL classroom in Vietnam. This study addresses that gap by exploring one teacher’s perspectives on using tailored YouTube Shorts to teach vocabulary meanings to fifth graders in Hai Phong, and by documenting her mediating role when ultra-short video is embedded in a structured instructional sequence.

The study is guided by the following research question: *“What are the teacher’s perspectives on the use of tailored YouTube Shorts in teaching vocabulary meanings to fifth graders?”*

## 2. LITERATURE REVIEW

Vocabulary knowledge is widely regarded as a strong predictor of overall language proficiency [1] and a central strand of instruction rather than a marginal sub-skill [2]; the primary years matter particularly, since the lexical base built there underlies later literacy and content learning [7]. Young learners are still developing cognitively and linguistically and depend more than adults on contextual information and implicit learning [6, 12], so they learn best when word meaning is tied to concrete experience rather than to abstract explanation. Translation and rote memorisation are thus poorly matched to how they learn.

This study applies a Meaning-Use-Form (MUF) instructional sequence, treated here not as a fixed method but as a researcher-adapted sequence operationalising word knowledge within a sociocultural view of learning. Its ordering follows a broader principle in communicative and task-based pedagogy, in which attention to form is embedded within, or follows, meaning-focused activity [13, 14]. The sequence is also compatible with Larsen-Freeman’s [15] account of language as interconnected dimensions of meaning, form, and use, arranged here to suit young learners, who benefit from concrete, contextualised experience [7]; Nation’s [2] description of word knowledge (form, meaning, use) informs the dimensions addressed but does not prescribe this order. In practice, meaning is established first through multimodal input; learners then engage in guided use of the word; explicit attention to pronunciation and written form follows.

The sequence is likewise consistent with sociocultural theory [16], in which learning is mediated through cultural tools and social interaction rather than produced by exposure to input alone. Video used merely as entertainment wastes its learning potential, whereas video around which the teacher pauses, questions, and co-constructs meaning can become a powerful

mediational tool [17]. This mediation corresponds to Vygotsky's [16] idea of learning as a process of guidance from a more knowledgeable other, and Wood, Bruner, and Ross's [18] idea of scaffolding that is tailored to the learner's level and gradually removed. Without such mediation, there is a possibility of "watching" and "learning" being separated.

The research conducted on YouTube shows that there are positive outcomes of using YouTube: better vocabulary learning and attitudes [8]; higher motivation with authentic language exposure [5]; higher level of enjoyment and participation among young learners [17]; higher level of vocabulary mastery when using video with Total Physical Response (TPR) [19]; and uneven attention across verbal and visual channels in multimodal materials [20]. In these studies, multimodal input appears to be effective when images, sound, and movement serve to stabilise meaning and foreground the target word for reuse, rather than when these elements merely co-occur in the input.

The number of processing turns that a single 10-15-second Short can sustain has not been determined, though there is some evidence in video that repeated viewing of dual-subtitled video enhances vocabulary learning more than single viewing [10], something that is seldom studied in shorter forms. The major concern in English teaching for primary school students in Vietnam is its feasibility within documented resource and time constraints, not its general value [11] since vocabulary teaching is still largely based on translation [3]. This study therefore focuses on the teacher's perspective as the person responsible for mediation.

### 3. RESEARCH METHODOLOGY

#### 3.1. Research design

The study was of a classroom-based action research (AR) design with two cycles of planning, action, observation, and reflection as developed by Kemmis and McTaggart [21]. Instead of conducting experiments to examine the effectiveness of YouTube Shorts in a controlled setting, it sought to investigate the teacher's perceptions and develop her instructional practice and instructional practice during an iterative cycle of reflection and adaptation [22] as the teacher adapted and incorporated a specific type of video into one particular classroom. "Improvement" was defined not as gains in students' vocabulary scores but as improvements in the teacher's mediating practice (e.g., an improved MUF sequence or more appropriate scaffolding strategies).

The study involved two sets of 5 lessons (in total 10 lessons in 10 weeks). Two cycles were judged appropriate as a single cycle would not have allowed any meaningful problems to be identified and then addressed, while more cycles would have been impractical in the school timetable, and would have risked repetition of data. YouTube Shorts were incorporated into lessons designed and delivered in Cycle 1 and observation focused on qualitative data (classroom observation, video recording and field notes) relevant to the teacher's perspective. The plan for Cycle 2 was then modified with additional pause and predict opportunities, more opportunities for peer interaction, more scaffolding prompts, and more frequent word recycling based on the reflections from Cycle 1.

#### 3.2. Participants, setting, and teaching materials

The teacher-researcher, who worked with the participants and designed the videos, taught the lessons, observed the classroom, and reflected on both cycles, was the participant relevant to the research question. She had more than ten years of experience in teaching English in primary school and was acquainted with the grade 5 curriculum and students' learning needs.

The study was carried out in one intact class in Grade 5 with 44 students (aged 10-11) of a public primary school in Hai Phong, Vietnam, who have studied English for approximately three years (on the national curriculum) and who have reached an English proficiency level close to A1 on the Common European Framework of Reference for Languages (CEFR). The sessions were 35 minutes each and were conducted based on the General Education Programme of the Ministry of Education and Training, and the lessons used *Wonderful World 5* in a setting typical of public primary EFL teaching in Vietnam where classes are large, lessons are short, and there is a heavy focus on translation for the purpose of learning vocabulary.

The teaching material was a collection of 40 YouTube Shorts created by the teacher that were 10-15 seconds in length and based on one target word from the Grade 5 syllabus. Each video conveyed meaning through familiar actions, objects, and situations, and the written form appeared only once meaning had been established. Language was controlled at A1 level, distractions were minimised, and all clips were vetted by the supervisors and by peer teachers before use. In the video for "butterfly", for example, a boy watches a butterfly and says, "Look. It's so pretty. It's flying." The movement is

repeated slowly, after which the written form is displayed with slow, clear pronunciation. No explanation or translation was given: students were expected to infer meaning through observation, discussion, and guided interaction. Table 1 summarises the design principles applied across all 40 videos.

Table 1. Design principles for tailored YouTube Shorts

| Design criterion      | Description                               |
|-----------------------|-------------------------------------------|
| Target vocabulary     | One target word per video                 |
| Duration              | Approximately 10 - 15 seconds             |
| Language level        | A1-level English                          |
| Visual context        | Familiar actions and daily objects        |
| Meaning presentation  | Meaning presented before written form     |
| Visual design         | Limited distractions, simple visual focus |
| Pacing                | Slow and clear actions and narration      |
| Teacher mediation     | Pausing, questioning, discussion          |
| Pronunciation support | Slow, clear pronunciation provided        |
| Classroom suitability | Used under teacher-controlled conditions  |

### 3.3. Research instruments

Three data collection instruments were employed to document the teacher's instructional practice and reflections throughout the study. The first instrument was a classroom observation checklist which had three parts. Part A recorded instructional implementation (use of Shorts as meaning-focused input, pausing and guiding questions, guided-use activities such as pair work, TPR, or role-play, scaffolding, and brief attention to form); Part B recorded engagement as low, medium, or high and Part C consisted of short reflective notes on what worked, what was difficult, and planned adjustments. Classroom observation was chosen as the main data collection method because it is a valuable means of helping teachers reflect on and develop their practice [23]. Every lesson was also video-recorded with two cameras to verify the observations, and field notes combined a descriptive part (context, sequence, interaction, behaviour) with a reflective part (effectiveness, engagement, difficulties, improvements).

### 3.4. Data collection

Observation was carried out under natural conditions over 10 weeks, one lesson per week (10 lessons of 35 minutes, coded Lessons 1 - 10). The teacher acted as both instructor and observer; records were cross-checked against recordings and field notes to avoid reliance on memory. Observation focused on how the teacher

presented and paused the videos, what questions she asked, how activities were designed, and how students inferred meaning from visual and contextual cues. At the end of each cycle, she reflected on what had worked, what was difficult, and what adjustments were needed.

### 3.5. Data analysis

The qualitative data were analysed using Braun and Clarke's [24] six-phase thematic analysis: familiarisation, initial coding, searching for themes, reviewing themes against the data, defining and naming themes, and producing the report. Coding was largely inductive, with the themes and categories derived primarily from the data. Only findings from the teacher's observational, field-note, and reflective data are reported here.

## 4. RESULTS AND DISCUSSION

### 4.1. Multimodal input helped the teacher reduce reliance on translation

Field notes and observation checklists show that the teacher clearly recognised what multimodal input (image, action, context) contributed to students' ability to infer meaning without lengthy explanation in Vietnamese. In Lesson 1, students identified the verb "fly" from a butterfly's wing movement; in Lesson 3 they correctly guessed "puppy" from an image of a small dog; in Lessons 8 and 10 they inferred "bitter" and "mountain" from images and context. Students also distinguished easily confused items such as "run" and "sail", or "river" and "waterfall", by attending to differences in movement and setting.

The reflective notes return repeatedly to the same point: the visual context supplied by the videos reduced the amount of Vietnamese explanation required and left students more room to work meaning out for themselves. After Lesson 5 the teacher wrote that "the multimedia approach significantly reduced the time needed for explanation, allowing more time for speaking practice". After Lesson 9, when teaching the more abstract antonym pairs bright/dark and easy to see/difficult to see, she described the videos as "a very effective way to elicit the vocabulary without translating", an indication that the benefit of multimodal input extended beyond concrete items to some abstract ones as well.

Video was used consistently at the "Meaning" stage across both cycles, supplying the context from which students inferred the target word, with the written form introduced only after meaning had been established. The reflective notes state this helped students "understand

the concept visually before moving to the challenging pronunciation of words like binoculars" (Lesson 1) and "grasp the long compound word sleeping bag through visual association before formal practice" (Lesson 2). The "Use" and "Form" stages, by contrast, depended on activities that did not involve video: pair work, TPR, choral and individual drilling, and rhythmic clapping for multisyllabic words. Table 2 summarises the implementation of seven instructional components recorded across the 10 observed lessons.

All seven components in Table 2 were implemented in every lesson. This uniformity indicates that the planned Meaning-Use-Form sequence was delivered as intended from the first lesson; it does not indicate that nothing developed across the two cycles. Change was visible instead in overall engagement (Part B): medium in Lessons 1 and 2, the only two lessons with technical interruptions in loading or switching clips, and high in each of the remaining eight, including all five lessons of Cycle 2. This pattern suggests that the improvement the teacher perceived lay not in adding instructional components, present from the outset, but in the quality of the mediation delivered within them, and particularly in how far students were required to predict, discuss, and justify meanings before she confirmed them.

Table 2. Instructional implementation across the observed lessons

| No. | Observation focus                                                 | Representative evidence                                               |
|-----|-------------------------------------------------------------------|-----------------------------------------------------------------------|
| 1   | YouTube Shorts used as meaning-focused input                      | Lesson 1: a butterfly clip shown before the word was named.           |
| 2   | Teacher paused the video and asked guiding questions              | Lesson 10: paused at the waterfall; "Is the water going up or down?"  |
| 3   | Students attempted to infer word meaning from visuals             | Lesson 6: "riding a bike" called out before "cycle" was introduced.   |
| 4   | Guided-use activities (pair work, TPR, role-play)                 | Lesson 2: mimed sleeping in a sleeping bag; mime-and-guess in pairs.  |
| 5   | Students orally reused the target vocabulary                      | Lesson 8: "It's spicy" / "It's sweet" in choral and group repetition. |
| 6   | Teacher provided scaffolding when needed                          | Lesson 1: gestures and sentence frames on the board.                  |
| 7   | Brief attention to pronunciation/spelling after the meaning stage | Lesson 7: rhythmic clapping to practise "Bas-ket-ball."               |

The teacher perceived the videos as multimodal input that enabled students to infer word meanings before any

Vietnamese translation was supplied, thereby reducing reliance on translation. This stands against the translation-dominated vocabulary instruction documented in Vietnamese schools [3]. This finding is broadly consistent with previous research on multimodal and video-based vocabulary learning: Almurashi [5] and Kabooaha and Elyas [8] similarly report that video-based input increases learners' motivation, while Tragant Mestres and Pellicer-Sánchez [20] show that young learners distribute their attention unevenly across the verbal and visual channels of multimodal input. The present findings extend this line of research by suggesting that an ultra-short format such as tailored YouTube Shorts, and not only full-length video, can support this stage of meaning construction in a large, translation-oriented Vietnamese primary classroom, though the videos were always used within the Meaning-Use-Form sequence and alongside the teacher's deliberate withholding of translation.

#### 4.2. Teacher mediation in supporting meaning construction through the videos

Alongside her recognition of what multimodal input contributed, the notes reveal a consistent second perception: the videos alone were not sufficient to support students' meaning construction. Reflecting on Lesson 5, the teacher wrote, "The interaction was mostly teacher-led; I need to find ways for students to ask each other questions rather than just answering me." From then on, she gave hints rather than answers, encouraged students to infer meaning through discussion, and asked them to justify what they proposed. The checklists record concrete forms of scaffolding: gestures and sentence frames (Lesson 1); narrowing options with a choice question such as "Is it a tiger or a cheetah?" (Lesson 3); descriptive adjectives to support sentence-building (Lesson 5); directing attention to salient visual details (Lesson 7); gesture combined with repeated modelling when a student could not produce "sunny" (Lesson 9); and modelling pronunciation with hand movements (Lesson 10).

This perception, and the practice that followed from it, developed over the two cycles. In Cycle 1 the notes dwell mainly on what the videos did for the teacher's own instruction, chiefly reducing the time spent explaining in Vietnamese, while some students still waited for her direct explanation instead of attempting to infer meaning themselves (Lessons 1, 2, 4). She therefore adjusted Cycle 2 in three ways: more pause-

and-predict questions before confirming meaning, more structured pair and group interaction, and clearer sentence frames and scaffolding prompts. The main changes in the teacher's mediating practice from Cycle 1 to Cycle 2, together with supporting classroom evidence, are summarised in Table 3.

Table 3. The Development of Teacher's Mediating Practice between Cycle 1 and Cycle 2

| Issue in Cycle 1                                                 | Adjustment in Cycle 2                  | Evidence in Cycle 2                                                                    |
|------------------------------------------------------------------|----------------------------------------|----------------------------------------------------------------------------------------|
| Students waited for the teacher's explanation (Lessons 1, 2, 4). | More pause-and-predict questions.      | Independently named verbs (Lesson 6), weather/nature words (9-10) before confirmation. |
| Peer interaction limited to repeating model sentences (3, 5).    | More structured pair/group activities. | Asked/answered about video content (6-7); expressed preferences with support (8-10).   |
| Interaction mainly teacher-led; students hesitant (5).           | Clearer sentence frames and prompts.   | Longer, supported responses in guided speaking (8-10).                                 |

Despite these adjustments, several instructional challenges remained unresolved across the two cycles. Pronunciation difficulties (final consonants and consonant clusters, multisyllabic items, silent letters) were recorded in 9 of the 10 lessons; noise during choral repetition and pair work was recorded in all 10, despite a proposed stop signal such as a bell or clapping pattern; and students seated at the back or sides participated less in every lesson. The teacher's own position recurred as a concern: moving to the back focused the students seated there (after Lesson 3), while staying near the projector risked overlooking the far-back corners (after Lesson 6). A further difficulty emerged in Cycle 2, once tasks moved to extended output: several students could not produce full sentences without a model on the board (after Lesson 9). Other constraints included one clip judged "a bit fast" (Lesson 6) and screen glare affecting students at the sides (Lessons 5, 7). Collectively, these observations indicate that the development of the teacher's mediating practice was partial rather than complete.

The second finding is that the teacher did not regard YouTube Shorts alone as sufficient; in her account, active mediation through questioning, modelling, scaffolding, and feedback was a necessary condition for moving students from observation to comprehension. The teacher treated the videos less as a self-sufficient tool than as a source of input that required deliberate mediation before it became learning.

This is consistent with Vygotsky's [16] concept of the Zone of Proximal Development, in which the teacher acts as the more knowledgeable other, and with Wood, Bruner, and Ross's [18] concept of scaffolding, evident here in the teacher's flexible adjustment of support to students' actual needs. It also aligns with Tohamba's [17] observation that video-based instruction realises its potential when teachers actively guide learners' attention rather than using video as entertainment.

Her understanding of what the videos were for developed across the two cycles, from an initial focus in Cycle 1 on reducing her own explanation time to a stronger concern in Cycle 2 with creating conditions for students to propose meanings before confirmation. This shift was shaped progressively through practice and reflection, consistent with the nature of action research [21, 22].

In practice, this meant that the assumption of the benefits of the technology in terms of learning outcomes had to be questioned, as the result of the videos largely depended on how they were used pedagogically rather than on their properties. Given a 35-minute lesson and a class of 44 students in this public primary EFL context, distributing mediation evenly was a genuine challenge, as reflected in students seated farther away participating less in all 10 lessons.

## 5. RECOMMENDATIONS AND SUGGESTIONS

Based on the observations made in this study, the teacher recommends the following four suggestions:

First, YouTube Shorts should be used as input in the "Meaning" stage of a structured sequence (e.g., MUF) and not used as a stand-alone activity. "Use" and "Form" should follow in pair work, TPR, and pronunciation exercises.

Second, teachers should make deliberate moves towards mediation by pausing the video, asking pause-and-predict questions, and asking students to give reasons before confirming the meaning. As suggested by Cycle 2, this appears to help students shift from waiting for explanation towards active inference.

Third, in large classes with short lessons, teachers should distribute mediation more evenly among students seated further away, for instance by varying their own position or using a turn taking device.

Fourth, systematic reflection after each lesson and cycle offers a practical route to developing mediating skills; in this study it was reflection on classroom

evidence, not the videos themselves, that produced the changes between cycles.

These recommendations derive from one teacher's practice under an action research design, not a large sample or a comparison group, and are offered as practical suggestions for comparable contexts rather than generalisable prescriptions.

## 6. CONCLUSION

This study explored one teacher's perspectives on using tailored YouTube Shorts to teach English vocabulary meanings to fifth graders in Hai Phong, through a two-cycle action research design drawing on classroom observation, video recordings, and field notes.

Tailored YouTube Shorts functioned, from the teacher's perspective, as a valuable multimodal mediational tool that supported the meaning-construction stage and helped reduce reliance on Vietnamese translation. That value, however, did not come from the video itself: it came from how the teacher mediated its use through questioning, modelling, scaffolding, and feedback within a structured Meaning-Use-Form sequence. Neither the perception nor the practice was fixed; both were refined progressively through systematic reflection across the two cycles.

The teacher perceived a gap between "watching" and "learning": viewing a video alone did not necessarily lead students to demonstrate an understanding of word meaning; such understanding was supported through her active mediation.

The study's contribution is concrete, practice-based evidence of the teacher's mediating role when an under-researched format - ultra-short clips of 10 - 15 seconds - is integrated into vocabulary teaching in a large-class, short-lesson public primary EFL classroom in Vietnam. Theoretically, it specifies the role of ultra-short video within one stage of the vocabulary-teaching sequence rather than treating video as a general solution; practically, it offers primary teachers concrete guidance on mediating short video purposefully.

Several limitations of the present study should be acknowledged when interpreting these findings. The study reflects one teacher's perspectives in one classroom over 10 weeks. Her dual role may have shaped how observations were interpreted, though video recording, field notes, and supervisor consultation strengthened trustworthiness. Finally, students' vocabulary outcomes were not measured with an objective instrument, so findings should be read as the

teacher's perceptions and interpretations, not direct evidence of learning outcomes.

Future research could examine several teachers across classrooms or schools, or combine teacher-perception data with objective vocabulary measures, to clarify the relationship between mediating practice and actual learning. Given the study's focus on one teacher and one classroom, the findings are context-specific and should be interpreted as practical insights for comparable settings rather than as generalisable conclusions.

## REFERENCES

- [1]. Alqahtani M., "The importance of vocabulary in language learning and how to be taught," *International Journal of Teaching and Education*, 3(3), 21-34, 2015. doi: 10.20472/TE.2015.3.3.002.
- [2]. Nation I. S. P., *Learning Vocabulary in Another Language*, 3rd ed. Cambridge University Press, Cambridge, 2022.
- [3]. Vu D. V., Peters E., "Vocabulary in English language learning, teaching, and testing in Vietnam: A review," *Education Sciences*, 11(9), Article 563, 2021. doi: 10.3390/educsci11090563.
- [4]. Truong P. H., Kim A. D., "The influence of TikTok on young generation in Vietnam," in *Proceedings of the 10th European Conference on Social Media (ECSM 2023)*, 310-317, 18-19 May 2023. doi: 10.34190/ecsm.10.1.1151.
- [5]. Almurashi W. A., "The effective use of YouTube videos for teaching English language in classrooms as supplementary material at Taibah University in Alula," *International Journal of English Language and Linguistics Research*, 4(3), 32-47, 2016.
- [6]. Juhana, "Teaching English to young learners: Some points to be considered," *Asian Journal of Education and e-Learning*, 2(1), 43-46, 2014.
- [7]. Linse C., *Practical English Language Teaching: Young Learners*. McGraw-Hill, New York, 2005.
- [8]. Kaboocha R., Elyas T., "The impacts of using YouTube videos on learning vocabulary in Saudi EFL classrooms," in *ICERI2015: 8th International Conference of Education, Research and Innovation*, 3525-3531, 16-18 Nov. 2015.
- [9]. Zaidi A., Awaludin F. A., Abd Karim R., Che Ghani N. F., Abdul Rani M. S., Ibrahim N., "University students' perceptions of YouTube usage in (ESL) classrooms," *International Journal of Academic Research in Business and Social Sciences*, 8(1), 534-545, 2018. doi: 10.6007/IJARBS/v8-i1/3826.
- [10]. Lo S., "Vocabulary learning through viewing dual-subtitled videos: Immediate repetition versus spaced repetition as an enhancement strategy," *ReCALL*, 36(2), 152-167, 2024. doi: 10.1017/S0958344024000053.
- [11]. Do T. T. T., Sellars M., Le T. T., "Primary English language education policy in Vietnam's disadvantaged areas: Implementation barriers," *Education Sciences*, 12(7), Article 445, 2022. doi: 10.3390/educsci12070445.

- [12]. Cameron L., *Teaching Languages to Young Learners*. Cambridge University Press, Cambridge, 2001.
- [13]. Long M. H., "Focus on form: A design feature in language teaching methodology," in de Bot K., Ginsberg R., Kramsch C. (Eds.), *Foreign Language Research in Cross-Cultural Perspective*. John Benjamins, Amsterdam, 39-52, 1991.
- [14]. Willis J., *A Framework for Task-Based Learning*. Longman, Harlow, 1996.
- [15]. Larsen-Freeman D., *Teaching Language: From Grammar to Grammaticizing*. Heinle & Heinle, Boston, 2003.
- [16]. Vygotsky L. S., *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge, MA, 1978.
- [17]. Tohamba C. P. P., "The use of YouTube in improving English vocabulary mastery in early childhood: Systematic literature review," *Journal of Instructional and Development Researches*, 5(5), 500-510, 2025. doi: 10.53621/jider.v5i5.595.
- [18]. Wood D., Bruner J. S., Ross G., "The role of tutoring in problem solving," *Journal of Child Psychology and Psychiatry*, 17(2), 89-100, 1976. doi: 10.1111/j.1469-7610.1976.tb00381.x.
- [19]. Lutviana R., Mafulah S., "The use of video and TPR to improve students' vocabulary mastery," *EnJourMe (English Journal of Merdeka)*, 2(2), 89-97, 2017. doi: 10.26905/enjourme.v2i2.1970.
- [20]. Tragant Mestres E., Pellicer-Sánchez A., "Young EFL learners' processing of multimodal input: Examining learners' eye movements," *System*, 80, 212-223, 2019. doi: 10.1016/j.system.2018.12.002.
- [21]. Kemmis S., McTaggart R., *The Action Research Planner*, 3rd ed. Deakin University Press, Victoria, 1988.
- [22]. Burns A., *Doing Action Research in English Language Teaching: A Guide for Practitioners*. Routledge, New York, 2010.
- [23]. Tarusha F., Bushi J., "The role of classroom observation, its impact on improving teacher's teaching practices," *European Journal of Theoretical and Applied Sciences*, 2(2), 718-723, 2024. doi: 10.59324/ejtas.2024.2(2).63.
- [24]. Braun V., Clarke V., "Using thematic analysis in psychology," *Qualitative Research in Psychology*, 3(2), 77-101, 2006. doi: 10.1191/1478088706qp063oa.

---

#### THÔNG TIN TÁC GIẢ

**Nguyễn Thị Hương<sup>1</sup>, Bùi Nguyên Khánh<sup>1</sup>, Phạm Xuân Quỳnh<sup>2</sup>**

<sup>1</sup>Trường Ngoại ngữ - Du lịch, Đại học Công nghiệp Hà Nội

<sup>2</sup>Trường Tiểu học Lai Cách, thành phố Hải Phòng