

Beyond Repetition: Multimodal and Social Mediation of Vocabulary Learning through Minecraft Gameplay Videos among Young EFL Learners

Aulia Kumalasari S Putri¹, Joseph Ernest Mambu¹

¹Universitas Kristen Satya Wacana, Indonesia

Received: August 4, 2025 | Revised: July 14, 2026 | Accepted: August 5, 2026

Correspondence: auliakumalasari010@gmail.com

Abstract

Although digital game-based language learning has received increasing attention, how watching gameplay videos creates vocabulary-learning opportunities for young EFL learners at home remains underexplored. This study examined how subtitled Minecraft gameplay videos on YouTube and family and peer interaction mediated children's engagement with English vocabulary. A qualitative case study with descriptive corpus support tracked the regular home viewing of two siblings aged seven and eight for 30 days. Data comprised viewing histories, observational notes, video subtitles, and interviews with the children and their parents. AntConc analysis identified 50 frequently occurring lexical items, which informed a subsequent vocabulary-elicitation interview. Triangulation revealed that, across the interviews, the children recognized or contextually explained 31 target items. Recognition did not consistently correspond to raw frequency: words linked to visually salient actions, game objectives, emotional moments, or familiar experiences were more memorable than some highly recurrent but contextually less distinctive items. Subtitles appeared to connect spoken and written forms, while gameplay visuals supported meaning inference. Co-viewing, sibling or peer discussion, and the reuse of game-related expressions during play extended engagement beyond the screen. The findings suggest that the vocabulary-learning affordances of gameplay videos arise not from repetition alone but from the interaction of multimodal salience, learner interest, and social mediation. Given the two-case design and absence of baseline testing, these results indicate context-specific learning opportunities rather than causal effects. The study positions guided co-viewing and contextual discussion as promising ways to enrich informal digital EFL learning at home.

Keywords: extramural English; informal digital learning; Minecraft; multimodal learning; vocabulary recognition; young EFL learners

1. Introduction

Vocabulary knowledge is central to young EFL learners' communicative development because it supports comprehension and production across listening, speaking, reading, and writing (Janebi Enayat & Derakhshan, 2021; Kyle & Crossley, 2016; Zhang & Zhang, 2022). Its development depends not only on explicit instruction but also on repeated encounters with words across varied and meaningful contexts, conditions that classroom instruction may not consistently provide, particularly where opportunities to use English in everyday life are limited (Muñoz & Cadierno, 2021; Peters et al., 2019). The growing presence of YouTube, digital games, social media, and other online platforms in children's lives has therefore made informal or extramural digital activity an important source of contact with English beyond school (De Wilde et al., 2020; Jensen, 2017; Leona et al., 2021; Pires et al., 2021). Gameplay videos are especially relevant within this ecology because they can coordinate spoken commentary and on-screen text with visible objects, purposeful actions, and unfolding events. Such coordination may allow viewers to encounter lexical forms in contexts that provide immediate clues to meaning, while the entertainment value of the videos may sustain voluntary engagement over time (Lee & Dressman, 2018; Montero Perez et al., 2018; Peters, 2019). For young EFL learners who regularly watch subtitled Minecraft gameplay videos at home, vocabulary encounters may thus occur through an everyday media practice rather than a planned instructional activity. Nevertheless, the learning opportunities associated with watching gameplay cannot simply

be inferred from research on playing digital games because viewing and playing organize learner agency, feedback, and linguistic input in substantially different ways.

Research on digital games and extramural English has shown that gaming can expose learners to recurring and specialized lexical items in goal-directed contexts in which spoken and written language is coordinated with visual objects, player actions, objectives, feedback, and social interaction (Acquah & Katz, 2020; Chien, 2019; Toh & Lim, 2022). Among children and adolescents, greater engagement with English-mediated out-of-school gaming has repeatedly been associated with larger receptive or productive vocabularies and, in some studies, broader English proficiency, although the strength of these relationships varies according to gaming time, game type, input modality, and learner group (Jensen, 2017; Sundqvist, 2019; Sundqvist & Wikström, 2015). These findings suggest that interest-driven gaming can complement classroom learning by extending contact with English into self-selected activities in which reading instructions, solving problems, and coordinating actions give language an immediate functional purpose (Scholz, 2017; Sundqvist, 2019). The evidence base, however, remains predominantly player-centered. Major reviews principally synthesize active gameplay and indicate that formal interventions have often used custom-built or educational games, with university learners common in the broader literature and formal school settings dominant in studies involving younger participants (Hanghøj et al., 2022; Hung et al., 2018; Xu et al., 2020). A related program of controlled research distinguished "watchers" from players, but the watchers observed two classmates playing during scheduled weekly treatments over five weeks rather than independently selecting creator-produced gameplay videos for home viewing (Ebrahimzadeh, 2017; Ebrahimzadeh & Alavi, 2016, 2017). Active players make decisions and receive feedback contingent on their actions, whereas gameplay-video viewers encounter a preselected and often edited sequence designed for an audience, in which commentary, subtitles or other on-screen language, visual action, and game events are coordinated by the content creator (Burwell & Miller, 2016; Shoda, 2022; Stuft & von Gillern, 2021; Fievez et al., 2020). Gameplay-video viewing therefore merits investigation as a distinct form of multimodal informal English exposure rather than as a reduced version of active play.

The theoretical distinctiveness of gameplay-video viewing lies in this creator-mediated configuration. Let's Play scholarship characterizes gameplay videos as hybrid game-video texts in which recorded play is reshaped through commentary, performance, editing, and platform conventions (Dizon, 2024; Nguyen, 2016; Piittinen, 2018). Viewers of prerecorded videos do not determine the represented in-game decisions or receive feedback contingent on their own inputs. Instead, they interpret visible objects, character movements, goal-directed actions, affective reactions, and outcomes alongside spoken commentary and written language (Gekker, 2018; Piittinen, 2018; Shoda, 2022). The cognitive theory of multimedia learning proposes that learning from coordinated verbal and visual information involves selecting relevant words and images, organizing them into coherent representations, and integrating those representations with one another and with prior knowledge. Research on spatial and temporal contiguity likewise supports coordinating corresponding sources of information, while also showing that the benefits depend on how information is presented and on learner characteristics (Liu et al., 2022; Mayer, 2017, 2024; Schroeder & Cenkcı, 2018). Captioning research provides more specific evidence that displaying spoken L2 input in written form can direct attention to lexical forms and support incidental recognition or learning of word forms and meanings, although outcomes vary according to prior vocabulary knowledge, word frequency, image-word congruence, and captioning format (Montero Perez et al., 2015, 2018; Peters, 2019). Minecraft offers a particularly useful setting in which to examine these processes because recurring activities such as building, crafting, collecting, navigating, and surviving connect language with visible entities and consequential actions. Minecraft-related input also includes both high-frequency vocabulary and less frequent game-specific terms denoting blocks, items, actions, and environmental features (Ståhl et al., 2025; Weisi & Hajizadeh, 2025; Zheng et al., 2015). Although viewers lack direct control over the game, interpreting these coordinated cues can still be cognitively active. The relevant question is therefore not whether viewing is categorically active or passive, but how the particular configuration of spoken, written, visual, and situational information creates opportunities for children to notice and interpret vocabulary.

Within this multimodal configuration, lexical frequency and multimodal salience should be treated as related but analytically distinct. Repeated encounters can strengthen familiarity with a word and provide additional occasions for form-meaning mapping, yet frequency in the video corpus represents the amount of exposure available rather than direct evidence that a child attended to, recognized, or learned a word (Frances et al., 2020;

Peters et al., 2019; Uchihara et al., 2019). Research on captioned audiovisual input shows that the contribution of repetition is conditioned by factors such as prior vocabulary knowledge, captioning format, and the semantic congruence between a lexical item and the accompanying image (Montero Perez et al., 2018; Peters, 2019). In gameplay videos, a word may therefore become especially noticeable when its spoken and written forms coincide with a clearly visible object or action, when it is relevant to an immediate game objective, or when it occurs within a distinctive and memorable episode. Conversely, a highly recurrent item may remain weakly attended to if it has no clear visual referent or salient function in the unfolding activity. This distinction is important for the present study because a frequency list can describe the lexical environment offered by the videos, but only evidence from the children can indicate which items they recognized or could explain in relation to a context. Combining corpus-derived frequency information with vocabulary elicitation thus makes it possible to investigate whether the most frequently available words were necessarily those most successfully recognized and to consider how multimodal context may help explain any divergence.

Subtitles and social interaction add two complementary forms of mediation to this account. Subtitles foreground orthographic forms within the audiovisual stream, whereas co-viewers and co-players may direct attention, elaborate meanings, and connect events on the screen with shared experiences and subsequent activities (Lantolf & Xi, 2023; Peters, 2019; Taylor et al., 2024). L2 captioning research, including studies involving primary-school learners, indicates that temporally aligned written forms can increase attention to target words and support word-form recognition and form-meaning learning, particularly when recurring words are accompanied by semantically informative imagery (Montero Perez et al., 2015; Peters, 2019; Teng, 2019; Vanderplank, 2016). These affordances are conditional rather than automatic: outcomes vary with learners' prior knowledge, the frequency and type of target vocabulary, captioning format, and image-word congruence, and most available evidence concerns general captioned audiovisual materials rather than children's self-selected gameplay videos at home (Montero Perez et al., 2018; Pujadas & Muñoz, 2019; Teng, 2019). Vocabulary engagement also extends beyond the relationship between an individual viewer and the screen. Developmental-media research indicates that structured or responsive adult co-viewing can guide attention and support some aspects of word learning, although effects vary across learners and outcome measures and the average advantage of adult-child co-use appears modest (Neuman et al., 2020; Samudra et al., 2019; Strouse et al., 2018). Research on sibling mediation and children's Minecraft practices further demonstrates that siblings and peers can act as mentors, co-players, problem-solving partners, and audiences for game-related knowledge across online and face-to-face settings (Adler & Elias, 2025; Dezuanni et al., 2015; Wernholm, 2021). These interactions can create occasions to ask questions, explain meanings, repeat expressions, and reuse language during later game-related play. From a sociocultural perspective, vocabulary engagement is therefore mediated jointly by culturally organized signs and digital artifacts and by assistance from other people, rather than resulting from exposure alone (Lantolf & Poehner, 2023; Lantolf & Xi, 2023; Lantolf et al., 2021; Xi & Lantolf, 2021). Subtitles, gameplay imagery, family assistance, peer discussion, and shared play can consequently be examined as an interconnected configuration through which children notice, interpret, and recontextualize English vocabulary in the home.

Despite these converging lines of research, limited attention has been devoted to how young EFL learners engage with vocabulary while voluntarily watching subtitled gameplay videos in natural home environments. Existing studies have focused primarily on active gameplay, older learners, formal interventions, or general captioned audiovisual materials. Consequently, comparatively little is known about how lexical frequency interacts with multimodal salience, subtitles, learner interest, and family or peer interaction in children's everyday gameplay-video viewing. Research has also rarely combined a corpus-based description of the lexical input available in such videos with qualitative evidence of children's vocabulary recognition, contextual interpretation, and socially mediated engagement. To address these gaps, the present qualitative case study, supported by descriptive corpus analysis, followed two young EFL learners during 30 days of regular home viewing of subtitled Minecraft gameplay videos. Specifically, it investigated which English lexical items occurred frequently in the videos, how the children recognized and contextually explained the selected target items, and how the videos' multimodal features and family or peer interaction mediated their engagement with English vocabulary. By distinguishing gameplay-video viewing from active gameplay and relating corpus frequency to the children's demonstrated recognition and contextual interpretations, the study extends research on extramural English to a form of digital activity that is common in children's everyday lives but remains insufficiently examined. It also

advances a socially situated account of informal vocabulary engagement in which repetition operates alongside multimodal salience and interpersonal mediation. Because the study did not include baseline testing or a comparison group, however, its contribution lies in identifying context-specific vocabulary-learning opportunities and mediational processes rather than establishing causal vocabulary gains.

2. Method

2.1 Research Design

This study adopted a qualitatively driven single-case study design, supplemented by descriptive corpus analysis, to examine how young EFL learners engaged with English vocabulary while watching subtitled *Minecraft* gameplay videos at home. The case was defined as the children's gameplay-video viewing within a single family and was bounded by two sibling participants and a 30-day observation period. A case study design was appropriate because it enabled an intensive, context-sensitive examination of the phenomenon in its natural setting using multiple sources of evidence (Harrison et al., 2017; Rashid et al., 2019; Yazan, 2015). Observations and interviews constituted the primary data sources, providing qualitative evidence of the children's vocabulary recognition, contextual interpretations, and engagement mediated through family interaction. The subtitle corpus played a complementary descriptive role: frequency analysis mapped the recurring lexical input available in the videos and informed the selection of target words for the vocabulary-elicitation interviews. Accordingly, frequency counts were not treated as evidence of attention, recognition, or vocabulary acquisition; rather, they were used to characterize the linguistic input and establish a systematic link between the video content and the children's qualitative responses.

2.2 Participants and Research Setting

The participants were two elementary-school-aged siblings, aged seven and eight, who were purposively selected because they regularly watched *Minecraft* gameplay videos on YouTube, were familiar with the game, and routinely engaged with digital media at home. In qualitative inquiry, purposive sampling emphasizes participants' direct relevance to the phenomenon under investigation rather than statistical representativeness (Gentles et al., 2015). The siblings were therefore considered information-rich participants, while their shared gameplay-video-viewing practices within a single household formed the bounded case; they were not treated as two separate cases. Data collection took place in the family home to minimize disruption to the children's usual viewing routines and to capture family-mediated interactions within their everyday context. Examining children's digital practices at home can reveal situated behaviors and interpersonal interactions that surveys or decontextualized tasks may overlook. At the same time, home-based research requires appropriate safeguards for participants' privacy and reflexive consideration of how researcher presence and family relationships may influence the practices being observed (Sandberg & Gillen, 2021).

2.3 Data Sources and Collection

Data were collected over 30 consecutive days without imposing a researcher-designed viewing program. Instead, the study examined subtitled *Minecraft* gameplay videos that the children watched as part of their ordinary YouTube routines. Videos were eligible for inclusion if they were viewed during the observation period, contained English-language gameplay commentary and subtitles, and were appropriate for the children's ages. Four complementary sources of data were collected. First, the children's YouTube viewing history was reviewed to establish a record of the videos watched and identify those meeting the inclusion criteria. Second, observational field notes documented the children's responses to vocabulary, spontaneous use of English, interpretations of gameplay events, and interactions with siblings or other family members during or after viewing. Third, the English subtitles from the eligible videos were compiled into a small, purpose-built corpus representing the lexical input available in the sampled content. Fourth, at the end of the observation period, each child participated separately in a structured, age-appropriate vocabulary-elicitation interview.

The interview protocol comprised 50 recurrent content words selected through frequency analysis of the subtitle corpus, including action verbs, descriptive adjectives, and game-related nouns. For each target word, the children were asked whether they recognized it, what they believed it meant, and whether they could describe a gameplay situation in which they had encountered it. Both participants received the same core prompts to ensure consistency, although the researcher provided age-appropriate clarification when needed. Using a consistent

interview protocol helped maintain alignment between the elicitation questions and the study's research questions (Castillo-Montoya, 2016). Because interviewing children requires accessible language and careful attention to differences in power, communication, and perceived adult expectations, the researcher avoided indicating preferred answers and allowed the participants sufficient time to respond in their own words (Kutrovátz, 2017). The responses were analyzed as evidence of the children's vocabulary recognition and contextual interpretation at the time of elicitation. They were not regarded as conclusive evidence that the words had been acquired exclusively through watching the videos, because the study did not include a pretest, comparison group, or delayed retention measure. Accordingly, the interviews provided context-specific evidence of recognition and meaning-making rather than a causal measure of vocabulary acquisition. The interviews were conducted in Indonesian to ensure the children could respond comfortably in their first language. All interview excerpts presented in this article were translated into English by the researcher for publication while preserving their original meanings.

2.4 Data Analysis

The subtitle files from all eligible videos were compiled into a small, purpose-built corpus and analyzed using AntConc version 4.3.1 (Anthony, 2024). The Word List tool was used to generate raw frequency rankings, after which the Concordance tool was used to examine recurrent lexical items in their immediate contexts and determine how they functioned within the gameplay commentary. Considering concordance evidence alongside frequency counts reduced the risk of interpreting lexical relevance solely on the basis of numerical prominence (Touri & Koteyko, 2015). The researcher subsequently identified recurrent content words and organized them into analytically relevant categories, including action verbs, descriptive adjectives, and game-specific nouns. Each frequency count was interpreted only as the number of occurrences, and therefore potential encounters, within the sampled subtitles. It was not treated as evidence that the children had attended to, understood, or learned the word. The corpus findings informed the selection of 50 target words for the vocabulary-elicitation interviews, while the concordance lines provided contextual information for interpreting the children's responses.

The observational field notes and interview responses were analyzed through repeated reading and iterative coding. Three broad dimensions derived from the research questions guided the initial analysis: vocabulary recognition, contextual interpretation, and socially mediated engagement. Within these dimensions, the codes were refined, combined, or expanded as additional patterns emerged from the data. The coded evidence was examined first within each participant and then compared across the two siblings to identify shared and differing patterns while retaining the study's treatment of the household as a single bounded case. Throughout the analysis, a transparent connection was maintained among the research questions, analytic codes, supporting evidence, and resulting interpretations to strengthen the credibility of the qualitative findings (Nowell et al., 2017).

Finally, a case-oriented analytical matrix was used to integrate evidence from the four data sources. The viewing history established which videos formed part of the children's regular exposure; the subtitle corpus characterized the recurring lexical input available in those videos; the observational notes documented vocabulary-related behavior and family interaction; and the interviews elicited evidence of vocabulary recognition and contextual interpretation. Cross-source triangulation was used to identify convergence, complementarity, and divergence among these forms of evidence (Fusch et al., 2018). Agreement among the sources was not treated as objective confirmation that vocabulary had been acquired through video viewing. Instead, triangulation supported a more contextualized interpretation of how the available lexical input, the children's responses, and their social interactions were related within this particular case.

2.5 Ethical Considerations and Methodological Scope

Written informed consent was obtained from the children's parent or legal guardian before data collection commenced. The study examined only *Minecraft* gameplay videos that formed part of the children's ordinary, parent-supervised viewing and did not direct them to access additional or unfamiliar online content. To protect their privacy, the children's names, YouTube account information, and all other potentially identifying details were excluded from the analytical dataset and research report. These safeguards were particularly important because qualitative research involving children in domestic and digital-media environments raises interconnected concerns regarding child vulnerability, online privacy, family relationships, and researcher access to everyday household practices (Kutrovátz, 2017; Sandberg & Gillen, 2021). Consistent with its qualitative single-

case design, purposive selection of two participants, and absence of experimental controls, the study was intended to provide an in-depth, contextualized account of vocabulary exposure, recognition, interpretation, and socially mediated engagement within one family setting. Its findings were therefore not intended to be statistically generalized to broader populations or interpreted as causal evidence that watching gameplay videos resulted in vocabulary acquisition.

3. Results

This study examined how two siblings, aged seven and eight, engaged with English vocabulary while watching subtitled Minecraft gameplay videos on YouTube. Observational field notes, descriptive corpus analysis using AntConc, and individual structured follow-up interviews were combined to explore the relationship between recurring lexical input in the videos and the children's recognition and contextual interpretation of selected words. The corpus analysis focused on recurrent gameplay-related content words, particularly action verbs and nouns. Raw frequency rankings were generated from the subtitles of the videos watched during the study period, and eligible items were selected to form a 50-word interview list. These included verbs such as go, build, and jump and nouns such as block, sword, and zombie. The adjective new, which appeared in the initial frequency results, was excluded from the verb-noun categorization used in the analysis. Importantly, frequency was interpreted as the number of times a word occurred in the sampled subtitles and, therefore, the number of potential encounters available during viewing. It was not treated as evidence that the children necessarily noticed, understood, or learned the word.

Table 1 presents the recurrent English verbs and nouns identified in the subtitle corpus. Frequency refers to the raw number of times each word occurred across the subtitle transcripts of the Minecraft gameplay videos watched during the study period. These counts characterize the lexical input available to the children rather than directly measuring their attention to or knowledge of individual words. The items are organized into verb and noun categories, while the color coding indicates their recognition status based on the children's responses during the follow-up interviews. This presentation enables comparison between the frequency of words in the available input and the children's demonstrated recognition of those words.

Table 1
Frequently Occurring Verbs and Nouns Identified Through AntConc Analysis

No.	Verb	Frequency	Noun	Frequency
1.	Get	445	Music	244
2.	Go	333	Minecraft	179
3.	See	290	Tutorial	144
4.	Let	289	Block	129
5.	Know	285	Mob	93
6.	Think	203	Water	90
7.	Make	159	Place	94
8.	Want	159	Game	62
9.	Come	156	Villager	54
10.	Did	147	World	50
11.	Take	85	House	38
12.	Find	101	Item	39
13.	Put	78	Zombie	59
14.	Use	78	Diamond	44
15.	Give	79	Sword	25
16.	Say	64	Fire	20
17.	Try	66	Time	20
18.	Run	20	Craft	44
19.	Move	22		
20.	Play	24		

No.	Verb	Frequency	Noun	Frequency
21	Build	72		
22	Doing	72		
23	Help	63		
24	Start	35		
25	Break	48		
26	Stop	29		
27	Spawn	67		
28	Keep	35		
29	Leave	21		
30	Open	21		
31	Test	24		

The color coding in Table 1 distinguishes vocabulary items coded as recognized from those not recognized during the structured interviews, thereby facilitating comparison between word frequency in the subtitle corpus and the children's elicited vocabulary recognition. Highlighted items indicate words coded as recognized based on the participants' responses to prompts asking them to explain the meaning of each word, use it appropriately, or relate it to a relevant gameplay situation. Because the interviews were conducted after the observation period, the responses provide evidence of the children's vocabulary recognition and contextual understanding at the time of elicitation. They do not, however, establish that the words were acquired solely through watching the videos, as the influence of prior knowledge and other sources of English exposure cannot be excluded. Comparing subtitle frequency with interview responses therefore reveals whether recurrent lexical items available in the sampled videos corresponded with the vocabulary recognized by the children. This comparison provides insight into the potential vocabulary-learning opportunities associated with informal gameplay-video viewing without treating either word frequency or interview recognition as proof of vocabulary acquisition.

Table 2
Vocabulary Items Recognized by the Participants in the Interview

No.	Verb	Noun
1.	Get	Music
2.	Go	Minecraft
3.	See	Block
4.	Let	Time
5.	Know	Game
6.	Think	Fire
7.	Move	World
8.	Want	Water
9.	Play	Place
10.	Build	Craft
11.	Open	House
12.	Start	Villager
13.	Stop	Zombie
14.	Use	Diamond
15.	Give	
16.	Say	
17.	Run	

As shown in Table 2, however, the participants' recognition of these words did not follow their frequency rankings. They recognized a combination of high- and mid-frequency words, while several highly frequent words remained unrecognized during the interviews. Thus, frequency of occurrence alone did not consistently correspond with vocabulary recognition. In explaining the words they recognized, the participants frequently connected them with visible actions, memorable gameplay events, or familiar experiences outside the videos

3.1 Vocabulary Recognition and Contextual Explanation

The interview responses showed that the participants recognized and contextually explained a range of vocabulary associated with *Minecraft*. Recognized words included *zombie*, *diamond*, *craft*, *sword*, *item*, *move*, *build*, *use*, and *jump*. Rather than explaining these words through formal definitions, the children usually connected them with visible objects, character actions, game instructions, or memorable events. Words such as *zombie* and *sword* were associated with fighting scenes, whereas *diamond*, *item*, and *craft* were explained in relation to objects or activities within the game. Similarly, *build*, *move*, *use*, and *jump* were connected with actions performed by characters or instructions heard during gameplay. Expressions such as “*build a house*” were recalled as complete action-based phrases rather than as isolated lexical items (INT-C1; INT-C2).

Recognition did not consistently correspond to the frequency ranking generated from the subtitle corpus. Some highly frequent words, particularly *get* and *let*, were not readily recognized or explained during the interviews. In contrast, less frequent but game-specific words, including *craft*, *sword*, *zombie*, and *diamond*, were recognized and linked to particular objects, objectives, or gameplay situations. The participants’ responses therefore showed that the words occurring most frequently in the subtitles were not necessarily those they could explain most clearly. Words connected with identifiable game objects and actions appeared more prominently in their interview explanations than several high-frequency general verbs. The participants also drew on experiences beyond the gameplay videos when interpreting particular words and expressions. One child connected *go* with a command used by a taekwondo coach, while the other associated “*let’s go*” with an expression frequently used by their mother (INT-C1; INT-C2). These responses indicated that their recognition of English vocabulary was informed by multiple everyday encounters rather than by YouTube viewing alone. The children also reported noticing words in both spoken commentary and subtitles, allowing them to refer to the sound and written form of some expressions when explaining how they recognized them.

3.2 Assistance and Co-viewing with Siblings and Friends

The interviews indicated that watching *Minecraft* videos was not always an individual activity. Both participants described seeking assistance, co-viewing, and discussing video content with siblings or friends. One child explained, “*I ask my brother or friend. But sometimes we watch YouTube together.*” (INT-C1). This response showed that siblings and friends served both as sources of assistance and as viewing companions. The same participant further described how a sibling provided explanations during viewing: “*I watch with my brother! He tells me what is happening.*” (INT-C1). Assistance therefore involved more than simply watching together; it also included explanations of events that the child did not immediately understand. The second participant emphasized the social and entertaining aspects of watching with a peer: “*With my friend! We like the funny parts.*” (INT-C2). Although the two focal participants did not usually watch the selected videos together, both reported sharing *Minecraft*-related viewing experiences with other children. Their conversations concerned memorable events, characters, and actions, including fighting zombies, finding diamonds, using swords, and constructing houses. They also reported repeating expressions such as “*build a house*,” “*jump*,” and “*run*” while playing or talking about the game (INT-C1; INT-C2). The interview evidence thus documented several forms of social engagement: asking questions, receiving explanations, watching together, discussing entertaining scenes, and reusing game-related vocabulary during play.

3.3 Perceived Familiarity with Spoken English and Video Content

During the vocabulary-elicitation interviews, the participants explained words such as *build*, *jump*, *sword*, and *diamond* without requiring translations or additional explanatory prompts. Their explanations referred to the corresponding actions, objects, or events in *Minecraft*, demonstrating recognition and contextual understanding at the time of the interviews. One participant also reported that pausing the videos had become less necessary because many recurring words were already familiar (INT-C1). This statement represented the participant’s perception of increased familiarity and was not treated as an independently measured improvement in listening comprehension. The children also described how they navigated English-language video content. They reported selecting *Minecraft* videos, using video menus, replaying tutorials, and returning to particular sections when they wanted to see an action again. Replying was especially associated with tutorials and gameplay sequences that they considered interesting or useful. These interview accounts showed that the children actively selected and revisited content rather than merely watching each video once from beginning to end (INT-C1; INT-C2).

3.4 Vocabulary Use and Parental Involvement at Home

The parent interview supplemented the children's accounts by describing their use of English during everyday activities. The parent reported hearing the children say words and expressions such as "*build a house*," "*jump*," and "*run*" while playing or reenacting events from *Minecraft* (INT-P1). According to the parent, the children sometimes transferred expressions heard in the videos into conversations and game-related play outside the immediate viewing activity. Parental involvement was described primarily in terms of supervision and occasional co-viewing. The parent explained, "*Sometimes, we watch with them to understand what they are learning.*" (INT-P1). The parent also reported initially being concerned about the amount of time the children spent watching online videos but later noticing their use of English words during play. These accounts indicated that parental participation included regulating access, observing the content, occasionally joining the children during viewing, and noticing their subsequent language use.

4. Discussion

This study examined how vocabulary occurring in subtitled *Minecraft* gameplay videos corresponded with the vocabulary recognized and contextually explained by two young EFL learners in a family environment. Four interconnected findings emerged. First, the videos provided recurrent exposure to general verbs and game-related nouns, but frequency did not consistently correspond with recognition. Second, words associated with visible objects, actions, and meaningful gameplay events appeared more accessible than several highly frequent general words. Third, the children actively selected, replayed, and discussed the content as part of their informal digital activities. Finally, siblings, friends, and parents contributed differently to the social environment surrounding the viewing. These findings suggest that vocabulary-learning opportunities arose through the interaction of repetition, contextual salience, multimodal representation, prior familiarity, learner interest, and social mediation.

4.1 Frequency of Exposure and Selective Vocabulary Recognition

The subtitle analysis revealed a combination of common English verbs, such as *get*, *go*, *see*, *let*, and *know*, and game-related nouns, including *block*, *mob*, *diamond*, *sword*, and *zombie*. This lexical profile reflects the distinctive language of *Minecraft* gameplay discourse, in which general English is combined with vocabulary referring to game objects, actions, mechanics, and objectives (Chien, 2019). The videos therefore provided potential exposure to both broadly useful vocabulary and specialized words that may rarely appear in the children's formal English lessons. However, the interview findings showed that the participants did not recognize the words according to their frequency rankings. Several highly frequent words, particularly *get* and *let*, were not readily explained, whereas less frequent words such as *sword*, *diamond*, *zombie*, and *craft* were connected with identifiable objects or actions. Frequency consequently represented the availability of lexical input rather than a direct measure of vocabulary uptake. Although repeated encounters generally increase the likelihood of incidental vocabulary learning, the relationship is probabilistic rather than automatic (Teng, 2019; Uchihara et al., 2019).

The contrast between frequency and recognition may partly reflect differences in word characteristics and contextual support. General verbs can carry multiple meanings and occur in varied grammatical constructions, making them difficult to associate with one observable event. Game-specific nouns, by contrast, often possess stable visual referents, while verbs such as *craft* and *build* describe central gameplay activities. Research on audiovisual vocabulary learning similarly indicates that frequency interacts with prior knowledge, visual representation, lexical characteristics, and the clarity of the surrounding context (Fievez et al., 2020). Thus, a less frequent but visually salient word may be more readily interpreted than a frequent word whose meaning varies across situations. The functional importance of particular words within *Minecraft* may also have strengthened their salience. Understanding *diamond*, *sword*, *zombie*, *craft*, or *build* helps viewers follow objectives, dangers, and actions central to the game. Their meanings are embedded in relationships among language, objects, actions, and goals. This situated connection distinguishes game-related vocabulary from words presented as isolated items and helps explain why domain-specific words sometimes became more memorable despite occurring less frequently (Zheng et al., 2015). The children's explanations also showed that recognition could not be attributed exclusively to the selected videos. One participant associated *go* with a command used by a taekwondo coach, while the other connected *let's go* with an expression frequently used by their mother. These responses suggest that the children combined encounters from videos with language experienced in family communication,

extracurricular activities, games, and other media. Encountering a word across diverse meaningful contexts can strengthen its accessibility and enrich its semantic representation (Frances et al., 2020). Nevertheless, because the participants' different sources of exposure could not be separated, the results demonstrate recognition following the viewing period rather than acquisition caused exclusively by YouTube gameplay videos.

4.2 Multimodal Context and Gameplay Relevance

The participants usually explained recognized vocabulary through visible actions, objects, characters, or memorable gameplay events. They connected *build* with constructing a house, *sword* and *zombie* with fighting, and *diamond* with locating a valuable object. Some vocabulary was recalled through action-based expressions, such as *build a house*, rather than through isolated definitions. The children's interpretations therefore appeared to draw simultaneously on spoken commentary, subtitles, visual imagery, movement, and knowledge of the game. The Cognitive Theory of Multimedia Learning helps explain how these different modes may have supported interpretation. Multimedia learning involves selecting and organizing verbal and pictorial information and integrating it with relevant prior knowledge (Mayer, 2017, 2024). In subtitled gameplay videos, spoken commentary provides auditory-verbal information, subtitles display written forms, and the gameplay presents corresponding objects and actions. When these elements are closely aligned, they offer complementary cues through which learners can interpret recurring vocabulary.

Research on captioned audiovisual materials similarly indicates that captions can make lexical forms more accessible and help learners connect spoken words with their orthographic representations. Visual imagery can further support form-meaning connections when it appears near the spoken occurrence and clearly depicts the relevant action or object (Montero Perez et al., 2015, 2018; Peters, 2019). The participants' clearer explanations of visually identifiable words are compatible with this account. Nevertheless, the present study did not independently manipulate subtitles, imagery, or spoken commentary. It cannot establish which mode contributed most strongly to recognition. Captions should be understood as resources that learners may use while processing audiovisual content, not as features that automatically produce vocabulary learning (Vanderplank, 2016). The most defensible interpretation is therefore that the coordination of written, spoken, and visual information created multiple opportunities for contextual interpretation.

The nature of the videos also shaped these opportunities. The children were not watching instructional materials designed to teach English. *Let's Play* videos combine gameplay footage, spontaneous commentary, narration, entertainment, and digital performance (Burwell & Miller, 2016; Shoda, 2022). English was consequently encountered while the children followed actions, objectives, problems, and humorous events. Words central to understanding these events may have attracted greater attention than words that occurred frequently but contributed less visibly to the gameplay. This distinction is important when relating the findings to digital game-based learning. Active gameplay can support vocabulary engagement through decision-making, problem solving, feedback, and goal-directed action (Acquah & Katz, 2020; Weisi & Hajizadeh, 2025). Watching recorded gameplay does not provide the same degree of interactivity because viewers do not control the avatar or determine what occurs. Nevertheless, gameplay videos retain contextual features of games, including recognizable objects, visible actions, objectives, and domain-specific language. The present study therefore illustrates the potential of game-related viewing without treating it as equivalent to active game participation. One participant reported that pausing had become less necessary because many recurring words were already familiar. This account suggests increased subjective familiarity with language commonly used in the videos. However, it does not independently demonstrate improvement in general listening comprehension because the study did not administer a standardized listening assessment. The statement is better interpreted as indicating that recurring language had become more manageable within the familiar context of *Minecraft* viewing.

4.3 Informal Viewing, Interest, and Learner Agency

The children's engagement with the videos was voluntary and driven primarily by their interest in *Minecraft*. They selected English-language content, navigated video menus, replayed tutorials, revisited interesting scenes, and requested assistance when they did not understand particular events. English was therefore encountered as part of an enjoyable activity rather than as the explicit object of a formal lesson. This pattern reflects broader research on informal exposure showing that young learners encounter English through gaming, online videos, streaming platforms, and other self-selected activities outside school. Such exposure may be particularly

meaningful when it is sustained by personal interest and connected with recognizable purposes or activities (De Wilde et al., 2020; Dizon, 2024; Peters et al., 2019). In the present case, the children watched to understand and enjoy *Minecraft*, while vocabulary exposure emerged as part of pursuing that interest. Their selection and replaying of videos also demonstrated learner agency. The participants decided what to watch, returned to useful or entertaining sequences, and sought explanations when necessary. Informal digital learning allows learners to exercise control over content, timing, repetition, and participation, although such engagement remains influenced by the social and material environment in which it occurs (Lee, 2021; Toh & Lim, 2022). The participants were therefore self-directed in their viewing choices without being entirely independent of assistance from others.

Enjoyment appeared to help sustain this engagement. The participants referred to funny scenes, fighting zombies, finding diamonds, constructing houses, and watching with siblings or friends. Previous game-based learning research has linked enjoyment and motivation with continued participation in vocabulary-rich digital activities (Ebrahimzadeh & Alavi, 2016, 2017). Studies of children's engagement with online video content have likewise suggested that interest and entertainment can create opportunities for repeated English exposure. However, motivation was not formally measured in this study. The findings indicate expressed enjoyment and voluntary participation rather than a measurable improvement in language-learning motivation. The children's ability to locate content, navigate menus, and replay selected segments also documented the digital practices through which they managed their exposure. Gameplay videos and gaming communities involve navigation, multimodal interpretation, and coordination of textual, auditory, and visual information (Dezuanni et al., 2015; Wernholm, 2021). These practices should not, however, be interpreted as digital-literacy gains caused by the study because the participants' abilities before the observation period were not assessed.

4.4 Sibling, Peer, and Parental Mediation

The children's viewing did not occur entirely in isolation. Both participants described watching with siblings or friends, requesting assistance, receiving explanations, and discussing memorable gameplay events. One child reported that a brother explained what was happening, while the other emphasized sharing funny scenes with a friend. Some English expressions, including *build a house*, *jump*, and *run*, were subsequently used during play. Vocabulary encountered in the videos therefore circulated through co-viewing, explanation, conversation, and game-related reenactment. From a sociocultural perspective, language engagement is mediated by interactions with other people and by the symbolic and technological resources available within an activity (Lantolf & Xi, 2023). In this case, spoken commentary, subtitles, gameplay imagery, replay functions, siblings, and friends provided interconnected resources. When the children did not understand an event, they could observe the corresponding action, replay the scene, or request an explanation. Meaning was therefore constructed within a broader social and digital activity rather than derived from subtitles alone.

Sibling involvement was particularly important because it provided immediate and accessible assistance. Older siblings can mediate younger children's digital-media use by explaining content, supporting navigation, and participating in shared viewing or play (Adler & Elias, 2025). Collaborative engagement with *Minecraft* can similarly create situated opportunities for discussion, shared problem solving, and language use (Ståhl et al., 2025). The assistance reported here was informal, but it may have helped the participants connect unfamiliar expressions with recognizable gameplay events. Friends contributed differently by providing companionship, shared enjoyment, and opportunities to discuss or reenact the content. Shared exploration of online videos can encourage children to talk about what they watch and reuse vocabulary related to their common interests. Such interaction does not establish stable productive mastery, but it shows that some expressions gained relevance beyond the immediate viewing event. Parental involvement was more indirect. Parents primarily regulated screen access, supervised the content, occasionally watched with the children, and noticed their use of English during play. Research on children's digital-media use distinguishes among restrictive mediation, supervision, active discussion, and adult-child co-use (Neumann & Herodotou, 2020). The parental practices observed here were closer to supervision and occasional co-use than to systematic language instruction.

Active adult-child interaction can support learning when adults explain unfamiliar material, direct attention toward relevant information, or connect screen representations with children's experiences (Neuman et al., 2020; Samudra et al., 2019; Strouse et al., 2018). The present evidence, however, does not show that parents regularly taught vocabulary or deliberately directed attention toward subtitles. Their contribution was primarily

environmental: establishing access conditions, monitoring viewing, responding when necessary, and acknowledging the children's subsequent English use. Evidence across digital-learning contexts suggests that adult-child co-use has a modest positive relationship with learning, but its influence varies according to the quality, purpose, and frequency of interaction (Taylor et al., 2024). Supportive family environments may facilitate informal engagement without transforming entertainment into formal instruction. In the present case, siblings and friends appeared to provide more immediate explanations and shared discussion, while parents maintained the broader conditions within which viewing occurred.

4.5 Integrated Interpretation and Pedagogical Implications

The findings indicate that no single factor adequately explains the participants' vocabulary recognition. Frequency increased the possibility of encountering a word, but functional and visual relevance made some words more interpretable. Subtitles displayed written forms, spoken commentary supplied auditory forms, and gameplay footage presented associated objects and actions. Previous experiences allowed the children to connect video language with familiar expressions, while interaction with siblings, friends, and parents provided opportunities for explanation, discussion, and reuse. The principal interpretive contribution of the study is therefore the distinction among lexical availability, contextual interpretability, and elicited recognition. Corpus frequency represented the vocabulary available within the sampled subtitle input. Multimodal and gameplay contexts made some of that vocabulary more readily interpretable. The interviews then identified words the participants could recognize and explain at a particular point in time. These categories should not be conflated: availability does not guarantee attention, and elicited recognition does not establish the original source, depth, or durability of vocabulary knowledge.

Pedagogically, gameplay videos may function as supplementary resources when selected vocabulary corresponds clearly with visible objects, actions, or objectives. Teachers and parents could ask children to identify game objects, explain character actions, connect spoken expressions with subtitles, retell gameplay events, or use selected expressions in off-screen activities. These practices could direct attention toward language while preserving the interest that motivates children to watch the content. Gameplay videos should not, however, be regarded as inherently educational simply because they contain English. Their potential value depends on content suitability, comprehensibility, alignment between language and imagery, familiarity with the game, opportunities for interaction, and appropriate management of screen time. Adult guidance may help children engage more purposefully with the content without turning an informal recreational activity into a compulsory language lesson.

These interpretations must remain within the evidential limits of the study. The investigation involved only two participants from one family over a limited period and included no pretest, comparison group, experimental manipulation, or delayed retention assessment. The children's previous exposure through school, family communication, extracurricular activities, games, and other media could not be controlled. Corpus frequency also measured occurrences in the subtitle files rather than verified attention to individual words. Consequently, interview recognition demonstrated accessible lexical knowledge during elicitation but did not establish when, where, or how securely each word had been learned. Future research could combine naturalistic observation with pretests, immediate post-tests, and delayed assessments. Comparisons between captioned and uncaptioned videos, active gameplay and gameplay viewing, or individual and socially mediated viewing could clarify the respective contributions of repetition, imagery, captions, interactivity, and family support. Systematic records of pausing, replaying, viewing duration, and attention to subtitles would also help distinguish the vocabulary available in the videos from the input actually noticed by learners.

5. Conclusion

This study explored how subtitled Minecraft gameplay videos created opportunities for English vocabulary engagement among two young EFL learners in a family environment. The findings show that lexical frequency alone did not determine which words the children recognized. Although the videos repeatedly presented common verbs, several less frequent game-related words were more readily explained because they referred to visible objects, actions, or goals central to the gameplay. Vocabulary recognition therefore appeared to depend on the interaction of repeated exposure, contextual relevance, multimodal representation, and prior familiarity. The study further demonstrates that gameplay-video viewing was both self-directed and socially situated. The children selected and replayed content according to their interests, used available contextual cues, and sought

assistance when necessary. Spoken commentary, subtitles, and gameplay imagery provided complementary resources for interpreting vocabulary, while siblings and friends offered immediate explanations and opportunities for discussion or reuse. Parents contributed more indirectly by supervising access, monitoring content, and maintaining a supportive viewing environment.

Pedagogically, subtitled gameplay videos may serve as supplementary resources when their language is age-appropriate and clearly connected with visible actions, objects, and objectives. Teachers and parents can strengthen their value by encouraging children to identify objects, explain actions, connect spoken and written forms, retell gameplay events, or reuse selected expressions in other activities. Nevertheless, their educational value depends on appropriate content selection, comprehensibility, purposeful mediation, and responsible management of screen time. These conclusions remain limited by the involvement of only two participants from one family, the short observation period, and the absence of pretests, a comparison group, and delayed retention measures. The influence of prior exposure through school, family communication, extracurricular activities, games, and other media could not be isolated. Future studies should involve more diverse participants, assess vocabulary knowledge over time, and compare captioned and uncaptioned viewing, active gameplay and gameplay-video viewing, and individual and socially mediated viewing. Such research would clarify how available lexical input becomes noticed, interpreted, retained, and eventually used by young EFL learners.

References

- Acquah, E. O., & Katz, H. T. (2020). Digital game-based L2 learning outcomes for primary through high-school students: A systematic literature review. *Computers & Education*, 143, Article 103667. <https://doi.org/10.1016/j.compedu.2019.103667>
- Adler, O., & Elias, N. (2025). "My brother teaches me everything": Sibling mediation of young Israeli children's media use. *Journal of Children and Media*, 19(2), 327–343. <https://doi.org/10.1080/17482798.2024.2414443>
- Anthony, L. (2024). *AntConc* (Version 4.3.1) [Computer software]. Waseda University. <https://www.laurenceanthony.net/software/antconc/>
- Burwell, C., & Miller, T. (2016). Let's Play: Exploring literacy practices in an emerging videogame paratext. *E-Learning and Digital Media*, 13(3–4), 109–125. <https://doi.org/10.1177/2042753016677858>
- Castillo-Montoya, M. (2016). Preparing for interview research: The interview protocol refinement framework. *The Qualitative Report*, 21(5), 811–831. <https://doi.org/10.46743/2160-3715/2016.2337>
- Chien, Y.-C. (2019). The language of massively multiplayer online gamers: A study of vocabulary in Minecraft gameplay. *TESL-EJ*, 23(3). <https://tesl-ej.org/wordpress/issues/volume23/ej91/ej91int/>
- De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure: Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, 23(1), 171–185. <https://doi.org/10.1017/S1366728918001062>
- Dezuanni, M., O'Mara, J., & Beavis, C. (2015). 'Redstone is like electricity': Children's performative representations in and around Minecraft. *E-Learning and Digital Media*, 12(2), 147–163. <https://doi.org/10.1177/2042753014568176>
- Dizon, G. (2024). "Watching video is maybe just like a hobby": A case study of video streaming for informal, self-directed foreign language learning. *The EUROCALL Review*, 31(2), 34–44. <https://doi.org/10.4995/eurocall.2024.20738>
- Ebrahimzadeh, M. (2017). Readers, players, and watchers: EFL students' vocabulary acquisition through digital video games. *English Language Teaching*, 10(2), 1–18. <https://doi.org/10.5539/elt.v10n2p1>
- Ebrahimzadeh, M., & Alavi, S. (2016). Motivating EFL students: E-learning enjoyment as a predictor of vocabulary learning through digital video games. *Cogent Education*, 3(1), Article 1255400. <https://doi.org/10.1080/2331186X.2016.1255400>

- Ebrahimzadeh, M., & Alavi, S. (2017). The effect of digital video games on EFL students' language-learning motivation. *Teaching English with Technology*, 17(2), 87–112. <https://files.eric.ed.gov/fulltext/EJ1140688.pdf>
- Fievez, I., Perez, M. M., Cornillie, F., & Desmet, P. (2020). Vocabulary learning through viewing captioned or subtitled videos and the role of Learner- and Word-Related factors. *CALICO Journal*, 37(3), 233–253. <https://doi.org/10.1558/cj.39370>
- Frances, C., Martin, C. D., & Duñabeitia, J. A. (2020). The effects of contextual diversity on incidental vocabulary learning in the native and a foreign language. *Scientific Reports*, 10, Article 13967. <https://doi.org/10.1038/s41598-020-70922-1>
- Fusch, P., Fusch, G. E., & Ness, L. R. (2018). Denzin's paradigm shift: Revisiting triangulation in qualitative research. *Journal of Social Change*, 10(1), 19–32. <https://doi.org/10.5590/JOSC.2018.10.1.02>
- Gekker, A. (2018). Let's not play: Interpassivity as resistance in 'Let's Play' videos. *Journal of Gaming & Virtual Worlds*, 10(3), 219–242. https://doi.org/10.1386/jgvw.10.3.219_1
- Gentles, S. J., Charles, C., Ploeg, J., & McKibbin, K. A. (2015). Sampling in qualitative research: Insights from an overview of the methods literature. *The Qualitative Report*, 20(11), 1772–1789. <https://doi.org/10.46743/2160-3715/2015.2373>
- Hanghøj, T., Kabel, K., & Hannibal Jensen, S. (2022). Digital games, literacy and language learning in L1 and L2: A comparative review. *L1-Educational Studies in Language and Literature*, 22(2), 1–44. <https://doi.org/10.21248/l1esll.2022.22.2.363>
- Harrison, H., Birks, M., Franklin, R., & Mills, J. (2017). Case study research: Foundations and methodological orientations. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 18(1), Article 19. <https://doi.org/10.17169/fqs-18.1.2655>
- Hung, H.-T., Yang, J.-C., Hwang, G.-J., Chu, H.-C., & Wang, C.-C. (2018). A scoping review of research on digital game-based language learning. *Computers & Education*, 126, 89–104. <https://doi.org/10.1016/j.compedu.2018.07.001>
- Janebi Enayat, M., & Derakhshan, A. (2021). Vocabulary size and depth as predictors of second language speaking ability. *System*, 99, Article 102521. <https://doi.org/10.1016/j.system.2021.102521>
- Jensen, S. H. (2017). Gaming as an English language learning resource among young children in Denmark. *CALICO Journal*, 34(1), 1–19. <https://doi.org/10.1558/cj.29519>
- Kutrovátz, K. (2017). Conducting qualitative interviews with children: Methodological and ethical challenges. *Corvinus Journal of Sociology and Social Policy*, 8(2), 65–88. <https://doi.org/10.14267/CJSSP.2017.2.04>
- Kyle, K., & Crossley, S. A. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, 34, 12–24. <https://doi.org/10.1016/j.jslw.2016.10.003>
- Lantolf, J. P., & Poehner, M. E. (2023). Sociocultural theory and classroom second language learning in the East Asian context: Introduction to the special issue. *The Modern Language Journal*, 107(S1), 3–23. <https://doi.org/10.1111/modl.12816>
- Lantolf, J. P., & Xi, J. (2023). Digital language learning: A sociocultural theory perspective. *TESOL Quarterly*, 57(2), 702–715. <https://doi.org/10.1002/tesq.3218>
- Lantolf, J. P., Xi, J., & Minakova, V. (2021). Sociocultural theory and concept-based language instruction. *Language Teaching*, 54(3), 327–342. <https://doi.org/10.1017/S0261444820000348>
- Lee, J. S. (2021). *Informal digital learning of English: Research to practice*. Routledge. <https://doi.org/10.4324/9781003043454>
- Lee, J. S., & Dressman, M. (2018). When IDLE hands make an English workshop: Informal digital learning of English and language proficiency. *TESOL Quarterly*, 52(2), 435–445. <https://doi.org/10.1002/tesq.422>
- Leona, N. L., van Koert, M. J. H., van der Molen, M. W., Rispen, J. E., Tijms, J., & Snellings, P. (2021). Explaining individual differences in young English language learners' vocabulary knowledge: The role of extramural

- English exposure and motivation. *System*, 96, Article 102402. <https://doi.org/10.1016/j.system.2020.102402>
- Liu, T.-C., Lin, Y.-C., & Paas, F. (2022). A new application of the temporal contiguity effect in designing narrated slideshows. *Educational Technology Research and Development*, 70(1), 59–72. <https://doi.org/10.1007/s11423-021-10076-7>
- Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, 33(5), 403–423. <https://doi.org/10.1111/jcal.12197>
- Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review*, 36, Article 8. <https://doi.org/10.1007/s10648-023-09842-1>
- Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal*, 99(2), 308–328. <https://doi.org/10.1111/modl.12215>
- Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning*, 31(1–2), 1–26. <https://doi.org/10.1080/09588221.2017.1375960>
- Muñoz, C., & Cadierno, T. (2021). How do differences in exposure affect English language learning? A comparison of teenagers in two learning environments. *Studies in Second Language Learning and Teaching*, 11(2), 185–212. <https://doi.org/10.14746/ssllt.2021.11.2.2>
- Neuman, S. B., Samudra, P. G., Wong, K. M., & Kaefer, T. (2020). Scaffolding attention and partial word learning through interactive co-viewing of educational media: An eye-tracking study with low-income preschoolers. *Journal of Educational Psychology*, 112(6), 1100–1110. <https://doi.org/10.1037/edu0000428>
- 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>
- Nguyen, J. (2016). Performing as video game players in Let's Plays. *Transformative Works and Cultures*, 22. <https://doi.org/10.3983/twc.2016.0698>
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1), 1–13. <https://doi.org/10.1177/1609406917733847>
- Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, 53(4), 1008–1032. <https://doi.org/10.1002/tesq.531>
- Peters, E., Noreillie, A.-S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages. *Language Learning*, 69(3), 747–782. <https://doi.org/10.1111/lang.12351>
- Piittinen, S. (2018). Morality in Let's Play narrations: Moral evaluations of Gothic monsters in gameplay videos of *Fallout 3*. *New Media & Society*, 20(12), 4671–4688. <https://doi.org/10.1177/1461444818779754>
- Pires, F., Masanet, M.-J., & Scolari, C. A. (2021). What are teens doing with YouTube? Practices, uses and metaphors of the most popular audio-visual platform. *Information, Communication & Society*, 24(9), 1175–1191. <https://doi.org/10.1080/1369118X.2019.1672766>
- Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, 47(4), 479–496. <https://doi.org/10.1080/09571736.2019.1616806>
- Rashid, Y., Rashid, A., Warraich, M. A., Sabir, S. S., & Waseem, A. (2019). Case study method: A step-by-step guide for business researchers. *International Journal of Qualitative Methods*, 18, 1–13. <https://doi.org/10.1177/1609406919862424>

- Samudra, P. G., Flynn, R. M., & Wong, K. M. (2019). Coviewing educational media: Does coviewing help low-income preschoolers learn auditory and audiovisual vocabulary associations? *AERA Open*, 5(2), 1–12. <https://doi.org/10.1177/2332858419853238>
- Sandberg, H., & Gillen, J. (2021). Investigating the digital media engagements of very young children at home: Reflecting on methodology and ethics. *Communications*, 46(3), 332–351. <https://doi.org/10.1515/commun-2021-0046>
- Scholz, K. (2017). Encouraging free play: Extramural digital game-based language learning as a complex adaptive system. *CALICO Journal*, 34(1), 39–57. <https://doi.org/10.1558/cj.29527>
- Schroeder, N. L., & Cenkci, A. T. (2018). Spatial contiguity and spatial split-attention effects in multimedia learning environments: A meta-analysis. *Educational Psychology Review*, 30(3), 679–701. <https://doi.org/10.1007/s10648-018-9435-9>
- Shoda, V. P. (2022). Let's play videos in literacy practice: From let's play to let's learn. *E-Learning and Digital Media*, 19(5), 515–536. <https://doi.org/10.1177/20427530221108015>
- Ståhl, M., Hansell, K., Bäck, S., & Wingren, M. (2025). Affordances for in-game interaction and language learning through children's collaborative play in Minecraft. *International Journal of Game-Based Learning*, 15(1), 1–23. <https://doi.org/10.4018/IJGBL.370559>
- Strouse, G. A., Troseth, G. L., O'Doherty, K. D., & Saylor, M. M. (2018). Co-viewing supports toddlers' word learning from contingent and noncontingent video. *Journal of Experimental Child Psychology*, 166, 310–326. <https://doi.org/10.1016/j.jecp.2017.09.005>
- Stufft, C. J., & von Gillern, S. (2021). Fostering multimodal analyses of video games: Reflective writing in the middle school. *Journal of Adolescent & Adult Literacy*, 65(3), 245–255. <https://doi.org/10.1002/jaal.1198>
- Sundqvist, P. (2019). Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary. *Language Learning & Technology*, 23(1), 87–113. <https://doi.org/10.64152/10125/44674>
- Sundqvist, P., & Wikström, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System*, 51, 65–76. <https://doi.org/10.1016/j.system.2015.04.001>
- Taylor, G., Sala, G., Kolak, J., Gerhardstein, P., & Lingwood, J. (2024). Does adult-child co-use during digital media use improve children's learning aged 0–6 years? A systematic review with meta-analysis. *Educational Research Review*, 44, Article 100614. <https://doi.org/10.1016/j.edurev.2024.100614>
- Teng, F. (2019). Incidental vocabulary learning for primary school students: The effects of L2 caption type and word exposure frequency. *The Australian Educational Researcher*, 46(1), 113–136. <https://doi.org/10.1007/s13384-018-0279-6>
- Toh, W., & Lim, F. V. (2022). Learning in digital play: A dual case study of video gamers' independent play. *Research and Practice in Technology Enhanced Learning*, 17, Article 6. <https://doi.org/10.1186/s41039-022-00182-2>
- Touri, M., & Koteyko, N. (2015). Using corpus linguistic software in the extraction of news frames: Towards a dynamic process of frame analysis in journalistic texts. *International Journal of Social Research Methodology*, 18(6), 601–616. <https://doi.org/10.1080/13645579.2014.929878>
- Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, 69(3), 559–599. <https://doi.org/10.1111/lang.12343>
- Vanderplank, R. (2016). 'Effects of' and 'effects with' captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Language Teaching*, 49(2), 235–250. <https://doi.org/10.1017/S0261444813000207>
- Weisi, H., & Hajizadeh, S. (2025). Mining for words: The effect of Minecraft on incidental vocabulary learning of young EFL learners. *Entertainment Computing*, 52, Article 100801. <https://doi.org/10.1016/j.entcom.2024.100801>

- Wernholm, M. (2021). Children's out-of-school learning in digital gaming communities. *Designs for Learning*, 13(1), 8–19. <https://doi.org/10.16993/dfl.164>
- Xi, J., & Lantolf, J. P. (2021). Scaffolding and the zone of proximal development: A problematic relationship. *Journal for the Theory of Social Behaviour*, 51(1), 25–48. <https://doi.org/10.1111/jtsb.12260>
- Xu, Z., Chen, Z., Eutsler, L., Geng, Z., & Kogut, A. (2020). A scoping review of digital game-based technology on English language learning. *Educational Technology Research and Development*, 68, 877–904. <https://doi.org/10.1007/s11423-019-09702-2>
- Yazan, B. (2015). Three approaches to case study methods in education: Yin, Merriam, and Stake. *The Qualitative Report*, 20(2), 134–152. <https://doi.org/10.46743/2160-3715/2015.2102>
- Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, 26(4), 696–725. <https://doi.org/10.1177/1362168820913998>
- Zheng, D., Bischoff, M., & Gilliland, B. (2015). Vocabulary learning in massively multiplayer online games: Context and action before words. *Educational Technology Research and Development*, 63(5), 771–790. <https://doi.org/10.1007/s11423-015-9387-4>