The Multimodal Affordances of Commercial-off-the-shelf Video Games That Support Vocabulary Learning

by

Jean Charlebois

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Arts

in

School of Linguistics and Language Studies

Carleton University
Ottawa, Ontario

Abstract

Engagement with video games by second-language (L2) users has been shown to be a substantial source of English-language exposure linked to increased L2 vocabulary proficiency (Lindgren & Munoz, 2013). Word learning through gaming has been attributed to the affordances of video games (Kuppins, 2010), although few studies have investigated how digital games presents vocabulary learning opportunities. This research expands on the work carried out by Rodgers (2018), which applied the principles of multimedia learning (Mayer, 2021) to television, by applying a multimodal analysis to 10 commercial games to examine how spoken dialogue is supported through visual, textual, and procedural modes. This was done by conducting a corpus analysis of gameplay recordings to classify the multimodal affordances for vocabulary. Results indicate that the imagery in digital games occurs similarly to narrative television and is further benefitted by additional modalities, demonstrating that video gaming could be a suitable medium for extramural learning

Keywords: Digital Game-based Language Learning, Multimedia Learning, Vocabulary Acquisition

Acknowledgements

To my students who inspired this thesis - keep playing, keep learning.

I want to thank the School of Linguistics and Language Studies department; this MA program has nurtured a keen interest in language-related research.

Thank you to my cohort of graduate students who, while remotely, have enriched course content through sharing their perspectives and engaging in insightful discussions. My virtual classroom community has been integral in making quarantine a less isolated affair.

I am grateful to my friends and family, especially my partner Kayleigh for providing me with almost a decade of support in my professional and academic endeavours. None of which would have been possible without her.

Lastly, an enormous thank you to Dr. Michael Rodgers, who has been paramount in instilling an interest in me for this line of research. I owe much of my professional career as a language teacher and my burgeoning position in academia to his supervision and counsel since 2016.

Table of Contents

Abstract ..... ***ii***

Acknowledgements ..... ***iii***

Table of Contents ..... ***iv***

List of Tables ..... ***vii***

List of Appendices ..... ***ix***

Chapter 1: Introduction ***1***

Chapter 2: Literature Review ***4***

2.1 Theoretical Framework **6**

Incidental Language Learning 6

Cognitive Theory of Multimedia Learning 9

2.2 Language Learning Through Meaning-focused Input **14**

Extensive Input 14

2.3 Factors That Can Affect Vocabulary Acquisition **18**

Frequency of Occurrence 18

2.4 Digital Game-based Language Learning **19**

Digital Games as Multimodal Texts 23

2.5 The Present Study **25**

Chapter 3: Methods ***27***

3.1 Video Games **27**

Selection Criteria 27

Walkthroughs 28

Genre 29

3.2 Target Word Selection **30**

Lexical Profile of Video Game Texts 31

Method 33

3.3 Multimodal Supports for Target Words **37**

Visual Mode 38

Written Mode 40

Procedural Mode 42

Chapter 4: Results ***46***

4.1 First Research Question **46**

Visual Mode 46

Written Mode 50

4.2 Second Research Question **51**

4.3 Third Research Question **55**

Comparison Between Genres 55

Comparison Between Media 59

Chapter 5: Discussion ***62***

5.1 Findings 62

What is the occurrence of visual and textual supports for the aural input of target words in commercial-off-the-shelf video games within a 10-second period?..... 62

What are the opportunities for target words to be interacted with in a virtual environment presented in commercial-off-the-shelf video games? 64

How does concurrent or near-concurrent visual and aural input compare between two video game genres and television programs? 67

5.2 Limitations and Future Research into vocabulary learning through video games 69

Visual Mode 72

Textual Mode 73

Procedural Mode 74

5.2 Pedagogical Considerations 77

Before gameplay 78

During Gameplay 80

After Gameplay 82

Chapter 6: Conclusion 83

References..... 85

Appendices..... *Error! Bookmark not defined.*

List of Tables

Table 3.1. Video game texts…….……..……34

Table 3.2. Number of target items selected based on their frequency of occurrence in each game…….……37

Table 4.1. Occurrence of visual support in relation to spoken target word provision……..…..47

Table 4.2. The degree of imagery and spoken word occurrence……..48

Table 4.3. Visually supported target words……49

Table 4.4. Occurrence of textual support in relation to spoken target word provision……..50

Table 4.5. Target word encounters through the procedural mode……..52

Table 4.6. Encounters of manipulation of target word……..54

Table 4.7. Skewness and kurtosis for aural-visual occurrence between genres…….……55

Table 4.8. Descriptive statistics for audio-visual concurrence in two video game genres……..57

Table 4.9 Mann-Whitney U test for audio-visual concurrence in video game genres……57

Table 4.10 Descriptive Statistics for audio-textual concurrence in video game genres……57

Table 4.11 Mann-Whitney U test for audio-textual concurrence in video game genres……..58

Table 4.12 Descriptive Statistics for procedural manipulations with target words……..58

Table 4.13 Mann-Whitney U test for procedural manipulations with target words……59

Table 4.14 Descriptive statistics supportive imagery between video games and television……..59

Table 4.15 Mann-Whitney U test for supportive imagery between video games and television. ……60

Table 4.16 Descriptive statistics supportive imagery between narrative fiction and video games.. ……60

Table 4.17 Mann-Whitney U test for supportive imagery between video games and television ……61

List of Appendices

Appendix A. Pre-gaming teacher intervention example……97

Appendix B. Post-gaming teacher intervention example……..98

Chapter 1: Introduction

Exposure to authentic, but comprehensible, language input is a key component for successful language learning (Nunan, 2002). However, for many language learners, particularly in an English as a foreign Language (EFL) context, exposure to the substantial quantities of authentic language may be difficult to achieve. One pedagogical intervention to addressed this, is by establishing an extensive reading programme in the language classroom (Nation, 2015). These programmes allow learners to choose appropriate material to extend their exposure to the target language beyond the classroom (Nation, 2015). This notion of extramural reading has been expanded to the extramural viewing of television (Rodgers, 2016; Webb, 2015). This consumption of audio-visual input has benefitted second-language (L2) learning in such way as: "improved listening skills, increased listening fluency, acquisition of new vocabulary, deepened knowledge of previously known vocabulary, and increased motivation between towards language learning" (Rodgers, 2018, pp. 191-192). These benefits of extramural input of English language could be further extended to other sources of audio-visual media, such as video games.

Digital gaming has seen consistent growth, with 2021 seeing a global revenue of $189.3 billion USD (Wijman, 2021). Despite video game developers releasing games in a myriad of languages, many gamers choose to play video games in their target language as rich source of authentic language input (Reinhardt & Han, 2021). English-mediated digital gaming is one of the few accessible sources of contact with their target language outside of the classroom (Sundqvist, 2021). In fact, engagement with video games by second-language (L2) users has been shown to be a valuable source of input. For example, in the Swedish (Sundqvist & Sylven,

2014), Belgian (Peters, 2018), and broader European (Lindgren & Munoz, 2013) contexts, gaming has been shown to be a substantial source of English-language exposure which has been linked to increased L2 proficiency (Lindgren & Munoz, 2013).

Meta-analyses into the field of digital game-based language learning (DGBLL) have identified the positive effects that digital gaming practices have had on language learning, particularly in vocabulary (Dixon et al., 2022; Tsai & Tsai, 2018; Chen et al., 2018). Second-language acquisition (SLA) research is invested in capitalizing on the potential affective motivational boon that video games could bring to the second-language classroom (Reinhardt, 2019). However, research into video games is still in its early stages (York et al., 2020). While game-based intervention studies have indicated positive vocabulary growth (Kuppins, 2010; Peterson, 2011), the reasoning behind this learning has been speculative (York et al., 2020). With this in mind, the present study seeks to provide an empirical basis for further investigations into vocabulary learning through playing video games.

Under the cognitive theory of multimedia learning, Mayer has established a wealth of empirical evidence for the psychological underpinnings for learning through various multimedia sources (Mayer, 2021). Mayer established educational design principles for multimedia messages (Mayer, 2021). One of these principles, the temporal contiguity principle, suggests that learning is benefitted from simultaneous presentation of corresponding words and imagery (Mayer, 2021). Rodgers (2018) applied this principle to investigate the supportive visuals within television programmes and found that this medium presented audio-visual congruity for vocabulary words and their visual referents. The congruity of associated words and images has led to the potential learning of previously unknown words (Rodgers, 2018;

Ghebghoub, 2021). This finding has established the potential for extramural viewing practices that could extend learning beyond the language classroom. The present study expands on the work carried out by Rodgers (2018) by applying Mayer's (2021) temporal contiguity principle to the medium of video games. The findings of which could provide empirical evidence to account for the vocabulary learning which has been observed in digital game-based language learning research.

Chapter 2: Literature Review

Incidental language learning, or learning through linguistic exposure, requires vast amounts of meaning-focused input (Nation, 2007). Due to limited classroom time, it is unlikely that language learners (LLs) will be exposed to enough input (Nation, 2007). Thus, it would behoove vocabulary research to explore all media for extramural linguistic exposure (Schmitt & Schmitt, 2020). Different media can possess affordances that could foster lexical development (Webb & Nation, 2017; Schmitt, 2020). This research will explore the potential for video games as a source of multimodal linguistic input.

Narrative video games host lexical demands that are comparable to watching a television program (Rodgers & Heidt, 2020). However, video games possess multimodal affordances that could differentiate them from other sources of language input (Reinhardt, 2019; Rodgers & Heidt, 2020). The auditory, visual, text-based, and interactive affordances presented through a game's narrative could be a potential source of learning unknown vocabulary items. Similar research has shown potential in learning vocabulary supported through concurrent visual and aural input in television programs (Rodgers, 2018). This effect aligns with the multimedia principle (i.e., temporal contiguity principle; Mayer et al., 1999; Ginns, 2006), which states that simultaneous spoken text and on-screen images are beneficial for learning (Mayer et al.,1999; Ginns, 2006). Within the context of video games, the temporal contiguity principle supports the claim that the co-occurring spoken dialog and on-screen images within video games can foster similar vocabulary learning to television (Rodgers & Heidt, 2020). Moreover, video games possess a type of interaction that is not possible within other input media (Calleja, 2011). Studies have started to observe the language learning effects

of game-enhanced from commercial off-the-shelf (COTS) video games (Reinhardt, 2019), which have yielded favourable results on vocabulary learning (Dixon et al., 2022; Tsai & Tsai, 2018; Chen et al., 2018). While gaming has been used supplementally in the language classroom to strengthen vocabulary knowledge (Franciosi, 2016; Bregni, 2017; McFayden, 2020), this research will explore incidental learning vocabulary through meaning-focused input.

Learning context can be a mediating factor in second language acquisition (Munoz, 2014). Munoz (2014) noted that "young learners' slow pace of learning […] may be related to the scarcity of input in typical [foreign language] learning settings" (p. 475). Thus, it is perhaps unlikely that learners studying in a foreign language classroom can achieve the input required for linguistic development (Nation, 2013; Munoz 2014). To supplement linguistic input, learners may need to engage in extramural activities to provide additional sources of language input (Nation, 2013). To this end, extensive reading has been a recommended practice to provide learners with this quantity of input (Nation, 2013). Extensive reading is a language learning practice in which learners read vast amounts of texts, either for enjoyment or to learn content in a second language (Day & Bamford, 1998). In an English as a foreign language context, linguistic exposure of the target language may be sparse (Munoz, 2014). Extensive reading programs, instructor-facilitated reading outside of class time, have proved to be a valuable source of suitable linguistic exposure both in contexts where learners are unable to engage in day-to-day interactions in a target language and to bolster the language exposure in contexts where the target language is frequently used (Nation, 2013). This learning could extend to other activities which could be a source of language input (Webb & Nation, 2017).

Research has explored the potential vocabulary gains through extensive input of various media such as, reading stories (Waring & Takaki, 2003), listening to stories (Elley, 1989), watching television (Webb & Rodgers, 2009b), and watching movies (Webb & Rodgers, 2009a). Despite the widespread interest in playing video games, however, research has been slow to investigate the possibility of videogames as a source of linguistic input (Reinhardt, 2019). The lexical exposure of video games could be comparable to other audio-visual media, such as television (Rodgers & Heidt, 2020). However, there may features intrinsic within video games that could provide additional support over other audio-visual mediums, such as supportive imagery, text, and gameplay procedures (deHaan et al., 2010).

2.1 Theoretical Framework

Incidental Language Learning

Understanding the size and growth of second language acquisition may be essential to observe how a native speaker's first language substantially increases (Horst, Cobb, & Meara, 1998). Waring and Nation (2004) estimated that native speakers of English know roughly 10,000 to 13,000-word families by adulthood. When these native speakers are children, they would likely know how to use 4,000 to 5,000-word families in speech production before attending elementary school (Waring & Nation, 2004). This sharp increase may be attributed to learning through classroom lessons and developing the child's literacy in their first language (Nagy et al., 1985). There are two distinctions of vocabulary acquisition, incidental and intentional (Nation, 2007). Incidentally learned vocabulary is seemingly picked up, whereas vocabulary gains from intentional learning are systematically taught to the learner (Nation, 2007). Both types of vocabulary learning occur in first and second language learners (Horst,

Cobb, & Meara, 1998). Studies of young second language learners have shown that these children learn new vocabulary at similar rates to native-speaker children (Jamison, 1978, as cited in Horst, Cobb, & Meara, 1998). In contrast, case studies with individual adult learners are seemingly not primed for acquisition as children are and may learn a language more successfully through intentional means (Lardiere, 2007). However, determining that children learn languages incidentally whereas adults learn languages intentionally may be an oversimplification. This thesis features studies which investigated the efficacy of incidental learning on vocabulary development across a myriad of input types in children and adult learner populations.

Krashen (1982) claimed that comprehensible input alone could facilitate language development in the Input Hypothesis. Language is acquired rather than learned (Krashen, 1982). The linguistic elements of language input, such as grammatical rules or vocabulary, are osmosed through exposure either through reading or listening (Krashen, 1982). When language learners read for enjoyment, their attention is on meaning rather than the linguistic forms presented in the text (Krashen, 1982). Some foundational aspects of extended reading lend credence to Krashen's (1985) hypothesis. Namely, that large quantities of input aid learners in acquiring vocabulary (Nation & Waring, 2004). A caveat of the Input Hypothesis (Krashen, 1985) is that input should be comprehensible to the learner and allow for opportunities of lexical growth by providing new vocabulary words through comprehensible linguistic input. Regarding reading, learners should know between 95% and 98% of the words within a text (Hu & Nation, 2000). These benchmarks ensure that the text would not frustrate the learners but provide input fostering lexical growth (Hu & Nation, 2000). Krashen (1985) asserted a bold claim that

input and a positive affect were the essential elements for successful language acquisition. This claim has been disputed (Ellis, 2015). However, a substantial amount of research supports the efficacy of extensive reading on vocabulary development (Waring & Nation, 2004). The following section may indicate that extensive input can address many learners' linguistic needs. However, the amount of time and linguistic input required for substantial vocabulary gains is relatively high (Nation, 2014).

In studying native-speaker children, Nagy, Herman, and Anderson (1985) observed that incidental vocabulary learning is a significant contributor to the development of a first language. It was estimated that native speakers are exposed to roughly one million words a year (Nagy et al., 1985). To meet this same level of language input through the reading of graded readers, language learners would need to read approximately 166 6,000-word graded readers in a year (Webb & Nation, 2017). Contrary to Krashen's (1985) emphasis exclusively on input, Nation (2007) suggests that while meaning-focused input is a crucial part of language acquisition, it should only consist of 25% of a language curriculum. When learners are exposed to linguistic input, they may incidentally learn lexical information while their attention is focused on comprehending the input (Krashen, 1989; Nagy et al., 1985). Nagy, Herman, and Anderson (1985), while observing incidental learning of vocabulary gains of native-speaking children, coined the Incidental Learning Hypothesis. It was argued that through engaging in free-reading, children would incidentally pick up linguistic knowledge while their focus was on understanding the text's story (Nagy et al., 1985). The caveat to this is that it should be comprehensible to the reader or listener (Krashen, 1985; Mesmer, 2008).

Cognitive Theory of Multimedia Learning

To investigate the role in which multimodality plays in the learning processes, Mayer's (2001) Multimedia Learning Principles lend credence to the learning potential of multimedia presentations as a result of the synthesis of associated mixed-media messages. Mayer's extensive research catalogue has established evidence-based support for the multimedia principle, the thesis of which is summed up as "people learn more deeply from words and pictures than from words alone" (Plass et al., 2020, p. 57).

Assumptions. Several assumptions of the Multimedia Principle allude to Mayer's theory regarding the learning processes and the cognitive involvement therein, which draws from cognitive-informed research, namely Baddeley's (1999) model of working memory (Mayer, 2001). The first of which is the dual-coding theory, in which incoming information is processed in two separate systems: one for verbal information, such as spoken and written texts, and another for non-verbal information, such as imagery or sounds (Clark & Paivo, 1991). Baddeley (1999) proposed an amendment to this conceptualization in their model of working memory by focusing on the sensory modality in which the information is delivered to the learner. In this stance, learners process visual materials, such as imagery or text, through their eyes and spoken language and background sounds through their ears (Baddeley, 1999). These two channels have been distinguished as the visuospatial sketchpad. Visual input is stored into the working memory, and the phonological loop is where the auditory information is briefly stored (Baddeley, 1999). This latter model was used by Mayer (2021) in devising the multimedia learning theory, as they were interested in a learner's cognitive interplay between these two channels to create a multimodal whole. For example, an academic lecture, while auditory input

through the ears, could be visualized mentally by the student through the construction of mental images and thus processed through the visual channel (Mayer, 2021)

The second assumption of the cognitive theory of multimedia learning is that the working memory contains a limited capacity for information input that can be processed in each of the channels mentioned above (Mayer, 2021). When sensory information is taken in by a learner, either through their eyes or ears, only a fraction of that information can be held by the learner (Cowan, 2010). Mental representations of visual information, such as static and moving imagery or written text, is fleetingly stored on the visual-spatial sketchpad (Baddeley, 1999). Sensory input taken in through the ears is stored temporarily through the phonological loop (Baddeley, 1999). The quantity of sensory information input varies between individuals' memory processing capacity, which, on average, could be somewhere between four and seven chunks of information (Cowan, 2010). Due to this limited processing capacity, the central executive system can employ metacognitive strategies to build connections between pieces of incoming information, thus limiting the cognitive load that the sensory input would demand of the learners' cognitive system (Baddeley, 1999). This interconnection between input types is a crucial component of the multimedia learning theory (Mayer, 2021).

The third assumption for this theory is that the learner actively processes incoming information (Mayer, 2021). When viewing a multimedia presentation, learners are pushed into active processing strategies such as paying attention to relevant information, organizing it, and integrating it into their larger mental schema (Mayer, 2021). These active processing strategies can be used to create a knowledge structure, a way in which the information is mentally represented, such as a cause-and-effect, comparison, generalization, enumeration, and

classification structures (Cook & Mayer, 1988, p. 59). Understanding multimedia input encourages learners to actively construct these knowledge structures to make sense of the sensory-rich material (Mayer, 2021).

In their Cognitive Theory of Multimedia Learning, Mayer (2021) outlines five processes that a learner must engage in for learning to occur. The following processes are of particular interest of the present study as vocabulary learning is the core value of this endeavour. While Mayer's (2021) work has informed educational multimedia design, presenting material to learners coherently to construct the above mental structures, the present study focuses on pre-existing material designed for entertainment purposes. The hypothesis is that commercial-off-the-shelf video games present players with lexical information that aid language learners in selecting, organizing, and integrating target vocabulary.

The first process involves the learner selecting relevant words from a multimedia presentation, which could include a combination of sensory input such as verbal input and on-screen or written text (Mayer, 2021). Learners can attend to some of the words presented in a multimedia message as they pass through the auditory sensory memory, in the case of verbal input, or the visual sensory stores if the linguistic input is written text (Mayer, 2021). As dictated in the limited capacity assumption, learners are limited in which words they can mentally store. Through executive functioning, learners select the relevant words to actively make sense of the incoming message (Mayer, 2021).

The second process that the learners could engage in, is in the selection of relevant images. In this process, the sensory input of pictorial representations is visually collected and processed to the viewer's working memory. Like the above process, viewers have a limited

processing capacity for this channel. Learners are unlikely to process an entire complex image or animation thus they may only attend to specific details in part. As learners actively seek to create meaning from imagery, akin to verbal input, they are likely to pay attention to the most relevant information to make sense of the incoming message (Mayer, 2021).

The subsequent two processes Mayer (2021) asserts is that the incoming information, either verbal or pictorial, is then organized within the learner's working memory through two separate channels. In the verbal model, words, either spoken or written, are organized by making connections to make sense of the linguistic information (Mayer, 2021). Active processing plays a key role in this process to aid the learners in making sense through constructing relationships between the incoming messages to create a holistic understanding of the information presented. Imagery is organized in the working memory to build connections between incoming pictorial message segments (Mayer, 2021). These connections form the basis of the pictorial model, which is a mental representation that aids the learner in making meaning for the incoming imagery (Mayer, 2021).

The final process, and most vital for the purposes of the present study, is that of the integration of word-based and picture-based representations. In this process, the learner connects the separate verbal and pictorial models into an integrated representation that draws upon the learner's established schema from their long-term memory (Mayer, 2021). Like the above processes, engaging in this active processing requires the learner's motivation to comprehend the incoming multimedia message. This integration, while requiring a sizeable cognitive demand, allows for long-term encoding of the information being presented (Mayer, 2021)

Mitigating Extraneous Processing. The above processes may overwhelm a learner's working memory capacity. One such cognitive demand is extraneous processing, which is cognitive processing because of poor instructional design (Mayer, 2021, p. 66). The learner may be involved in excessive processing to make sense of the multimedia message, leading to a lack of learning. Thus, one of the goals of Mayer's (2012) Cognitive Theory of Multimedia Learning is to reduce the cognitive load through the informed design of instructional multimedia instruction. Mayer (2021) outlines 12 design principles for instructional multimedia materials that mitigate the cognitive load for learners. Educational video games, while informed by instructional design, have not yielded the same learning potential as their entertainment-based counterparts (Dixon et al., 2022). Thus, this study focused on commercial-off-the-shelf games that were not designed to meet an educational goal but may still adhere to the design principles of multimedia learning intrinsically (Mayer, 2014).

The present study is concerned with supportive imagery presented through video games for the purposes of vocabulary learning. Thus, Mayer's (2012) temporal contiguity principle, which states that narration and corresponding graphics should be presented simultaneously (p. 186), guided the design of this study. In a study by Mayer and Anderson (1991), it was found that their participants scored higher on a transfer test when spoken words and corresponding imagery were presented simultaneously rather than sequentially. The temporal distance between text, either spoken or written, and imagery had a significant effect on the cognitive demand of the material. Concurrent presentation of language and imagery prevents unnecessary processing involved in the integration of the verbal and pictorial models in the learner's working memory (Mayer et al., 2020).

2.2 Language Learning Through Meaning-focused Input

Nation (2007) established five facilitating conditions necessary for language learning through meaning-focused input: large quantities, familiarity, comprehensibility, interest, and opportunities to gain knowledge. Video games have the potential to satisfy each of Nation's (2007) conditions (Rodgers & Heidt, 2020). Should language learners know the 3,000 most frequently used words within the British National Corpus and the Corpus of Contemporary American English (BNC/COCA; Nation, 2017), video games provide comprehensible input that could be processed in large quantities, similarly to that of a season of a television drama (i.e., large quantities and comprehensibility; Rodgers & Heidt, 2020). Television has been shown to aid LLs in developing fluency of high-frequency vocabulary while also providing contextual support to learn unknown words (Webb & Rodgers, 2009a; Rodgers, 2018). Like television, video games satisfy Nation's (2007) condition that meaning-focused input provides opportunities to gain knowledge of vocabulary. Nation (2007) suggested that language learners can learn lexical information through context clues and previous knowledge. This research considers the multimodality of visual, auditory, and interactive aspects of video games as contextual aids for language learning.

Extensive Input

Day and Bamford (1998) outline various principles that extensive reading programs should follow. First, reading material should be easy (Day & Bamford, 1998). Krashen (1982) may identify this as comprehensible input as well, Hu and Nation (2008) determined that 95% lexical coverage of a text could lead to a pleasurable read as language learners are not likely to encounter too many unfamiliar words that would take away from a global understanding of the

text. Similar findings have been reached for other input types (Webb & Rodgers, 2011; van Zeeland & Schmitt, 2013). The second principle expressed a need for various reading materials on a wide range of topics (Day & Bamford, 1998). This is often a critique of extensive reading, as providing learners could be costly (Webb & Nation, 2017). However, the benefit of having a wealth of topics to read allows learners to explore texts that they enjoy and thus aid learners to develop a positive affect towards reading (Garnier & Schmitt, 2016). Day and Bamford (1998) suggest that the third principle is that learners should choose what they want to read to continue this line of reasoning. Another principal stated that learners should read as much as possible (Day & Bamford, 1998). An extensive reading program for language learners with a foundation of the 1,000 most-frequent words would require students to read roughly 200,000 words over 40 weeks (Nation, 2014). Day and Bamford (1998) suggest that reading should be related to pleasure, information, and general understanding. Extensive reading may contrast with intensive reading, where the learners' goal is to find specific information in a text (Lightbown et al., 2002). The following principle is that reading should be its own reward (Day & Bamford, 1998). Day and Bamford (1998) suggest that reading should be both quick and completed individually. Various studies have observed the individual fluency gains that learners have experienced through extensive reading (Iwahori, 2008). Finally, the last principle is that the teachers' role is the facilitator of the reading experience (Day & Bamford, 1998). Day and Bamford (1998) encouraged active readership from language learners. However, these same principles could be applied to other mediums of extensive consumption of linguistic input.

Reading. Most of the research regarding extensive input is mainly concerned with reading and the incidental acquisition of vocabulary. The vocabulary acquisition process can be

described in three steps: encountering, establishing, and elaborating (Nation, 2001). First, learners may encounter an unfamiliar word within a story. This requires attention on the learner's part to notice that this word is unknown to them (Nation, 2007). This can be achieved by providing learners with considerable comprehensible input (Krashen, 1982). Learners need to commit these new lexical items to their procedural memory (Nation, 2007). This may be fostered through various learning conditions discussed below, such as repetition and varied encounters (Nation, 2007). Waring and Takaki (2003) describe elaboration as expanding word knowledge that the learner has internally conceptualized. This could be knowing the grammatical rules, the derivational forms, and the semantic information (Waring & Takaki, 2003). Similarly, this can be supported through extensive reading practices that have been shown to provide repetition, varied encounters, and allow learners to guess from context (Nation, 2014).

Listening. Linguistic input through listening can come from various sources, such as radio, listening to audiobooks or podcasts. Evidence suggests learners can learn implicitly through listening to input (van Zeeland & Schmitt, 2013; Elley, 1989). As with reading, to benefit from implicit learning, a learner would still need to recognize 95% of the vocabulary present within the input (van Zeeland & Schmitt, 2013). This number could be reduced with linguistic support of interactive modifications (Long, 1983, as cited in Ellis, 2015), but the focus of this paper is on one-way input. Elley (1989) read a combination of children's stories written for native speakers and graded readers out loud to second-language learners to research the incidental acquisition of spoken input. It was observed, that while acquisition did occur from linguistic exposure, the amount of input provided was insufficient to see significant vocabulary

gains (Elley, 1989). Listening to authentic language input can be beneficial for vocabulary learning. A study by Nation (2006), looked at non-interactionally modified conversations between two native speakers, with language-learner participants listening for comprehension (Nation, 2006). Nation (2006) noted that listening to the informal language use in these conversations were favourable for second language development. This was determined to result from the more limited vocabulary often being recycled in informal conversation than in formal scripted language indicative of the supplemental audio recordings appended to many second-language textbooks at the time (Nation, 2006).

Viewing Audio-visual Input. Another form of listening input is watching television programmes and films. This potential source of authentic target language input is distinct from other listening practices as it features an audio-visual component (Rodgers, 2018). A significant strength of television and films are their accessibility and availability, even in an English as a foreign language context (Webb & Rodgers, 2009a; Webb & Rodgers, 2009b). Watching television and films is a popular extramural activity for young language learners (Lin, 2014). Lin (2014) indicated that streaming services have over 200-million viewers in Europe, and of these viewers, 84% of Europeans reported that they watch television daily. This emphasizes the importance that television and film have on language exposure (Lin, 2014)

The seemingly passive act of watching TV has been shown to foster language development when consumed in the target language. With the support of captions in the target language, exposure to television aided learners in developing proficiency in reading and listening activities (Lindgren & Munoz, 2013). Television has been shown to be a source of incidental vocabulary acquisition for language learners (Webb & Rodgers, 2009a; Peters &

Webb, 2018). However, the caveat remains that the linguistic input does need to be comprehensible to the viewer (Webb & Rodgers, 2009a). Webb and Rodgers (2009a) determined that 95% lexical coverage seems to be sufficient for comprehension, similarly, to reading and listening comprehension. However, there are inherent affordances of this medium and other sources of language exposure that aid learners in comprehension (Peters & Webb 2018).

2.3 Factors That Can Affect Vocabulary Acquisition

While 95% of lexical coverage may be necessary for listening comprehension, various factors have been shown to aid learners in comprehending new vocabulary. It is potentially unlikely that learners would know the most common 5,000 English vocabulary items to fully benefit from linguistic input from video games (Rodgers & Heidt, 2020). These factors, present within video games, may benefit learners to comprehend the language input. This could imply that, while 98% or 95% lexical coverage is ideal for vocabulary acquisition, this medium's affordances may lower the learners' lexical load.

Frequency of Occurrence

Using frequently occurring words within the British National Corpus and the Corpus of Contemporary American English, Nation (2014) determined that learners can progress to subsequent 1,000-word increments of low-frequency vocabulary items through reading. Nation and Wang (1999, as cited in Nation, 2014) conducted a corpus analysis of graded readers and concluded within a 1,000-level word frequency band. It was estimated that learners would repeatedly encounter a word 12 times while reading at a leisurely pace of 150 words per minute (Nation, 2014). To this end, should a learner who knows the first 1,000 word-frequency

level wish to progress their vocabulary to the second 1,000 word-frequency level, "learners would need to read 200,000 words per year at a rate of 33 minutes per week" (Webb & Nation, 2017, p. 91). These estimates are dependent on the readers using specifically graded readers that are written to their present linguistic abilities. One way that graded readers accommodate vocabulary repetition is to have these level-appropriate lexical items recycled over several encounters within multiple reading sessions with a text (Nation & Wang, 1999, as cited in Schmitt, 2020). In this, graded readers have since manufactured natural encounters of these vocabulary items to promote incidental learning (Nation, 2014).

In a previously mentioned study, it was observed that natural-authentic language use in conversations generally used a limited range of vocabulary (Nation, 2006). This meant that the language present in these samples consisted of the most-frequently-used word families (Van Zeeland & Schmitt, 2013), as well many lexical items were recycled by the speakers (Nation, 2006). The frequently occurring words could aid language learners in comprehension (Nation, 2006). Similarly, a study by Saragi, Nation, and Meister (1978, as cited in Nation & Waring, 2004) recorded that participants learned 98% of the words that they encountered six or more times. In contrast, tokens encountered less than six times were learned by only half the participants (p.17). Thus, it could be determined that frequent exposure to lexical items through reading and listening is a factor that aids in lexical development.

2.4 Digital Game-based Language Learning

A recent meta-analysis of digital game-based language learning concluded that video games support vocabulary learning (Dixon et al., 2022). Previous meta-analyses indicate that action games (d = 1.87) were more supportive of vocabulary acquisition than non-action games

(d = 0.705; Chen et al., 2018), and those studies comparing DGBLL with other non-gaming activities had a large effect size (d = 0.99) by measures of vocabulary learning (Tsai & Tsai, 2018). The most current meta-analysis on the topic of DGBLL reported a small-to-medium effect size regarding vocabulary gains through L2 gaming (d = 0.59; Dixon et al., 2022). Thus, there is a consistently positive effect on vocabulary learning through digital gaming practices. However, it should be noted that there is a considerable distance between the lower and upper confidence intervals of this metric (Lower = -0.09, Upper = 1.84; Dixon et al., 2022). As digital games vary in their purpose, the number of players, and the language used in play, these factors could account for the differences in the intervals.

Digital game-based language learning has focused on two broad categories of games for language learning: those intended for education and those designed for entertainment purposes (Reinhardt, 2019). Contrary to the intention of games' purpose, studies which focused on entertainment games seem to produce greater effect sizes (N= 12, Md = 2.03, SE = 0.37) than those that used games designed for educational purposes (N= 17, Md = 1.13, SE = 0.15; Dixon et al., 2022). This echoes the argument that games designed solely for educational purposes lack meaningful L2 language as the gaming industry may have low investment in designing pedagogically innovative products (Reinhardt, 2019; Thorne et al., 2012). Educational games, such as Duolingo, often rely on grammar translation activities that may not foster learner engagement more than entertainment-based games (Loewen et al., 2019, as cited in Dixon et al., 2022).

Digital games, similar to their analog counterparts, can vary in the number of players that can engage in play (Reinhardt, 2019; York et al., 2021). Video games can be played alone

(i.e., single-player games), played with multiple players (i.e., multiplayer games) or, unlike analog games, played with thousands of others in a shared game environment (i.e., massively-multiplayer online games [MMO]; Dixon et al., 2022, p. 3). The latter has garnered the attention of SLA researchers, as the potential for learning may increase as opportunities for both players' linguistic input and output increases due to more interlocutors within a shared game space (Sundqvist, 2019). The meta-analysis of DGLL indicated that, contrary to the previous statement, the number of players has mixed results on L2 development. Single-player games presented a more positive effect on vocabulary acquisition (n = 17, d=1.45 SE = 0.217) than MMOs (n= 9, d = 0.59, SE = 0.272; Dixon et al., 2022). The more favourable learning in single-player games may be due to the cognitive demands of multiplayer and MMO games, due to the spontaneous use of both text and speech to communicate with other players (Reinders & Wattana, 2012).

Games can vary in the type of language in which the information is presented and used by the player (Reinhardt, 2019). Language input from digital games, both spoken and text-based, can work in conjunction to present the player with the ludic skills and abilities to interact and take part in the game's wider plot (Heidt, 2020). While multiple modes of linguistic input could potentially overwhelm language learners (Reindeers & Wattana, 2012), the alignment of these modes may work when in accordance to aid vocabulary development (Jensen, 2017). Single-player video games may require no linguistic output, whereas multiplayer and MMO games often require players to coordinate through spoken or text-based communication (Reinhardt, 2019). In the meta-analysis conducted by Dixon and colleagues (2022), it was observed that games requiring no linguistic output had the most positive effect on linguistic

goals (N= 14, d = 1.60, SE = 0.244). This finding could suggest that the simultaneous input and output required of multiplayer and MMO games could affect cognitive load making vocabulary less salient (Dixon et al., 2022). It should be noted, that single-player experiences offer temporal affordances as the game can be paused, allowing more time for the language to be processed should the learner require it (deHaan et al., 2010; Reinhardt, 2019).

The oft-cited research conducted by deHaan and colleagues (2010) has been used to critique the efficacy of entertainment video games for vocabulary learning. In this study, students with shared target language proficiencies were assigned to two groups; one would play the game, while the other group would view the game being played. In both immediate (r = 23.27; 11.18) and delayed post-test (16.03; 7.35) measures, watchers outperformed players (immediate post test: r = 7.42, 5.17; delayed post test: r= 5.15, 4.88; DeHaan, 2012a; DeHaan et al.,2010). The group who watched the game *PaRappa the Rapper 2* learned almost twice as much vocabulary as those who played it (deHaan et al., 2010). This led to the conclusion that the physical interactivity of playing a game prevented players from noticing vocabulary due to the increased cognitive load required for gameplay (deHaan et al., 2010). deHaan and colleagues (2010) point out that while the players were immersed in playing the game, they were not immersed in interacting with the language input from the game. A recent meta-analysis on digital game-based language learning flagged this study as an outsider when compared to other studies concerned with vocabulary learning through entertainment-based video games (Dixon et al., 2022). However, this variance in video game efficacy on vocabulary learning is worth considering. There could be multimodal presentation inherent in games that could positively or negatively affect vocabulary learning gains (Mayer, 2014).

Digital Games as Multimodal Texts

Multimodality uses multiple means to create meaning (Jewitt et al., 2016). Through gameplay, language learners can attend to all semiotic resources to create multimodal wholes (Jewitt et al.,2016). This multimodal analysis explores the semiotic resources within video games that may aid learners in constructing meaning through multimodal means, such as imagery, interaction, and aural input. Video games are inherently multimodal by design (Gee, 2007). That is, digital games are presented to the player by means of various semiotic modes, particularly through audio, textual, visual, and in-game interactions that encompass meaning through gameplay (Gee, 207; Ensslin & Balteiro, 2019). Ensslin and Balteiro refer to this as going "beyond the 'text" in their volume Approaches to Video Game Discourse: Lexis, Interaction, Textuality (2019, p. 225). Video games present players with a semiotically rich experience that can provide ample learning opportunities (Gee, 2007; Mayer, 2014). The interplay between the various modes can be mutually modifying for language learning purposes (Jewitt et al., 2016).

Visual Mode. The use of images can aid in referential comprehension when reading in a second language (Altarriba & Knickerbocker, 2011). However, compared to other aids in comprehension, such as direct translations, supportive imagery is less effective for second language learning (Altarriba & Knickerbocker, 2011). Comic books written for native speakers were also less effective than graded readers, despite providing learners with both words and images (Iwahori, 2008). However, through auditory input, the imagery seems to aid learners in comprehension (Rodgers, 2018). Spoken input alongside co-occurring images was seen to foster learning by having language learners adequately guess new vocabulary items through context (Rodgers, 2018). Despite the mixed findings for supportive imagery above, Rodgers' (2018)

study identified that television was in line with the Cognitive Theory of Multimedia Learning (Mayer, 2021). The temporality of simultaneous spoken text and on-screen images, beneficial for designing multimedia instructional materials, occurred naturalistically in television programs (Rodgers, 2018). Within the context of video games, multimedia learning could support the claim that the co-occurring spoken dialogue, captions, and on-screen images within video games can foster vocabulary learning (Mayer, 2014).

Procedural Mode. Video games present a relatively new medium of a virtual space. Thus, the nature of interaction may take on a new conceptualization (Hartwick, 2018). Within a virtual environment, language learners can both interact with each other and the environment itself (Hartwick, 2018). Peterson (2011) investigated the interaction between participants during a 60-minute session of an online game. The game possessed task-like qualities, in that there was a communicative goal in which participants needed to interact with one another to achieve the goal (Peterson, 2011). Within the game space, new vocabulary items were learned due to interactional modifications between learners (Peterson, 2011). Similar findings were concluded in comparing learner-to-learner interaction in virtual space versus interaction within a classroom environment (Hartwick, 2018). Hawreliak (2019) proposed that this interaction with the virtual elements provided in video games is a means of meaning-making. This gameplay is referred to as the procedural mode, which involves the active construction of meaning through carrying out the procedure of a games' programmed actions and subsequent reactions while adhering to the games' defined system, rules, and parameters (Hawreliak, 2019). Hawreliak (2019) suggests that this procedure of gameplay "should be viewed as a semiotic equal alongside established modes, such as text, image, and music" (p. 228). This

interaction with a video game's elements could aid learners in comprehending new vocabulary items. Video games afford a level of interaction that may not be achievable through reading or viewing television or films, which should warrant an investigation into the language learning potential of video gaming as a language learning tool (Rodgers & Heidt, 2020). In this regard, there could be opportunities to explore the rate of vocabulary acquisition through engaging with virtual environments.

2.5 The Present Study

The multimodally rich experience that video games can provide ample learning opportunities (Mayer, 2014). While language learning through gaming has been attributed to the affordances of video games (Kuppins, 2010; Peterson, 2011), few studies have actually investigated how the digital game format presents a favourable learning environment for vocabulary acquisition. The present study seeks to address this gap by investigating the multimodal learning opportunities present in commercial-off-the-shelf video games. This study draws on the work conducted by Rodgers (2018) which applied the Multimedia Principles (Mayer, 2021) to identify the concurrence of imagery and spoken text in television programs that present favourable conditions for vocabulary acquisition. This research is in line with a larger field of investigations exploring multimodal supports for learning (Jewitt et al., 2016). The methodology used to investigate employs a multimodal approach that explores how aural input of target words are supported visually through the imagery present within a digital game. As an extension to the initial study by Rodgers (2018), the present study applied a multimodal analysis to 10 commercial games to examine how spoken dialogue is supported through multiple means: co-occurrence of imagery, written text, and player-game interactions.

Investigating these semiotic resources could provide insight into the affordances of this medium on vocabulary learning that occurs authentically through the playing of video games. An additional extension to Rodgers' (2018) study is to perform a media comparison of the supportive imagery between digital games and television. As the imagery in television is an advantage to vocabulary learning (Ghebghoub, 2021; Rodgers, 2018), visuals presented in video games could lead to similar benefits. Thus, the following questions seek to investigate the multimodal affordances for vocabulary learning in commercial-off-the-shelf video games:

1. What is the occurrence of:
    a. Visual supports for aurally occurring target words in commercial off-the-shelf video games?
    b. Textual supports for aurally occurring target words in commercial off-the-shelf video games?
2. What are the opportunities for target words to be interacted with in a virtual environment presented in commercial-off-the-shelf video games?
3. How does concurrent or near-concurrent imagery and aural provision compare between:
    a. two video games genres
    b. television programmes?

Chapter 3: Methods

This chapter presents the methodology used for a multimodal analysis to investigate the semiotic supports for the potential learning of new vocabulary from commercial-off-the-shelf (COTS) video games. Studies have observed the effects of game-based language learning from the extramural engagement of video gaming (Reinhardt, 2019), which have yielded favourable results on vocabulary learning (Dixon et al., 2022; Tsai & Tsai, 2018; Chen et al., 2018). These investigations often credit favourable language learning affordances to the digital gaming format. However, little attention has been given to how video games present language learners with a supportive language-learning environment. Thus, this study used a multimodal analysis to identify how target items were supported through multiple means in the medium of video games. This chapter will discuss the investigation of the multimodal supports in authentic video games and the methodology used to collect them.

3.1 Video Games

Selection Criteria

The meta-analysis on digital game-based language learning conducted by Dixon et al. (2022) informed three criteria for the selection of video games for the present study. First, only single-player games were selected. These games involve a human taking the role of a player which involves embodying a virtual avatar which can interact with elements of the game world, such as non-playable characters and artifacts. The player can interact with non-playable characters by communicating in a preprogrammed dialog or in some instances engage in combat. Unlike a multiplayer game or an MMO, these non-playable characters are not piloted by other humans. In multiplayer or MMO games communication can be spontaneous and

carried out in real-time through text and speech. In contrast, the dialogue in single-player games is between the human player's avatar and other non-playable characters, thus the communication is pre-recorded dialogue supplied by voice actors. The pre-recorded dialogue present in this sample of games was the basis for a textual analysis.

The second selection category for digital games in this study were games which featured no player-supplied linguistic output. Converse to the hypothesis put forth by Sundqvist (2012), the more interlocutors present in a game space may not lead to increased learning opportunities (Dixon et al., 2022). This may be a result of the increased cognitive load placed on a language learner due to the spontaneity of real-time communication and simultaneous input and output (Reinders & Wattana, 2012).

The third criteria that showed favourable second-language learning potential was that games were selected by their intent for entertainment purposes (Dixon et al., 2022). The games in this study are across two distinct genres, action games and interactive fiction games. Both of which are commercial off-the-shelf games designed for entertainment rather than for educational purposes.

Walkthroughs

Video walkthroughs hosted on the video streaming website YouTube were chosen as a unit of analysis. These videos can be viewed to watch narrative of the game's story without requiring the viewers to engage in play. Video walkthroughs, initially intended to aid other players by modelling game segments, were determined to represent a typical player experience (Rodgers & Heidt, 2020). The videos in this study were uploaded by the YouTube user *MKIceAndFire* who recorded their playthrough of each of the games without providing

commentary. Some of these walkthroughs were single videos, while others were compiled into a playlist due to video length restrictions on YouTube. The video lengths in the playlists were added together for the purposes of data logging. As these videos are primarily for instructional purposes, many feature voiceovers that explain ludic elements to the viewers to develop their skill in a particular game. For this study, videos containing these commentaries were excluded, as the instructional dialogue would deter from the language presented in the game. Following the call for investigation into the language presented through video games by Schmitt (2020), Rodgers and Heidt (2021) determined that video games holistically presented players with low-frequency vocabulary that would be unlikely for a language learner know (i.e., 3,000-5,000 levels of the BNC/COCA). However, the narrative present within a video game's story presented players with language that could be at an approachable lexical coverage for a L2 learner (Rodgers & Heidt, 2021). The texts, comprised of the dialogue within video walkthroughs of each game's narrative plot, were transcribed to conduct a textual analysis for this study (See Table 3.1).

Genre

The field of DBGLL is still in its infancy and thus the sample of video games was exploratory. The ten games within the sample were distributed amongst two broad game genres. Five games are associated under the banner of action games which were composed of action-adventure games and action-roleplaying games (Reinhardt, 2019). Video games within this category are "sensorially rich 2D or 3D virtual simulations that afford multimodal goal-oriented interaction with content" (Cornillie, 2022, p. 275). While these games could be played alone or with others (Reinders & Wattana, 2014), the games in this sample are of the single-

player variety. The rationale for choosing this genre is that it is perhaps the prototypical example of contemporary gaming, in which human players are immersed in the game by embodying a virtual avatar which they can carry out actions within the digital game environments (Reinhardt, 2019; Cornillie, 2022). These interactions with the virtual space form the intention behind answering the second research question. The second game genre chosen was that of interactive fiction, which has been described as story in which the player can insert themselves into the plot by having virtual characters select dialogue choices which result in a non-linear, personalized narrative (Lee, 2019). Interactive fiction games were chosen in this study, as they represent a digital gaming genre that is perhaps the most comparable to television. The emphasis of gameplay is not on the actions the avatar is required to make, but on the non-linear story that the player can design unlike that of the presubscribed story in television or film (Lee, 2019). The rationale for selecting five games in these two categories was to observe if video game genre informed the quantity of multimodal supports such as imagery, text or procedurality, that could affect learning potential for vocabulary.

3.2 Target Word Selection

As each research questions focuses on the multimodal supports for vocabulary items, target word identification was a central component to this investigation. The spoken dialogue of each video game was the basis of the textual analysis. To this end, target vocabulary in this study was selected using similar methods to previous corpus-based studies that investigated the language within other audio-visual input sources, such as television (Rodgers & Webb, 2009a)

Lexical Profile of Video Game Texts

To determine which target items would be selected for this study, a lexical profile was conducted. Lexical profiling software, in conjunction with reference corpora or word lists, can be used to conduct corpus-driven analyses that identify the distribution of words within a text (Webb & Nation, 2017). This study compared the spoken dialogue present in each game to Nation's (2017) frequency-based word lists of the British National Corpus and Corpus of Contemporary American English (BNC/COCA), to determine which words occur at which frequency levels. The rationale for this procedure was twofold.

The primary reason behind conducting a lexical profile was to identify the frequency of target word occurrence within the spoken dialogue of a sample of COTS video games. This measure has been determined to be a contributing factor for potential word learning (Horst et al.,1998; Waring & Takaki, 2003; Webb, 2007). While research has indicated that repeated encounters with words positively affect incidental vocabulary learning, the minimum number of for such encounters for learning to occur is varied (Webb, 2007). Studies have shown that as few as 6 encounters (Rott, 1999) as many as 20 encounters (Waring & Takaki, 2003) for long-term learning to occur. Five encounters with target items in the spoken dialogue was chosen as the qualifying benchmark it has been shown to be sufficient for at least allowing learners to infer partial knowledge of a word from auditory input (Webb, 2007; Rodgers & Webb, 2009a). In a gaming-specific context, Rankin, Gold, and Gooch (2006) observed vocabulary gains in L2 users as a result of increased frequency of target items in an 18-hour session of an online roleplaying game. With language exposure from the online game, words encountered once were recognized by 35% of the learners, whereas words encountered five or more times were

recognized 55% of the time (Rankin et al., 2006). Additionally, to draw a comparison to Rodgers (2018), this study also used the benchmark of 5 or-more encounters for target item selection based on research on vocabulary acquisition through incidental listening (Rodgers & Webb, 2009a).

The second rationale for conducting a lexical profile was that a lexical analysis could be used to determine the relative appropriacy of video games as a tool for the potential to learn unknown words. Much research has been conducted to determine what percentage of vocabulary knowledge should a learner possess to comprehend most types of linguistic input (Webb & Nation, 2017). Based on the findings from Schmitt and Schmitt (2014), knowing the 3,000 most-frequent word families would enable students to comprehend at least 95% of the words in most spoken input types, such as conversation (van Zeeland & Schmitt, 2013), television shows (Rodgers & Webb, 2009b) and films (Webb & Rodgers, 2009a). In a gaming context, based on Rodgers and Heidt (2021), the lexical coverage needed to achieve 90% comprehension of video games was identified occur between the 3,000-4,000 levels of the BNC/COCA. However, investigations in the European context indicate that English-mediated gaming by young language learners is a substantial source of English-language exposure (Sundqvist & Sylven, 2014; Kuppins, 2010; Lindgren & Munoz, 2013). It may be unlikely that young learners may have the lexicon to reach the 3,000-4,000 benchmarks in their second language. However, in these same populations English-language proficiency has been attributed to this exposure to English-mediated media, such as video games or watching videos (Lindgren & Munoz, 2013), even in children who have not been enrolled in language classes (Jensen, 2017). Words beyond the 2,000 most frequently occurring words within the BNC/COCA (Nation,

2017) and beyond were included in this study, as it is likely that young language learners may engage with extramural gaming as their first exposure to English in an EFL context (Jensen, 2017). Through identifying vocabulary items' distribution within Nation's (2017) BNC/COCA, target items could be selected from the less frequent word lists. Similarly, to Rodgers (2018), by selecting these items, they would have the potential to be unknown by hypothetical language learners.

Method

Texts. To conduct a lexical profile of video game walkthroughs, a text-based format was needed for the software to analyze. Corpus-driven research concerned with television has an affordance of captions, which is an accurate text depiction of the spoken dialogue present in a television program that can be used for lexical analysis (d'Ydewalle et al., 1987). Video games, while also containing text-based captioning and subtitles, have this information encoded into the program's data files which require extensive decoding (Heidt, 2020). Thus, this study used a similar method to Rodgers and Heidt's 2021 study. In the follow-up analysis they employed a video walkthrough as a unit of analysis for the video game *Grand Theft Auto V* rather than the full game itself (Rodgers & Heidt, 2021*).* This present study was a methodological continuation of the follow-up analysis used by Rodgers and Heidt (2021), in that the texts used were comprised of the dialogue presented to the viewer as a player would likely experience it. For the purposes of this study, video walkthroughs were deemed an ecological alternative to recording participants engage in actual play. The lexical data transcribed from the spoken dialogue in these video game walkthroughs are displayed in table 3.1. The process in which this data was transcribed is in the following subsection.

Table 3.1

Video Game Texts

Title	Release Year	Genre	Tokens	Play Duration
Uncharted: Drake's Fortune	2007	Action	7,054	4:14:27
The Last of Us	2013	Action	19,203	11:48:16
Assassin's Creed: Origins	2017	Action	25,074	8:18:15
Ghost of Tsushima	2020	Action	34,588	12:26:07
The Elder Scrolls V: Skyrim	2011	Action	26,469	8:00:46
Beyond: Two Souls	2015	Interactive Fiction	15,079	8:10:32
Firewatch	2016	Interactive Fiction	12,370	3:20:23
Life is Strange	2015	Interactive Fiction	46,762	8:29:02
Twelve Minutes	2021	Interactive Fiction	8,320	2:31:00
Tales from the Borderlands	2014	Interactive Fiction	44,324	9:01:52

Transcription. Otter.ai, an automated speech-to-text software *Otter.ai* (Liang & Fu, 2021), was used to build the corpus for this study. Artificial intelligence has trained the software to produce accurate speech-to-text transcription through the use of a library of authentic speech samples (Liang & Fu, 2021). This automated transcription tool allowed for the extraction of the spoken dialogue from each walkthrough video file to a text format which could then be used for a corpus analysis of spoken dialogue present within the games.

While the Otter.ai transcription service was intended for a general consumer user-base (Trueman, 2022). It was a suitable tool for research purposes. In the proposal for this study, I had intended to use the Nvivo transcription tool (Version 12; OSR, 2020). However, the initial data collection yielded several errors in the transcription using Nvivo (Version 12; OSR, 2020). This could be due to the cacophony of the musical score, diegetic sound effects and spoken dialogue occurring simultaneously within the video game. Thus, I conducted a brief diagnostic test to compare the two transcription softwares. This was done by having both programs

transcribe the dialogue from a 5-minute segment of a game's video walkthrough. To serve as a reference standard for this assessment, I transcribed the segment manually. Transcription errors were marked in order to yield a word error rate percentage. Of the 137 words spoken in a five-minute segment of *Ghost of Tsushima*, this process yielded word-error rates of 6.57% for Otter.ai and 9.81% for Nvivo. As a result of this parametric testing, Otter.ai was chosen as the transcription tool for data collection in the present study. Errors that were still present after the automatic transcription process were fixed by listening to the associated timestamp and manually corrected with the proper word.

Lexical Analysis. The transcribed text files that were extracted from the video walkthroughs were compiled into separate texts that would serve as a representation of each game’s narrative. Using these transcribed text files, lexical profiles of each game were conducted using the free online software ANTWORD Profiler (Anthony, 2021). The created corpora for each game were compared with Nation’s (2017) BNC/COCA. This process identified the frequency of lexical items as they are used in accordance with the BNC/COCA word lists (Nation, 2017). The BNC/COCA (Nation, 2017) is a 25-level corpus. Each level is composed of 1,000-word families based on their frequency of appearance within spoken discourses, fiction, magazines, newspapers, and academic compositions (Nation, 2017). The rationale for choosing this corpus is that the BNC/COCA portrayed an authentic representation of the real-life English-language use indicative of everyday contexts (Nation, 2017). A further justification for using the BNC/COCA as a reference corpus is that the original research by Rodgers (2018), which was the inspiration for this extension study, used the BNC (Nation, 2006) as the reference corpus. Thus, to address the third research question, the findings for this study would be more in line to

make a fair comparison to narrative television and documentary genres. One of the selection criteria for target words was that they were selected from the 3,000-level wordlist of the BNC/COCA and beyond. This is a depart from the initial research conducted by Rodgers (2018) which used the 4,000-level word list as the benchmark. The rationale for this decision was that based on the extramural gaming habits of young language learners in the European context and the multimodal supports indicative of digital games which could support vocabulary learning through play.

The lexical profiling software ANTWORD Profiler (Anthony, 2021) was used to analyze each game's textual data against the 25-levels of the BNC/COCA (Nation, 2014). ANTWORD Profiler (Anthony, 2021) both identified target words for this study by their frequency of use within each game's text, as well as identified the lexical coverage of target words in accordance with the BNC/COCA (Nation, 2014). The frequency of occurrence of each potential target item was used to qualify the target words of this study. As mentioned previously, words featured with five-or-more occurrences were chosen as target items. The lexical analysis identified that, of the 249 target items across the ten games, occurrences of target word encounters ranged from five to 138. Moreover, in comparison with the BNC/COCA (Nation, 2014), target items were identified to be across an array of high to low frequency vocabulary. Target items, their number of occurrences in each game, and their associated frequency level in the BNC/COCA (Nation, 2014) are displayed in Appendix A and summarized in Table 3.2.

Table 3.2

Number of target items selected based on their frequency of occurrence in each game

Title	Genre	Target Items	Highest Freq.
Uncharted: Drake's Fortune	Action	9	23
The Last of Us	Action	14	32
The Elder Scrolls V: Skyrim	Action	28	20
Ghost of Tsushima	Action	48	138
Assassin's Creed Origins	Action	31	23
Beyond: Two Souls	Interactive Fiction	12	15
Firewatch	Interactive Fiction	10	14
Tales from the Borderlands	Interactive Fiction	29	39
Life is Strange	Interactive Fiction	58	32
Twelve Minutes	Interactive Fiction	7	16

3.3 Multimodal Supports for Target Words

With the target items identified within each text, the next step of this inquiry was to investigate how these target items were supported through multiple modes in the medium of video games. Multimodality is the use of multiple means to create meaning (Jewitt et al., 2016). In a gaming context, language learners can attend to all semiotic resources to create multimodal wholes through gameplay (Jewitt et al., 2016). This methodology was intended to be an extension to the initial research conducted by Rodgers (2018) which had identified visual supports for vocabulary learning through viewing television. The dialogue presented in television programs in conjunction with the supportive imagery of lexical items, is beneficial to LLs for acquiring unknown vocabulary (Durban et al., 2020, Ghebghoub, 2021). The present study expands this notion of supportive modes for vocabulary items by extending the analysis to additional modes such as text, and procedural gameplay. This research is in line with a larger field of investigations exploring multimodal supports for learning (Jewitt et al., 2016). The

methodology used to answer this question was a multimodal approach that involved the researcher to consider how aurally provided target words were supported by semiotic modes such as imagery, text, and procedurality within a digital game. This multimodal analysis explored the semiotic resources within video games that could aid learners in constructing meaning of vocabulary through multimodal means, such as imagery, written text, and procedures of gameplay.

To provide reliability measures in regard to the coding of each of the supportive modes, an external examiner was used to establish inter-rater reliability. Training in coding and cataloguing the visual, textual and procedural modes was provided to a native speaker of English who was familiar with the video streaming platform and digital gaming. 5.69% of the codes were collected and analyzed by the rater, to which there was 100% agreement between the coding conducted by external rater and the researcher's coding of the modes.

Visual Mode

Playing a video game requires the player to process both visual and verbal input. Under the dual-channel principle (Mayer, 2021), sensory information (e.g., auditory, and visual, in the case of this sample of video games) is processed as two separate channels. This information is stored in the player's sensory memory and, if not attended to, could be lost within a fraction of a second (i.e., the limited capacity principle; Mayer, 2014, p. 88). However, if the input is supported by simultaneous, or near-simultaneous, auditory-visual cues, the sensory information is more likely to be transferred to the observer's working memory (Mayer, 2014, p. 88). The multimodal flooding input inherent in COTS video games could be a source of extraneous processing. One way to mitigate these two sensory input channels' cognitive load is

the simultaneous provision of aural and visual referents (Mayer, 2014). Thus, learning can be enriched through simultaneous multimedia messages of corresponding visual and audio input (Mayer, 2014, p. 183). In the context of video games, this occurs from the concurrent visuals and spoken dialogue within the game's narrative (Mayer, 2021; Calleja, 2011).

Method. The inspiration for the present study investigated the relationship between aural and visual co-occurrence of target words within television programs (Rodgers, 2018). Thus, this study followed a similar procedure for data collection of the visual supports for vocabulary. In order to address the first research question, target word occurrence within the input was systematically recorded and timestamped using text-to-speech transcription. Target words that had been encountered five-or-more times through the duration of each video walkthrough were systematically logged. As carried out similarly by Rodgers (2018), aural provisions of target words were timestamped while viewing the content. The researcher took note of any on-screen visual depictions of the target word manually. These visual supports occurred as either a digital representation within the game's virtual environment or as on-screen text. All depictions of the visual representation of target words were timestamped for the duration of their on-screen appearances.

Imagery and aural provision target word referents were categorized occurring as either: concurrent, +/-2 seconds, +/- 5 seconds, and no occurrence within a 10-second timespan (Rodgers, 2018, p. 200; see table 4.2). The rationale for these categories was twofold. First, this thesis is an extension of a study conducted by Rodgers (2018), which investigated the alignment supportive imagery for vocabulary words in the spoken dialogue of television programs, thus the present study adopted a similar method of categorization to compare visual supports. The

categories of the +/-2 seconds and the +/- 5 seconds employed by Rodgers (2018) were based on a study investigating language learners' noticing of subtitles (d'Ydewalle et al., 1987, as cited in Rodgers, 2018), six seconds was the most extended length of time that subtitles remained on screen in television (d'Ydewalle et al., 1987, as cited in Rodgers, 2018), thus +/- 5 seconds was used to encapsulate a time frame no longer than 6 seconds. While the +/- 2 second was indicated as the shortest amount of exposure to subtitles in the study by d'Ydewalle, Van Rensbergen, and Pollet (1987, as cited in Rodgers, 2018). Language learners could still process this shorter provision of subtitles at rates similar to that of the 6-second timeframe (d'Ydewalle et al., 1987, as cited in Rodgers, 2018). The second rationale for the categories was that they were supported in a study conducted by Ghebghoub (2021), which tested the categorization of the alignment of imagery and spoken dialogue used in the initial study by Rodgers (2018). These instances are summarised in table 4.2 of the following chapter.

Written Mode

Aside from imagery, another visual support depicted in games is text. The cooccurrence of text and images provides a supportive environment for learning, often seen in the design of instructional materials such as textbooks or multimodal presentations (Mayer, 2021). As an extension to the initial study by Rodgers (2018), this study investigated the textual support of target words, as these could aid the learners in developing word knowledge through meaning-focused input (Nation, 2007). For this reason, this study analyzed textual concurrence with target word provision in the spoken dialogue. While subtitles and captioning were an option for these games, they were not included in the analysis. The rationale for this inquiry is to address

the first research question, as one of the affordances of games compared to television is that it may provide more text-based support

The games used in this study, while not text-reliant to engage in gameplay, use text to both support aspects of the narrative dietetically and to assist the players on a ludic level toward achievement of a goal within the game. Diegetic text varied in each of the games in this study. Some text took the form of collectible notes or even books that, while not central to the plot, provide the reader with insights into the holistic understanding of the narrative. Another way that diegetic text was used in the gameplay was through written riddles that the player was tasked with solving in order to progress the story. A game's user interface often depicted ludic text through tutorials, hints, in-game actions, labels, or equipment inventories. This text was intended to aid the player in the act of gameplay (Yorke et al., 2021; Heidt, 2020). Games analyzed for the purposes of this study presented ludic text in different quantities. Some games provided the player with a minimal interface that presented sparse text, while others maintained textual reminders of tasks or quests that the player needed to attend. Both diegetic and ludic text were included in data collection as they both provided supports for the identified target items.

Method. The same process was carried out for the occurrence of imagery in relation to text occurrence within each game (see table 4.4). The concurrent category aligns with the multimedia principle (i.e., temporal contiguity principle; Mayer et al., 1999), which states that simultaneous spoken text and on-screen images are beneficial for learning (Mayer et al., 1999). The time of any encounters where the on-screen text of a target word appeared was

logged as occurring either at concurrent, +/- 2 seconds, +/-5 seconds, or no occurrence. The duration of the text's on-screen appearance was recorded as well.

Procedural Mode

Playing a digital game requires communication between a player and the information presented by the game. This interaction is the “discourse that takes place between the player and the visual representation” (Sedig et al., 2017, p. 1) which culminates into an experience referred to as gameplay (Salen & Zimmerman, 2003). Gameplay requires players to actively interact with or manipulate various elements in the virtual environment as part of the procedure of play (Hawreliak, 2019). While this kind of interaction has been explored in the elicitation of LLs' engagement beyond the classroom (Kuppens 2010), few studies have speculated on this as additional support for vocabulary acquisition (Ginns, 2006).

Gameplay can vary across different games as they are influenced by game design elements such as information content, mechanics, graphics, goals, narrative, and rules (Salen & Zimmerman, 2003). Player-game interaction has been identified as an affordance for language learning (Reinhardt, 2019). However, the procedures of play, such as the actions and manipulations a player-character’s avatar performs, has only recently been considered as a semiotic resource (Hawreliak, 2019). The present study investigated the procedural mode within the sample of COTS video games to identify opportunities for the manipulation, presence, and engagement with target words as an additional support for vocabulary words.

Manipulation. This player-game interaction was operationalized as the player manually interacting with the game's interface, typically through the game's controller or keyboard (Sedig et al., 2017). This was observed in the video walkthroughs as the player expressing and

action to manipulate artifacts within the virtual environment of the game using a manual command through a controller. This could be in the manipulation of the player's avatar as in the target word [dodge] to have their character move out of the way of incoming danger or manipulating items from the environment as in [statue] to have their characters rotate a large stone sculpture. Compared to television or film, this category is unique to the video game medium and could benefit vocabulary development through bringing learner's attention to specific vocabulary items. The rationale for this category was that lexical items which could be attended to, by a player engaged in gameplay, would be more salient for a hypothetical language learner. Requiring an action on the part of a learner could aid them by implicitly increasing their awareness to the vocabulary item (Schmidt, 2001).

Presence. This element is characterised as the advertisement of the existence of an action to a player (Sedig et al., 2017, p. 10). In the games chosen for this study, avatars controlled by the players have the ability to explore the three-dimensional virtual environments at their own pace. As the narrative follows the perspective of the player's avatar's actions in the foreground, background content may not be intentionally interacted with through manual agency by the player. However, the presence of this content is still a part of the player's experience. For the purposes of this study, this presence was noted as target words that are present within the environment but couldn't be manipulated, such as in the target word [temple]. A player may observe or even explore a digital representation of a temple within the game space.

Ludic Engagement. The last category of observed player-game interactions within the playthroughs was that of goal-oriented, or ludic, interaction with the game's virtual elements.

Both target words and potential target words needed to be interacted with as a stipulation of gameplay to progress through the games' established narrative. This category involved the players in a combination of actions which may result in high-level cognition such as problem solving. An example of this is illustrated in the quality of interaction between the two lexical items [barrel] and [harmony] within the game *Uncharted: Drake's Fortune*. The two words in this example were encountered aurally once within the spoken narrative. Similarly, they both appeared textually, albeit ludically for [barrel] as a tooltip from the game to the player and diegetically for [harmony] as it appeared within a written text from the game's virtual world. The former example is optional for the player to engage with in gameplay because it largely depends on their playstyle. On the other hand, the latter example is a part of a problem-solving encounter where the player encountered a riddle that needed to be solved by having two separate bells ring simultaneously for a secret passage to open. These ludic activities are task-like in that they could involve a hypothetical language learner engaging with the target language to complete a goal (York et al, 2021). These task-like interactions could aid learners in identifying gaps in their present interlanguage by having learners implicitly engage with potentially unknown words similar to a structure-trapping task (Skehan & Foster, 2001).

Method. To address the second research question, this study recorded all the encounters of ludic interactions for target words from the spoken dialogue by viewing pre-recorded video walkthroughs of players engaged in completing each game's narrative. It should be noted that as video walkthroughs were being used as an ecological alternative rather than recording players engaged in interactions with target words, coding for these interactions were the prompt which followed the player action. The elicited programmed response from the

game would signal that an action was carried out by the player. Thus, these signaled interactions were used for data collection purposes to indicate that an action had been carried out by the player. These reactions from the game were timestamped and logged by manually denoting the type of gameplay element, its' time, and its' duration of occurrence in which said interaction was carried out. These findings will be explored in the following chapter.

Chapter 4: Results

This chapter outlines the results of the analyses and methods that were outlined in the previous chapter. The first research question, which concerned the supportive modes, was divided into two sub-questions to address visual and linguistic modes. This research question has been explored extensively through research that investigated multimedia learning (Mayer, 2021). The second research question adds to the discussion of multimodality for learning by addressing a distinct mode unique to video games. The procedural mode involves players creating meaning through engagement with the ludic elements of digital gaming. Finally, the third question compares two video game genres, Action games and interactive fiction games. Additionally, the third research question seeks to provide a media comparison between video games and television programs. A comparison with television is a valuable reference point at this time, as vocabulary research into the multimodal supports of video games is still in its early stages.

4.1 First Research Question

Visual Mode

The first research question, which explores the extent aurally provided target words are supported through the cooccurrence or the near-cooccurrence of multimodal supports, was addressed by analyzing the time of aural provision of spoken dialogue in relation to the visual presentation of images and text.

Imagery support was operationalized as the provision of a visual referent occurring within a 10-second timeframe of the target word being spoken in the dialogue of each video game. As the target words were timestamped, the occurrence of the visual referent for the

target word was recorded in relation to the spoken dialogue. The target word occurrence and the provision of the imagery are displayed in Table 4.1.

Table 4.1

Occurrence of visual support in relation to spoken target word provision

Title	Encounters with target words	Visual Encounters within 10 seconds	Percentage of visual encounters within 10 seconds
Uncharted Drake's Fortune	77	18	23.38%
The Last of Us	129	28	21.71%
The Elder Scrolls IV: Skyrim	205	8	3.9%
Ghost of Tsushima	708	117	16.53%
Assassin's Creed Origins	252	26	10.32%
Beyond Two Souls	88	16	18.18%
Firewatch	77	19	24.68%
Tales from the Borderlands	305	19	6.23%
Life is Strange	550	38	6.91%
Twelve Minutes	57	0	0%

As noted in Table 4.1, target words featured in *Firewatch* had the highest amount of visual support. Of the 77 target words encountered through the spoken dialogue, 24.68% of these encounters featured supportive imagery within a 10-second timeframe. Conversely, the game with the least amount of visual support was *Twelve Minutes*, which hosted a similar number of encounters (N= 57) with target vocabulary but none of them were supported with a visual referent within a 10-second timeframe.

A further elaboration on the degree to which imagery cooccurs with the spoken dialogue in single-player narrative video games is displayed in Table 4.2. The distribution of supportive imagery is bimodally distributed between the *Concurrent* (n=10, M=2.59, m=18) and the *No Occurrence* (n=10, M=27.1, m=185) groups. This dichotomous relationship echoes the original findings of Rodgers' (2018) investigation of supportive imagery in television

Table 4.2

The degree of imagery and spoken word occurrence

Title	Concurrent	+/- 2 Sec.	+/- 5 Sec.	No Occurrence
Uncharted: Drake's Fortune	14	2	2	59
The Last of Us	21	3	4	101
The Elder Scrolls V: Skyrim	6	0	2	197
Ghost of Tsushima	116	0	1	591
Assassin's Creed Origins	24	1	1	226
Beyond: Two Souls	10	3	3	72
Firewatch	19	0	0	58
Tales from the Borderlands	17	1	1	289
Life is Strange	30	4	4	512
Twelve Minutes	0	0	0	57

To highlight the extent to which the near-cooccurrence of a visual mode supports target vocabulary, Table 4.3 indicates how many audio-visual encounters with each target word tokens occur within a 10-second timeframe. The distribution of the visual-supported spoken target words is positively skewed (n= 69, M= 4.32) as the most common visual encounter was 1 (n= 22), and the least common value was for the target item [samurai] (n=1), which had 63 visually supported encounters. The target word [samurai] accounted for 56.26% of all the visually supported target word encounters in the game *Ghost of Tsushima.* This frequency may be due to the game's content, but more specifically, how the game displays a 3rd person point-of-view in which the player controls an avatar of a samurai. Thus, of the 63 times the target word [samurai] is spoken in the dialogue, 100% of these encounters are supported with concurrent visual support as there is always a virtual depiction of a samurai on screen.

Table 4.3

Visually supported target word tokens

Title*	Item	A/V Encounters**	Title	Item	A/V Encounters	Title	Item	A/V Encounters	Title	Item	A/V Encounters
UDF				forge	4		condenser	1		selfie	1
	symbol	3		monk	3	FW				surveillance	3
	generator	3		bandits	1		receiver	4		tornado	2
	diary	5		archer	7		helicopter	1		principal	2
	monastery	4		temple	2		scout	2		rewind	5
	statue	6		armour	32		creek	1		whale	1
LoU				samurai	63		trail	3		plank	3
	boost	5	ACO				hiking	4		padlock	2
	quarantine	1		villa	1		cave	3		crowbar	1
	ladder	8		ledger	1		hike	1		coordinate	1
	clicker	8		captain	2	TftB				gallery	2
	infected	6		symbol	4		trash	1			
TESV:S				fleet	4		bandit	4			
	forge	3		bandit	3		robot	14			
	scroll	1		pharaoh	2	LiS					
	jarl	4		temple	4		lighthouse	3			
GoT				priest	7		campus	3			
	fleet	1	BTS				bracelet	2			
	bamboo	1		pizza	4		code	1			
	stables	1		talisman	2		eclipse	1			
	aim	1		submarine	6		dorm	3			
	bandit	1		entity	3		drone	1			

Note. *audio-visual **Title coding: *Uncharted Drakes Fortune* (UDF); *Last of Us* (LoU); *The Elder Scrolls V: Skyrim* (TESV:S); *Ghost of Tsushima* (GoT); *Beyond Two Souls* (BTS); *Firewatch* (FW); *Tales from the Borderlands* (TftB); *Life is Strange* (LiS).

Written Mode

The next sub-question is regarding the way in which aurally occurring target words were supported by the accompaniment of text. This sub-question was approached through similar methods as the visual supports. That is, the on-screen text for target words were recorded as occurring within a 10-second time frame of the spoken target word. Unlike the supportive imagery, textual supports were not analyzed in subcategories of concurrent, or +/- 2 seconds, as this would be an example of Mayer's (2011) redundancy principle. Wherein, as spoken dialogue and written language are both considered texts (i.e., linguistic modes), the simultaneous presentation could overwhelm the learner (Mayer, 2011). The +/- 5 second boundary for accompanying text alongside spoken dialogue aligns with previous investigations regarding subtitles (d'Ydewalle et al., 1987). However, as mentioned in the previous chapter, the collected textual data was not the subtitles of the spoken dialogue but rather the diegetic and ludic text that appeared on screen through gameplay.

Table 4.4

Occurrence of textual support in relation to spoken target word provision

Title	Encounters with target words	Textual Encounters within +/- 5 seconds	Percentage of Textual encounters within +/- 5 seconds
Uncharted Drake's Fortune	77	12	15.58%
The Last of Us	129	15	11.62%
The Elder Scrolls V: Skyrim	205	80	39.02%
Ghost of Tsushima	708	143	20.19%
Assassin's Creed Origins	252	42	16.67%
Beyond Two Souls	88	6	6.81%
Firewatch	77	22	28.57%
Tales from the Borderlands	305	16	5.24%
Life is Strange	550	29	5.27%
Twelve Minutes	57	0	0%

The game *Elder Scrolls V: Skyrim* provided the most textual support for target item encounters (39.02%). A frequent communicative event in this game was that after a target item was said aloud by a virtual interlocutor the player could then select a dialogue option which featured the target word in a text format. This cycle of spoken and written target word occurrence could indicate the prevalence of text within this particular game. The digital game *Twelve Minutes* featured no supportive text before or after aural occurrence of a target word. The written format of these target words did appear, but beyond the temporality of the 10-second requirement of the present study.

4.2 Second Research Question

This research question is an additional extension to the initial research carried out by Rodgers (2018), in that a wider scope of multimodal supports for target vocabulary was explored. Mainly, a significant distinction for the medium of video games compared to that of television is the ability to interact with virtual elements within the narrative. This engagement with the details presented in the game is an additional means of meaning-making. This mode is referred to as the procedural mode. That is, the set of steps that involve actions and reactions that the player makes through the game's user interface, which culminates into a procedure of gameplay (Hawreliak, 2019).

The procedural mode was catalogued through the instances of player-game interactions. As discussed in the previous chapter, the procedural mode was categorized into three distinct procedures of gameplay, manual player-game interactions, environmental presence, and ludic engagement. The results of the categorization of this mode are displayed in Table 4.5. Manual interactions, which were catalogued as a number for each encounter,

consisted of actions that the player would take which concerned target items. For example, in the game *Firewatch* a player may use a receiver to communicate to a non-playable character. This instance would count as one manual interaction with the target token [receiver]. Presence was determined as a time duration, rather than a single instance. An example of this is the target item [dorm] in the game *Life is Strange*, which takes place on a school campus. The time duration in which the player spent within the virtual environment of a dormitory was recorded and the total was recorded (e.g., 39 minutes and 11 seconds). Ludic engagement, like presence, was represented as a time duration. This procedure of gameplay involved the player in engaging with target items to achieve a ludic goal. For example, the player in *Beyond Two Souls*, spent three minutes and 8 seconds piloting a representation for the target word [submarine].

Table 4.5

Target word encounters through the procedural mode

Title	Manual Interactions	Presence	Ludic Engagement
Uncharted Drake's Fortune	67	0:52:34	0:09:29
The Last of Us	194	-	4:37:20
The Elder Scrolls IV: Skyrim	35	0:10:03	0:08:46
Ghost of Tsushima	269	0:05:30	0:17:02
Assassin's Creed Origins	43	1:13:30	0:11:26
Beyond Two Souls	13	0:10:36	0:17:05
Firewatch	14	0:06:41	0:33:18
Tales from the Borderlands	13	0:00:27	0:34:11
Life is Strange	74	0:39:11	0:07:49
Twelve Minutes	33	-	-

The video game *Ghost of Tsushima* hosts the most interactions with target vocabulary (N=14, M= 19, SD= 28.7). In contrast, the games that supply the least of these interactions are *Beyond Two Souls* (N=5, M=3, SD = 1.52) and *Tales from the Borderlands* (N= 3, M= 4, SD= 4.04).

In the measure of presence through the virtual environment, *Assassin's Creed Origins* presents players with over an hour of target words represented through virtual environments (4,410 seconds, N=3, M= 1470, SD= 959). By measure of ludic interactions with target words, *The Last of Us* presented the most interactions of this type (16,640 seconds).

Table 4.6 demonstrates how each target item headword is supported through engagement with the procedural mode by measure of interactions. Each target word's part of speech is indicated as well. An additional extension of the research conducted by Rodgers (2018), which focused on concrete nouns as they would be represented visually, the present study included verbs as they could be represented visually as actions that player-characters perform in the game.

Table 4.6

Encounters of manipulation of target word tokens

Title	Target word	POS	interactions
UDF			
	symbol	n	7
	generator	n	1
	diary	n	49
	statue	n	7
	treasure	n	3
LoU			
	boost	v	5
	ladder	n	51
	infect (-ed)	n	138
TESIV: S			
	forge	v	3
	enchant	n	1
	wizard	n	2
	scroll	n	10
	armour	n	8
	claw	n	6
	jarl	n	5
GoT			
	fleet	n	1
	bamboo	n	6
	target	n	3
	sneak	v	24
	stables	n	5
	aim	v	48
	duel	v	14
	forge	n	4
	monk	n	3
	fort	n	7
	bandit (s)	n	27
	archer	n	9
	armour	n	6
	mongol	n	108
ACO			
	villa	n	2
	ledger	n	1
	captain	n	4
	symbol	n	2
	fleet	n	4
	hyena	n	5
	bandit	n	19
	pharaoh	n	2
	temple	n	4
	crocodile	n	3
BTS			
	pizza	n	1
	talisman	n	3
	submarine	n	1
	experiment	n	4
	entity	n	4
F			
	receiver	n	13
	helicopter	n	1
TftB			
	bandit	n	9
	robot	n	2
	vault	n	2
LiS			
	outfit	n	1
	lighthouse	n	3
	focus	v	5
	dorm	n	6
	drone	n	1
	selfie	n	1
	surveillance	n	2
	tornado	n	2
	rewind	v	52
	coordinate	n	1
TM			
	switch	n	13
	cop	n	8
	prove	v	4
	dessert	n	8

4.3 Third Research Question

Comparison Between Genres

One of the aims of this study was to compare multimodal supports between two video game genres. Of the 10 video games being analyzed, five are of the action genre, and the remaining belong to the interactive fiction genre. Statistical analysis was carried out to address this comparison to determine whether there is a significant difference between the multimodal supports (i.e., imagery and text) between each genre. Jamovi (OSR, 2022), an open-source statistical spreadsheet, was used to carry out this analysis.

Supportive Imagery. Table 4.7 shows skewness (1.94; 1.98) and kurtosis (3.77; 3.91) of the occurrence of imagery occurrence in relation to spoken target words. As Table 4.7 shows, the data is not normally distributed as these measures fall beyond the acceptable range of normality for skewness and kurtosis (Larson-Hall, 2016). Moreover, imagery occurrence yielded different sample sizes between the genre. Thus, a non-parametric test was chosen to examine this relationship between the two genres.

Table 4.7

Skewness and kurtosis for aural-visual occurrence between genres

Genre	**Skewness**	**Std. error skewness**	**Kurtosis**	**Std. error kurtosis**
Action	1.94	1.01	3.77	2.62
Interactive Fiction	1.98	1.01	3.91	2.62

A Chi-square test of association was used to assess the relationship between the two groups (i.e., action games and interactive fiction games) and how the visual mode (i.e., imagery of target words) occurred in relation to the spoken dialogue. One of the assumptions for this

test is that the data was nominal in nature (Larson-Hall, 2017). To address this, the variables were categorically organized into the groups*: Concurrent, +/- 2 Seconds, +/- 5 Seconds*, and *No Occurrence*. The next assumptions are that the variables are independent and mutually exclusive (Larson-Hall, 2016). These two assumptions were satisfied through the data collection and the data analysis phases of this study. The video games from which the data was collected were observed as distinct texts. Thus, imagery occurrence in one text would not affect occurrence in another. The timestamped data was organized into the closest occurrence group during data analysis. For example, an image of the target word occurring four seconds after the spoken dialogue would fall into the +/- 5 Second category rather than the +/- 2-second category to organize data into mutually exclusive groups. Thus, with the assumptions satisfied, the Chi-square test of association was conducted.

The chi-square likelihood identified that there was a significant difference between groups, X^2 (4, N=1101) = 534 (p= <.001). In addition, there was a medium effect size as determined through Cramers' V (v= 0.042, p = <.001). Thus, as the two genres being investigated are statistically different further statistical tests can be used to identify the significance of visual occurrences and the corresponding target words.

To determine the difference in visual support between each game genre, the three occurrence categories were collapsed into one variable for each of the target words. Table 4.8 displays the descriptive statistics for the visual support in relation to spoken provision of target words within the two video game categories, action and interactive fiction games.

Table 4.8

Descriptive statistics for audio-visual concurrence in two video game genres

Genre	N	Mean	SD	SE
Action	25	6.68	13.27	2.65
Interactive Fiction	30	2.77	2.53	0.46

Due to the non-parametric nature of the data, the Mann-Whitney U test was used to compare the two genres. The results of this statistical test are indicated in Table 4.9. The results indicate that there is not a statistically significant difference between the occurrence of visual support in relation to aural occurrence of target words between video game genres (U= 430, p = 0.276) with a very small effect size (r_{rb} = 0.168) using the second-language research interpretation guidelines suggested in Plonksy and Oswald (2014).

Table 4.9

Mann-Whitney U test for audio-visual concurrence in video game genres

	U	p	Effect size
Imagery Support	430	0.276	0.168

Note. The rank biserial correlation used for effect size in the Mann-Whitney test

Supportive text. The next comparison to be investigated was to compare the audio-textual concurrence between the two video game categories. The descriptive statistics for the textual support for target words is displayed in Table 4.10. To analyze this non-parametric data a Mann-Whitney U test was conducted, which is shown in Table 4.11.

Table 4.10

Descriptive Statistics for audio-textual concurrence in video game genres

Genre	N	Mean	SD	Skewness	Kurtosis
Action	35	8.37	7.78	2.05	5.44
Interactive Fiction	22	3.45	2.76	1.20	0.52

The Mann-Whitney U test indicated that there is a significantly significant difference between groups (p= 0.003; see Table 4.11). Text provision is more favourable in the action game genre (N = 35, M = 8.37, SD = 7.78) than in the interactive fiction genre (N = 22, M = 3.45, SD = 2.76). This test yielded a small effect size in the field-specific benchmark reporting (r_{rb} = 0.46; Plonsky & Oswald, 2014)

Table 4.11

Mann-Whitney U test for audio-textual concurrence in video game genres

	U	p	Effect size
Textual Support	208	0.003	0.46

Note. The rank biserial correlation used for effect size in the Mann-Whitney test

Supportive Procedures. The last mode compared between video game genres was the procedural mode, specifically the manual interactions that players made with target words. The descriptive statistics are shown in Table 4.12.

Table 4.12

Descriptive Statistics for procedural manipulations with target words

Genre	N	Mean	SD	Skewness	Kurtosis
Action	37	16.59	29.1	3.11	10.1
Interactive Fiction	21	5.81	11.1	4.01	17.1

A Mann-Whitney U test was used as a non-parametric alternative to the independent T test. This analysis, represented in Table 4.13, indicated that there was a significant difference between groups (U= 213, p = 0.004). The Mann-Whitney U test indicated a small effect size (r_{rb} = 0.45). Similar to text-based supports, action games featured more manual interactions with

target words (N = 37, M = 16.59, SD = 29.1) than the interactive fiction games (N = 21, M = 5.81, SD = 11.1).

Table 4.13

Mann-Whitney U test for procedural manipulations with target words

	U	p	Effect size
Procedural Support	213	0.004	0.45

Note. The rank biserial correlation used for effect size in the Mann-Whitney test

Comparison Between Media

The third research question also sought to investigate how the medium of video games compared to that of television by measure of visual support for vocabulary items. To conduct this analysis, the visual supports for vocabulary in video games were compared with the data from Rodgers (2018) which investigated the concurrence of imagery and spoken target words in television. The television data was collected to represent a sample of the documentary genre and the narrative fiction genre. To align with the results of Rodgers (2018), only words occurring on the BNC/COCA (Nation, 2017) 4,000-level word lists were included in the comparison analysis. As table 4.10 indicates, the data is non-parametric in nature due to the skewness (5.35; 2.42) and kurtosis (31.5; 7.93).

Table 4.14

Descriptive statistics supportive imagery between video games and television.

Media	N	Mean	SD	Skewness	Kurtosis
Video Games	55	4.55	9.25	5.35	31.5
Television	88	6.74	6.70	2.42	7.93

As the data was not normally distributed, the non-parametric Mann-Whitney U test was used to compare the concurrence of spoken-visual target word provision. The results of the test

are shown in Table 4.11. There is a statistically significant difference between the groups (U= 1479, p = <.001) with a low effect size (r_{rb} = 0.38). The presence of imagery is more prevalent in the sample of television programs (M = 6.74) than in the sample of video games (M = 4.55).

Table 4.15

Mann-Whitney U test for supportive imagery between video games and television

	U	p	Effect Size
Supportive Imagery	1479	<.001	0.389

Note. The rank biserial correlation used for effect size in the Mann-Whitney test

As indicated in Rodgers (2018), the documentary genre provided more supportive imagery for target words than in the narrative-fiction genre. The educational design of documentary films could be more congruent with Mayer's (2007) multimedia learning design principles, thus providing more supportive visuals. As the video games were designed for entertainment purposes, rather than educational, they may provide supportive imagery in line with that of the narrative fiction genre which is also intended for entertainment. Table 4.12 outlines the descriptive statistics for these two groups.

Table 4.16

Descriptive statistics supportive imagery between narrative fiction and video games

Media	N	Mean	SD	Skewness	Kurtosis
Narrative-fiction TV	22	2.23	2.41	1.35	1.87
Video games	55	4.55	9.25	5.35	31.5

The Mann-Whitney U test was used to compare the audio-visual concurrence between the two groups. The results of the analysis are outlined in Table 4.17. There is a small effect size as indicated by the rank biserial correlation (r_{rb} = 0.235). The results of the Mann-Whitney U

test indicate that the groups are not statistically significant difference between the narrative fiction program and the sample of COTS video games.

Table 4.17

Mann-Whitney U test for supportive imagery between video games and television

	U	p	Effect Size
Supportive Imagery	454	0.083	0.235

Note. The rank biserial correlation used for effect size in the Mann-Whitney test

Chapter 5: Discussion

This chapter will outline how target words spoken in the dialogue of commercial-off-the-shelf video games were supported through consonance of aligning communicative modes. These findings were uncovered through a multimodal linguistic-visual-procedural analysis of supportive semiotic resources encountered through playing each video game's narrative. The findings are discussed below, as well as the limitations of the present study and how directions of future research may overcome these.

5.1 Findings

What is the occurrence of visual and textual supports for the aural input of target words in commercial-off-the-shelf video games within a 10-second period?

Of the video games included in the study there was a varied range of concurrent visual support for spoken target words. This section will outline the various findings of supportive semiotic modes on target words, discuss video games which provided a bimodal distribution of low and high percentages of supportive encounters, and provide possible insights into why encounters with target words were either multimodally supported or unsupported.

Visual Support. The ten games selected in this study supplied varied encounters of visual support for the identified target words provided in the spoken dialogue. The games, The *Last of Us* (LoU) and Firewatch, hosted the most multimodal supports for target items within this sample of commercial-off-the-shelf games. Target word encounters in the dialogue of LoU were supported mainly by the visual mode (30.23%) but featured textual support as well (11.62%). Conversely, *Firewatch* was supported more by the textual mode (28.57%) and marginally less by the visual mode (24.68%). A possible reason for this may be the setting and

genre of both of these titles. LoU takes place in a post-apocalyptic version of modern-day United States, and *Firewatch* has the players embody a park ranger whose work involves surveying the Shoshone National Forest. A fundamental commonality between these two settings is that they are fictional narratives that take place in a virtual version of the real world. Thus, much of the imagery that players can see and interact with is somewhat reflective of reality. Being rooted in a real-world setting, much of the vocabulary present in these games are likely to appear in the BNC/COCA word lists. Using the BNC/COCA wordlists as a reference corpus for the lexical analysis would favour language from a real-world context for target word identification and selection. As well, both of these games possess elements of the survival sub-genre. As the name may indicate, survival games place players into survival situations they must overcome (Reinhardt, 2019). In games of this nature, mundane objects often become valuable resources that the player can use as tools or crafting components, leading to learning the vocabulary of the everyday objects through using them in meaningful scenarios (Franciosi et al., 2016).

Textual Support. The game with the least visual support was *Twelve Minutes* (TM), in which visual supports were supplied for 2.81% of all target word encounters in this game. Two games that scored similarly in visual support measures are *The Elder Scrolls IV: Skyrim* (TESVS; 3.90%) and *Assassin's Creed Origins* (ACO; 4.52%). A possible reason for the low visual support for target items in these three games could be the process in which the target items were selected. Both TESVS and TM had a marked increase in text-based concurrent support for spoken target words; 39.02% and 19.10%, respectively. These percentages could point to both of these games being more text-based. Thus, since a limitation of this study was that it is

concerned with how target words from only the spoken dialogue were supported, there could be many more potential target words from the in-game text that were unmeasured because they did not appear in the spoken dialogue. The game ACO saw a minor increase (7.09%) between the, which could be attributed to a similar conclusion. *Life is Strange* (LiS), was a game that provided many encounters with target vocabulary in the spoken dialogue (n = 550). However, not many of these encounters were supported with visual (6.91%) or textual (5.27%) modes. The last game, *Tales from the Borderlands* (TftB), did not have much visual (6.23%) or textual (5.25%) support.

What are the opportunities for target words to be interacted with in a virtual environment presented in commercial-off-the-shelf video games?

The procedural mode presents an opportunity for the potential of learning vocabulary through video games. At the time of this paper, research has not attributed the procedural elements of gameplay as a semiotic mode for the purposes of language learning.

Manipulation. Interactions requiring a manual action from the player through a controller or keyboard, afforded conscious engagement with target vocabulary is distinct to the medium of digital games. The rationale for investigating instances of this interaction between the player and the game is that engagement with target vocabulary could aid the learner in noticing the target items. If the claim that “attention appears necessary for understanding nearly every aspect of second and foreign language learning” (Schmidt, 2001, p. 6), it may stand to reason that attending to multimodal representations of vocabulary items, albeit ludically, could benefit lexical development. The item [infect(ed)] was the most interacted with the target word (N = 138). This is due to the content of the game, LoU, in which infected zombies

are the antagonists of the plot. Thus, a player spends a considerable amount of game time engaging with these elements. The second most procedurally interacted with target word is [rewind] in LiS. In this video game, the protagonist possesses the supernatural ability to rewind time. This ability is a core mechanic of gameplay, and as a result, the player performs this action several times throughout the narrative (n = 52). The content of a game and the gameplay procedures may make specific vocabulary more salient for a learner with the appropriate lexical coverage. While the procedural mode could be an affordance for noticing, it still requires the support of a linguistic mode for language learning to occur. The language used in the games of this study is mainly communicated in the spoken dialogue between in-game interlocutors. As well, there are varying amounts of text provided between games. These words are further supported through text, as the game provides text-based instructions to aid the player in developing a gaming literacy within their specific games. The target word [infect(ed)] and [rewind] appears in the dialogue of their respective games 32 times and textually eight times as ludic instructions. This alignment of procedural and linguistic modes makes the medium of video games a potentially valuable source of vocabulary development. Future investigations could include more of an emphasis on the written mode in conducting a corpus analysis of games as the present study was mainly concerned with spoken-dialogue and its supportive modes.

Presence. Exploring the spatial-temporal virtual environment was an additional procedural mode that was observed. Two games are notable in this, *Uncharted: Drake's Fortune* (UDF) and ACO. Target words that were represented in this mode are visual, but a distinctive feature is that players were able to procedurally explore the locale as a feature of

the game. In UDF, the player spent nearly an hour (52 minutes and 34 seconds) exploring a digital representation of the target word [monastery]. Similarly, the player in ACO spent considerable time (one hour, 13 minutes and 30 seconds) in representations of the target words [temple], [villa], and [pyramid]. This may be a sizable amount of time exposed to visuals of these spaces, but it is still possible that a learner may not make the form-meaning connection between the linguistic and the visual mode because attention is not required in these encounters with target words. Moreover, the temporality between target word provision and the visual stimulus may be too vast.

Ludic Engagement. The last category of the procedural mode observed was ludic engagement with target items. This category included game-specific interactions with vocabulary, such as solving a puzzle or carrying out a goal. The data collected in this category is a conservative estimate, as the level of attention given toward these tasks is unmeasured in the present study. For example, the target item [claw], in which an individual tasks the player-character to retrieve a golden claw that was stolen, it is unsure if the player was resolute in directly finding the stolen item, or if their attention was toward the series of obstacles that they needed to achieve. Puzzle-solving, in which target vocabulary played a key role, could be a valuable opportunity for language learning. In the game UDF, players are presented with a puzzle where they are required to manipulate statues in the proper sequence to open a secret passage. The players possess a riddle to aid them in choosing the correct sequence of statute movements. While the player carries out actions to solve the puzzle, in-game interlocutors provide diegetic commentary about the statues or the riddle that may aid the player without explicit direction. Under these circumstances, a language learner may learn the target item

[statue] through playing the game. In this encounter, the target item is represented visually as a digital representation of various statues, the characters discuss that perhaps something needs to be done about the statues to solve the puzzle, the word [statue] is written in the riddle, and the player makes various manipulations of the statues in order to open the secret passage to the treasure room. The described encounter took the character nine minutes and 29 seconds to solve, and in other games, many of these puzzle encounters did not exceed 30 minutes. However, it is time spent engaging with low-frequency vocabulary that can be learned through extramural enjoyment of video games.

How does concurrent or near-concurrent visual and aural input compare between two video game genres and television programs?

Between Genres. COTS video games, despite being intended for entertainment purposes, provide inherent educational supports for vocabulary items. The cooccurrence of these semiotic modes are varied across video game genre. Thus, video game genre does not seem to be a determining factor in providing supportive L2 learning environments.

The provision of concurrent audio-visual modes is similar between the two genres within this sample of digital games. The action and interactive fiction genres were not statistically different from one another (see tables 4.8 and 4.9). This could be a result of video games being a visual medium and thus provide associated imagery of narrative elements within both genres. Future investigations of the visual mode in video games could investigate other genres, such as educational games. This could lead to the distinction found in Rodgers (2018) where the educational documentary series provided more supportive imagery than the narrative fiction television programme. However, the results from Dixon and collogues (2022)

meta-analysis on the efficacy of vocabulary learning in video games indicate that educational games to not yield the same learning potential as their entertainment-based counterparts. Until future investigations are conducted, genre does not seem to be a discernable factor for supportive visuals like it is in television.

The statistical tests indicated that genre does, however, play a role in providing encounters with the supportive textual and procedural modes (See tables 4.10 – 4.13). The action game genre provided more text-based support than the interactive fiction genre. A rationale for this could be the player-character's inventories, in-game books and notes, and the in-game hints and tooltips that help the player engage in play. These features were notably absent in the sample of interactive fiction games which could have led to the higher score of textual supports in action games. In the measure of ludic interactions with target vocabulary, the action games presented more of these encounters than the interactive fiction games. This could allude to action games having a higher learning potential, as the interaction with target word referents in the digital space could make these words more salient. Moreover, these items, as they relate to gameplay, could be more frequent in paratexts regarding the specific games, such as forums, tutorials, or strategy guides.

Between Media. Audio-visual concurrence is provided similarly between narrative fiction television programmes and this sample of video games (see tables 4.16 - 4.17). The supportive imagery from Rodgers (2018) was compared with the supportive imagery present within this sample of video games, which yielded similar results. This promising finding could indicate that the learning potential of watching television could apply to playing video games. Television provides ample learning opportunities to second language users (Durban, Rodgers, &

Peters 2020), the present study provides some empirical evidence for similar learning opportunities through supportive imagery present in this audio-visual medium. However, video games to not provide as much visual support for target items as documentary television. This may be a result of documentaries being informed by educational design by means of audio-visual concurrence. While the present study did not include educational games, future analyses could find similar learning potential in these as in documentary television.

An important finding was that entertainment-based video games present visually supported vocabulary words similarly to narrative television. This could lead to the conclusion, that similar to television, that video games are a suitable source of extramural linguistic input. Extensive viewing programmes have been well documented to provide language learners with positive vocabulary learning (Webb, 2015; Rodgers, 2016). Vocabulary learning has been attributed to the audio-visual contiguity present in television programmes (Webb, 2015; Rodgers, 2016; Ghebghoub, 2021). The similar rate presentation of supportive imagery between narrative television and video games could indicate that this medium could be equally as successful in class-based implementations of extensive input programs.

5.2 Limitations and Future Research into vocabulary learning through video games

This section will outline various limitations of the present study while offering insights to future research endeavours toward vocabulary learning opportunities in videogames. The major limitation of the present study was that it only focused on the spoken dialogue as a unit of analysis, while the textual and procedural modes played an auxiliary role. This resulted in only observing the target words that were spoken in the dialogue. An example of this, is in the game *Twelve Minutes* which scored particularly low in comparison of visual, text, and

procedural modes. This was a result of the data collection methods, as this game provided many labels for the visual elements within the character's virtual apartment but thse referents were not included in the analysis because they were not mentioned in the spoken dialogue of the game. Thus, the most significant obstacle to overcome in the design of this research is the availability of sophisticated data collection methods for multimodal analysis. To date, tools such as object and text recognition software which could be used for multimodal video analysis have been developed; however, these remain economically unfeasible. Thus, the present study manually collected the data for the visually supportive modes (i.e., text, imagery, and procedurality). While limited, the present study points to potential learning opportunities for target words spoken in the narratives of several commercial-off-the-shelf video games designed for entertainment purposes.

Similar to the multimedia principles suggested by Mayer (2001) to reduce extraneous cognitive processing, Lemke (1998) encourages agreement between modes in a multimodal ensemble. This occurs when two or more modes work in tandem to present a holistic semiotic message. While the methodology of the present study was primarily informed by the cognitive theory of multimedia learning (Mayer, 2001), it is worth noting that research into multimodality, multimodal learning specifically (Jewitt et al., 2016), point to similar conclusions (Lauer, 2009). To this end, the present study has identified encounters of target words supported both through multiple media (i.e., imagery, text, and play) or multiple modalities (i.e., visual mode, linguistic mode, and procedural mode). As the previous chapters indicate, I have opted for the nomenclature of multimodality, as the term multimedia has been saturated with an emphasis on products developed for the public sphere (Lauer, 2009). Regardless of the

nomenclature employed, the alignment of concurrent and related semiotic messages has led to vocabulary learning opportunities (Rodgers, 2018; Peters, 2019; Ghebghoub, 2021; Peters & Webb, 2018). Thus, this study seeks to encourage future research to employ a multimodal perspective when assessing the efficacy of how these modes align to contribute to vocabulary learning.

Another important limitation that future investigations could address, is whether the visual and textual modes were noticed by the learner. The data collection methods made no distinction between whether a visual referent for a target word or associated text was in the foreground or background of the player's field of view on screen. It is likely that less-salient imagery or text in the background, unless required for a specific gameplay procedure, would go unattended by a learner. The implementation of eye-tracking tools to observe language learners engaged in play, could identify if these visuals or written text are noticed by the learner. Analyzing a language learner's gaze, could indicate whether aural provision of unknown words prompt them to actively seek for possible supportive imagery (Rodgers, 2018). In the gaming context, players can dictate the on-screen point-of-view, which could control the amount of exposure a player has to the supportive modes identified in this paper. The video walkthroughs in this study were recorded by experienced players not language learners. It is possible that a language learner would interact with the digital elements of the game differently to that of a native speaker of the game's language, notably regarding interactions with digital referents for unknown vocabulary. It would behoove future studies to record second-language players to analyze their interaction with the supportive modes in authentic play.

The sample of video games used in this study were chosen to reflect the conditions which demonstrated successful vocabulary learning in the meta-analyses on DGBLL (Dixon, et al. 2022). However, it is worth noting, that these video games represent one example of the various digital gaming formats, types, and genre that language learners could play in their target language. Most notably, none of these games provide opportunities for linguistic output, which could be achieved through multiplayer games (Reinhardt, 2019). Future investigations could study how the supportive modes identified in this paper are used between L2 interlocutors which could lead to valuable examples of negotiation of meaning (Long, 1996) toward the comprehension of the various elements in the digital game space.

Visual Mode

Future research endeavours concerning vocabulary acquisition would do well to compile a visual corpus of the on-screen elements presented to the individual who plays the game, such as the virtual objects within the game world. However, developing such a corpus for the purposes of vocabulary learning would likely be a lengthy undertaking. Due to the exploratory nature of the present study, I opted to investigate how one aspect (i.e., spoken target words) was supported across myriad titles and genres. Thus, the present study catalogued imagery in relation to the spoken target words rather than comparing the spoken-word and visual corpora for congruency. Advancements in the theory and practice of multimodal communication could benefit further investigations of language learning through video games (Gee, 2007). Developing powerful software tools has enabled researchers to collect and analyze multi-million-word corpora (Webb & Nation, 2017). Similar computational

advancements could be used to construct visual corpora (Carter & Adolphs, 2007). Using sophisticated tools like application programming interfaces, objects appearing on-screen can be detected and categorized. However, these applications show great promise for the construction of visual corpora, but during the present study's design, these tools were unavailable.

Textual Mode

Similar to how visual data was collected, the textual data was only collected in relation to target word occurrence within the spoken dialogue in each game. Future endeavours should account for the entirety of the text presented that could be presented on screen. While related cooccurring text and auditory input are beneficial in reducing the cognitive load of multimedia messages (Mayer, 2021), it only accounts for a portion of the written-linguistic mode that is available to the learner. In this study, some words occurred in high textual frequencies but infrequently appeared in the spoken dialogue. For example, the potentially learnable word [potion] was not included in the target words due to the inclusion criteria of five-or-more encounters (Frequency of occurrence = 3). However, the written word [potion] appeared 53 times on screen and each of these encounters featured a visual referent of the word. It is likely that learners under these conditions would learn this word but was excluded from this analysis due to the study's focus on support for spoken words. Moreover, the target word [armour], while meeting the inclusion criteria (Frequency of occurrence = 10), was represented textually more frequently with 100% visual-referent concurrence (Frequency of occurrence = 38). Again, as this word was supported through multiple means, these supports are likely to benefit in the learning of this vocabulary item.

Since textual data in this study was collected manually, it would be difficult to systematically log all text that appears on screen. Previous research has succeeded in extracting the text from the embedded files of the games system (Heidt, 2020). As this process would require knowledge in coding to extract the text, it may be more economical to record the text that is available to the learner through its occurrence on screen. Using advancements in automated text-recognition, future investigations could collect vast amounts of simultaneous textual input, similar to how the speech-to-text software was used in this study to collect the spoken dialogue. Like the above suggestions towards the construction of visual corpora, cloud-based application programming interfaces could extract text that occurs on screen to create a textual corpus of player-encountered text.

Procedural Mode

Procedurality, the player's actions and reactions through interaction with the virtual environment, is a particularly under-researched semiotic mode (Hawreliak, 2019). This engagement with the games' elements is an exciting depart from other audio-visual media. The present study used video walkthroughs that demonstrate the game's narrative from the beginning to the story's conclusion. While these video walkthroughs were chosen as an economical alternative to recording a participant playing through a game's narrative, the walkthroughs did not capture the agency that players have to complete their objective. Agency is seen as a significant distinction that sets apart video games from other audio-visual media (Gee, 2007; Mayer, 2014; Reinhardt, 2019). For example, in Ghost of Tsushima, the video walkthrough recorded the player receiving a quest, travelling to the designated area, and completing the task as quickly as possible. However, in authentic play, the player has the

freedom to explore the entire island of Tsushima so that the narrative may have the player 1) receive a quest, choose to: a) travel to the designated area, or b) explore nearby villages, or c) collect resources to craft better equipment, or d) complete a side quest, or so on. This procedure is designed to "make the player feel as though they are a part of an expansive, living game world" (Hawreliak, 2019). In the learning context, learners are afforded the opportunity to engage with the game's elements that learners are motivated to explore (Jewitt, 2013). If a component of the game's narrative is unknown or interesting to the learner, autonomous learners may shift their attention to investigate (Jewitt, 2013).

The gameplay procedures, such as interacting with the game's elements or manipulating the player's avatar to perform actions, are a semiotic resource that learners can encode and decode to make sense of the virtual space (Hawreliak, 2019). While not linguistic in nature, encounters with these words could potentially lead to vocabulary learning opportunities. In contrast to television, while concrete nouns in audio-visual media would likely be the most tangible target words (Rodgers, 2018), the interaction in video games could make an ideal learning environment for actions that are otherwise inaccessible to a classroom situation. However, based on the target word selection process, verbs did not appear to be a frequently discussed topic. For example, in the game Ghost of Tsushima about warring samurai, the player's character performs a series of sword fighting maneuvers. The action [parry] (FoO = 4) was performed 92 times and the [dodge] (FoO = 4) action was carried out 56 times. While both of these words were excluded for not meeting the inclusion criteria, the player still spent considerable time using these actions. In the initial design of this project, verbs were considered for target words. While not appearing much in the spoken dialogue of the narrative,

these ludic words are a core component of the game and thus may be used extensively in paratexts of the discourse around the game such as, tutorials, forums, or instruction manuals.

An interesting observation that was found during data collection phase, was how target words were used in the procedural mode to reflect task-like activities. Various identified target words were represented visually or textually and as virtual elements that needed to be engaged with to progress in the games' narrative. An example of this is for the target word [ladder] in the game The *Last of Us*. Protagonists explore post-apocalyptic environments such as ruined cities or towns in this game. Protagonists would often encounter obstacles, such as blocked passages or destroyed stairs, that can be overcome with a ladder. The game does not explicitly instruct the player to find a ladder in these situations. Instead, non-playable interlocutors will dietetically make suggestions for how to overcome the obstacle. This puzzle leaves the player-character to investigate the environment for possible solutions in order to progress. Should the player-character not find the solution, the cooperative interlocutor will suggest that the protagonist needs to find a ladder to reach the higher elevation. This exchange of in-game instructions between the player-character and the virtual companion could be a catalyst for interlanguage development if the target vocabulary is unknown to the learner. Episodes of this nature, through communication breakdown, may lead learners to identify gaps in their linguistic repertoire and make target vocabulary more salient (Long, 1996). Critics of tasks that push learners to notice gaps in their interlanguage, argue that such tasks may have a demotivational effect on the learner (Aston, 1986, as cited in Skehan & Foster, 2001). However, inherent motivational factors may outweigh the demotivational effects, such as wanting to progress through the unfolding narrative (Reinhardt, 2019). As well, through embodying a

digital avatar, language learners may subconsciously acquiesce to the identity of the protagonist rather than their status as a learner (Yee & Bailenson, 2007). Video games are designed to present a series of challenges to the player (Reinhardt, 2019). Thus in-game instructions are a valuable resource for developing gaming literacy in the specific game (Jewitt, 2013; Mayer, 2014). An exciting direction for future research would be investigating the extent of vocabulary learning opportunities that could arise from these diegetic instructional encounters in video games. This could be addressed through an intervention by first assessing the lexical coverage that language learners would have of a specific game to identify unknown vocabulary, have participants with a 90%-95% lexical coverage engage in the gameplay of these task-like encounters, and then conduct a post-test to determine if target items have been learned.

5.2 Pedagogical Considerations

There is an unfortunate trend of technological determination in DGBLL that research into the efficacy of video games as self-guided language learning tools has seemed to exclude teachers from the learning process (York et al., 2021). In their review of DGBLL literature, deHaan (2020, as cited by York et al., 2021) indicated that less than a quarter of research papers neglected to mention "interactions with students during or after gameplay, teaching materials, or practical advice" (pp. 1168-1169). Research into DGBLL has largely emphasized game design aspects that are studied in the lab-based contexts over how games can be used in the classroom and beyond to facilitate language learning (York et al., 2021). This trend may be a by-product of research being conducted by language-teaching researchers instead of language teachers themselves (York et al., 2021). To this end, while the present study claims the

vocabulary learning potential of digital games by identifying instances of pedagogical principles embedded into the design of commercial-off-the-shelf video games, these games should by no means serve as a substitute for learning in the classroom. Instead of looking at what games do, it would behoove researchers to look at what teachers can do with games prior to, during, and after gameplay. Research, especially in reading endeavours, have shown that language teacher interventions led to greater vocabulary learning (Teng, 2015). The results indicate that video games do possess qualities that could lead to learning outside of the classroom. However, the following subsections will provide examples for in-class pedagogical interventions in order to capitalize on the vocabulary learning potential of video games as part of an extensive playing programme.

Before gameplay

Prior to gameplay, language teachers facilitate learning prior to the lesson, which could increase the learning potential of games (York et al., 2019). Understanding the plot, context, gameplay elements of specific games can inform how a game can be incorporated into the classroom. This, in turn, can inform the design of materials that can supplement the potential learning from a game. An example of poor choices and design for the implementation of DGBLL comes from a study conducted by McCosker (2015). This study used the educational game *Furious Frogs* in the second-language classroom. McCosker, while intrigued by the potential for using games in the classroom, criticized this game as a language learning tool. The core mechanic of this particular game is to competitively select antonyms before your opponent can (McCosker, 2015). McCosker argued that the linguistic feature should intentionally be taught to the learner as it is a crucial gameplay component. In this example, the role of the teacher is

excluded. In choosing to use the game *Furious Frogs* in the classroom, a teacher should familiarize themselves with the game in order to make use of its pedagogical potential for language learning. Prior to using this in the classroom, a teacher could identify that antonyms are a key concept to take part in the game and thus should consider if their learners need to know this linguistic feature. Therefore, this process can inform the materials that could supplement the game before or after gameplay. While *Furious Frogs* may not impart on learners the linguistic knowledge necessary to play, teachers can design lessons that incorporate the concepts the game reinforces. Further, the game could be a suitable tool for fluency development, as per Nation's Four Strands (2007) approach, due to time pressures which could be a by-product of the competitive nature of the game design. Informed by language pedagogy, a teacher can consider the strengths and weaknesses of games and how they fit into the classroom (York et al., 2021).

The target vocabulary and their supportive modes identified for this study could benefit language teachers in incorporating these games into their curriculum. The identified target words could serve as a basis for a word list to set long-range vocabulary learning goals across a language course (Nation, 2016). While five encounters seem to be sufficient for partial learning through an audio-visual medium (Rodgers and Webb 2009; Peters, 2018), pedagogical intervention could be used to develop deeper lexical knowledge. Word knowledge of the target words can be strengthened by repeated temporal encounters of the different supportive modes for the representations of vocabulary items. To this end, as this paper identified various supportive modes present in COTS games, language teachers could conduct similar analyses of games they intend to bring into the classroom to determine how these games can strengthen

the intended vocabulary goals of a course. The various modes identified in this paper, such as imagery, could be used to create materials and activities to elicit learners' prior knowledge before engaging in gameplay. Pedagogical decisions before gameplay could further benefit the learning that can occur in video games.

In the enactment of an extensive play program, teachers could develop pre-task activities that involve learners in choosing the games that they intend to play (York et al., 2019). Grouped activities could have learners research the game by reading instruction manuals, reading online forums, or watching videos such as trailers or game walkthroughs (York et al., 2019). Teacher interventions can prime learners to identify useful vocabulary in these external media that may aid them in playing the games prior to engaging in gameplay (York et al., 2019). An example of how teachers can elicit engagement with various paratexts around the game is identified in Appendix A which has been adopted from the kotoba rollers framework which used board games in the L2 classroom (York et al., 2019)

During Gameplay

The decision to use games in a curriculum to supplement the learning in the classroom may require the instructor to take on a supportive role rather than a central one. Reflective of that which occurs in the task-based language class, the teacher can organize, manage, facilitate, motivate, and advise (Willis & Willis, 2007). In this way, teachers can serve as interactive guides that tailor learning to the learner better than solely through a video game (York et al., 2021). To this end, the teacher can draw the students' attention to both game-based and linguistic elements to promote learning. The games within this study are one-way linguistic input, there is no language output required by player. Teachers can therefor encourage language output by

engaging students in ludic discussions during gameplay, such as reacting or encouraging actions within the game to promote discussion. These curated episodes could elicit meaningful discussion that can be the basis of corrective feedback instances or make specific language features more salient. Teachers observing play while monitoring students can provide these instances of language-focused discussion that is tailored to their individual interlanguage development, rather than students passively intaking the material.

With the information gained from this study, language teachers could further benefit from the supportive modes that occur naturally in the game. In curating word lists and multimodal corpora of a video game a teacher intends to include in a language course, teachers can be primed to point out these and elicit engagement with the various modes. Visually, a student may fail to notice imagery of a specific target word. To which, a teacher can pose questions regarding the imagery to make the feature more salient or have the student elaborate on the item (e.g., "what was that over there?", or "what do you think this does?"). Teachers can have students read text that they encounter aloud to help them develop the form-meaning connection and develop phonetic competency for specific items. The diegetic texts within this study were largely optional for the player. A language teacher could use these in-game texts to generate activities that connect the diegetic texts to the game's broader narrative. The procedural mode could elicit meaningful interactions between a student and their teacher through discussing strategies, providing ludic feedback, and encouraging play. When observing their students in gameplay, teachers in a mediator role can produce discussion through developing literacy of a game's rules and actions (Poole, 2020). The above interactions

could occur between a teacher and their student and among classmates watching their peers engaged in gameplay.

After Gameplay

Video games could be used as tools to extend language learning beyond the classroom (York et al., 2021; Reinhardt, 2019). However, while much of the input of the game occurs outside of the classroom, classroom discussions around gameplay could increase learner output by engaging in meaningful discussions (York et al., 2021). This could be done by discussing or debriefing the events that occurred during play, planning future strategies, or specifying what will happen next in the narrative. Teachers can bring learners' attention to linguistic features that were supplied from in and outside the game and provide corrective feedback for utterances observed when monitoring play (York et al., 2021).

While this study identified aspects of a games' design that supported learning through multiple modes, language teachers can design material that capitalizes on the learning that can occur through gaming in an L2. Focusing on language pedagogy rather than on specific game features, teachers can develop approaches that extend across a myriad of games. To this end, Appendix B serves as an example of how a language teacher could devise post-gaming tasks that strengthen the learning that can occur from games as part of an extensive gaming program. The model is designed from the Literature Circle framework (Daniels, 2002), which has been adapted for use in English-language teaching curricula (Shelton-Strong, 2012). This classroom practice involves learners being assigned into various roles in which each student performs as an assignment (Daniels, 2002). Appendix B is an example of how this practice could be adapted to the extensive input of video games.

Chapter 6: Conclusion

The present study sought to provide an empirical basis for the findings of multiple meta-analyses regarding the efficacy of video games in second language learning. Research has highlighted the positive effect of video games on vocabulary learning specifically (Dixon et al., 2022; Tsai & Tsai, 2018; Chen et al., 2018). Through conducting a multimodal analysis of various supportive modes, I have identified that imagery, text, and procedures occur in relation to the aural presentation of low-frequency vocabulary. The findings of imagery congruency is in line with that of television. The similarity of the supportive multimodal contiguity for second-language vocabulary learning between video games and narrative television is promising. While there are additional modes present in video games that are not in television, the similar provision of visual support could perhaps allow future investigations into video games to be informed by the wealth of research conducted toward television.

The next step of this research could involve more sophisticated multimodal analysis tools, such as object and text recognition software to explore how target words are represented in different semiotic modes. These tools could allow for more games to be analyzed, thus establishing more definitive results as to the efficacy of playing video games as a second-language learner. Moreover, this research could inform future analyses which could include how these target words are learned by language learners, by the inclusion of eye-tracking tools to determine if these supportive modes are salient to the learner.

Finally, the results suggest that myriad modes, particularly that of imagery, presented in commercial-off-the-shelf video games provide a supportive vocabulary learning environment. As extensive viewing programs has been beneficial in providing language learners with sensory

rich and authentic target language input, the similar audio-visual contiguity found in video games could inform similar implementations in L2 and EFL classrooms.

References

Altarriba, J., & Knickerbocker, H. (2011). Acquiring second language vocabulary through the use of images and words. In Trofimovich, P. and McDonough, K. (eds.), *Applying Priming Methods to L2 Learning, Teaching and Research*, 21–47. John Benjamin.

Anthony, L. (2021). AntWordProfiler (Version 1.5.1) [Computer Software]. *Waseda University*. Available from https://www.laurenceanthony.net/software

Baddeley, A. (1999). *Essentials of human memory*. Psychology Press.

Bregni, S. (2017). The Italian digital classroom: Italian culture and literature through digital tools and social media. NEMLA Italian Studies, XXXIX(Special issue), 42–71.

Calleja, G. (2011). *In-game from immersion to incorporation*. MIT Press.

Carter, R., & Adolphs, S. (n.d.). *Linking the verbal and visual: new directions for corpus linguistics*.

Chen, M.H., Tseng, W.-T., & Hsiao, T.-Y. (2018). The effectiveness of digital game-based vocabulary learning: A framework-based view of meta-analysis. *British Journal of Educational Technology*, *49*(1), 69–77. https://doi.org/10.1111/bjet.12526

Chomsky, N. (1965). *Aspects of the theory of syntax* . M.I.T. Press

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149–210. https://doi.org/10.1007/BF01320076

Cook, & Mayer, R. E. (1988). Teaching Readers About the Structure of Scientific Text. *Journal of Educational Psychology*, *80*(4), 448–456. https://doi.org/10.1037/0022-0663.80.4.448

Cornillie, F. (2022). Digital Games and Technology-Mediated Gameful Environments for L2 Learning and Instruction. In *The Routledge Handbook of Second Language Acquisition and Technology*, 272–285. Routledge. https://doi.org/10.4324/9781351117586-24

Cowan. (2010). The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science: a Journal of the American Psychological Society*, *19* (1), 51–57. https://doi.org/10.1177/0963721409359277

d'Ydewalle, G., Praet, C., Verfaillie, K., & Rensbergen, J. V. (1991). Watching Subtitled Television: Automatic Reading Behavior. *Communication Research*, *18*(5), 650–666. https://doi.org/10.1177/009365091018005005

Daniels, H. (2002). *Literature circles : voice and choice in book clubs and reading groups* (2nd ed.). Stenhouse Publishers.

Day, R., & Bamford, J. (1998). Top Ten Principles for Teaching Extensive Reading. *Reading in a Foreign Language* 14(2), 136-141.

deHaan, J., Reed, W. M., & Kuwanda, K. (2010). The effect of interactivity with a music video game on second language vocabulary recall. *Language Learning & Technology*, *14*(2), 74–94. http://doi.org/10125/44215

deHaan, J., Reed, W. M., & Kuwanda, K. (2010). The effect of interactivity with a music video game on second language vocabulary recall. *Language Learning & Technology*, *14*(2), 74–94. http://doi.org/10125/44215

Dixon, D. H., Dixon, T., & Jordan, E. (2022). Second language (L2) gains through digital game-based language learning (DGBLL): A meta-analysis. *Language Learning & Technology*, 26(1), 1–25. http://hdl.handle.net/10125/73464

Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System (Linköping)*, *88*, 102166–13. https://doi.org/10.1016/j.system.2019.102166

Elley, W. B. (1989). Vocabulary Acquisition from Listening to Stories. In *Source: Reading Research Quarterly* (Vol. 24, Issue 2).

Ellis, R. (2015). *Understanding second language acquisition* (Second edi). Oxford University Press.

Ensslin, & Balteiro, I. (2019). *Approaches to Videogame Discourse: Lexis, Interaction, Textuality*. Bloomsbury Academic & Professional.

Franciosi, S. (2016). Acceptability of RPG Simulators for Foreign Language Training in Japanese Higher Education. *Simulation & Gaming*, *47*(1), 31–50. https://doi.org/10.1177/1046878115608621

Gee. (2007). *Good video games and good learning: collected essays on video games, learning, and literacy*. P. Lang.

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, 16(6), 511–525. https://doi.org/10.1016/j.learninstruc.2006.10.001

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, 16(6), 511–525. https://doi.org/10.1016/j.learninstruc.2006.10.001

Hartwick, P. (2018). Investigating research approaches: Classroom-based interaction studies in physical and virtual contexts. *ReCALL (Cambridge, England)*, *30*(2), 161–176. https://doi.org/10.1017/S0958344017000386

Hawreliak, J. (2019) in Ensslin, A., & Balteiro, I. (Eds.), *Approaches to videogame discourse: Lexis, interaction, textuality, 1st ed.,227-246*. Bloomsbury Academic & Professional.

Heidt, J. (2020). *"GG" Good Gaming Wordlists Insight into single-player commercial off-the-shelf games and action-roleplay game vocabulary* [Unpublished master's thesis]. Carleton University

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: Acquiring Second Language Vocabulary through Reading. *Reading in a Foreign Language*, 11, 207-223.

Hu, H., & Nation, I. (2000). Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language*, 13(1), 403-430.

Iwahori, Y. (2008). Developing Reading Fluency: A Study of Extensive Reading in EFL. *Reading in a Foreign Language*, 20(1), 70-91.

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

Jewitt, C. (2013). *Learning and communication in digital multimodal landscapes An inaugural professorial lecture by Carey Jewitt 13:23*.

Jewitt, C., O'Halloran, K. L., & Bezemer, J. J. (2016). *Introducing multimodality*. Routledge.

Krashen, S. (1985). *The Input Hypothesis: Issues and Implications*. Longman Press.

Kuppens, A. (2010). Incidental Foreign Language Acquisition from Media Exposure. *Learning, Media, and Technology*, 35, 65–85.

Lardiere, D. (2007). *Ultimate attainment in second language acquisition: a case study* . Lawrence Erlbaum.

Larson-Hall. (2016). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). Routledge. https://doi.org/10.4324/9781315775661

Lauer, C. (2009). Contending with Terms: "Multimodal" and "Multimedia" in the Academic and Public Spheres. *Computers and Composition*, *26*(4), 225–239. https://doi.org/10.1016/j.compcom.2009.09.001

Lee, S. M. (2019). Her Story or their own stories? Digital game-based learning, student creativity, and creative writing. *ReCALL*, *31*(3), 238–254. https://doi.org/10.1017/S0958344019000028

Liang, S. & Fu, Y. (2016) Otter.ai [Computer Software]. https://otter.ai/

Lightbown, P., Halter, R., White, J., & Horst, M. (2002) Comprehension-based Learning: The Limits of "Do It Yourself". *Canadian Modern Language Review*, 58, 427-464

Lin, P.M. 2014. Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences. *System,* 42: 164-176

Lindgren, E., & Munoz, C. (2013). The Influence of Exposure, Parents, and Linguistic Distance on Young European Learners' Foreign Language Comprehension. *International Journal of Multilingualism*, 10(1), 105-129

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), Handbook of second language acquisition (pp. 413-468). New York: Academic Press.

Mayer, & Anderson, R. B. (1991). Animations Need Narrations: An Experimental Test of a Dual-Coding Hypothesis. *Journal of Educational Psychology*, *83*(4), 484–490. https://doi.org/10.1037/0022-0663.83.4.484

Mayer, Fiorella, L., & Stull, A. (2020). Five ways to increase the effectiveness of instructional video. *Educational Technology Research and Development*, *68*(3), 837–852. https://doi.org/10.1007/s11423-020-09749-6

Mayer, R. (2019). *Computer games for learning: an evidence-based approach*.

Mayer, R. E. (2012). Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles. In *The Cambridge Handbook of Multimedia Learning* (pp. 183–200). Cambridge University Press. https://doi.org/10.1017/cbo9780511816819.013

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology*, 91(4), 638–643. https://doi.org/10.1037/0022-0663.91.4.638

Mayer. (2011). Towards a science of motivated learning in technology-supported environments. *Educational Technology Research and Development*, *59*(2), 301–308. https://doi.org/10.1007/s11423-011-9188-3

Mayer. (2021). *Multimedia learning* (Third edition.). Cambridge University Press.

McCosker, C. (2015). The effects of digital game based learning on the classroom dynamic. All Ireland Journal of Teaching and Learning in Higher Education, 7(2), 2311–2322.

McFadyen, N. (2020). Leveling up: A teacher's personal journey of bringing video games into a school andclassroom.Ludic Language Pedagogy,2,86–94.

Mesmer, H. A. E. (2008). Chapter 1: Using text analysis tools to match readers to texts. In *Tools for Matching Readers to Texts: Research-based Practices*, 1–10. Oxford University Press.

MKIceAndFire. (2014, June 22). *Beyond Two Souls Walkthrough HD* [Playlist]. YouTube. https://www.youtube.com/playlist?list=PL4vbGURud_HoeFKZDh7zB7oF2LWbZZdGH

MKIceAndFire. (2015, January 30). *Life is Strange Walkthrough HD* [Playlist]. YouTube. https://www.youtube.com/playlist?list=PL4vbGURud_HqV0_BNDG9AdW4zJzGQevTl

MKIceAndFire. (2016, February 9). *Firewatch | No Commentary | 100% Complete Gameplay Walkthrough* [Video]. YouTube. https://www.youtube.com/watch?v=Qm0iHU-K-HQ

MKIceAndFire. (2016, October 22). *SKYRIM SPECIAL EDITION Gameplay – First 2 Hours of SKYRIM Remastered [1080p HD PS4] - No Commentary* [Video]. YouTube. https://www.youtube.com/watch?v=K7NbJx6q-vk

MKIceAndFire. (2017, Oct 31). *ASSASSIN'S CREED ORIGINS Gameplay Walkthrough Part 1 [1080p HD PC MAX SETTINGS] - No Commentary* [Playlist]. YouTube. https://www.youtube.com/playlist?list=PL4vbGURud_HqugIgLZ3ZnRCDhmyyb9x_R

MKIceAndFire. (2020, April 15). *Tales from the Borderlands Walkthrough HD* [Playlist]. YouTube. https://www.youtube.com/playlist?list=PL4vbGURud_HoSJXqVRQQSLlswL9gszaD6

MKIceAndFire. (2020, May 11). *UNCHARTED DRAKE'S FORTUNE Gameplay Walkthrough Part 1 FULL GAME [4K 60FPS PS4 PRO] - No Commentary* [Video]. YouTube. https://www.youtube.com/watch?v=PfsLJcImiDE

MKIceAndFire. (2020, June 21). *THE LAST OF US REMASTERED Gameplay Walkthrough FULL GAME [1440P 60FPS PS4 PRO] - No Commentary* [Video]. YouTube. https://www.youtube.com/watch?v=zOET0ZMfayI

MKIceAndFire. (2020, July 27). *GHOST OF TSUSHIMA Gameplay Walkthrough Part 1 FULL GAME [1080P HD PS4 PRO] - No Commentary* [Video]. YouTube. https://www.youtube.com/watch?v=UOheOYV-H2U

MKIceAndFire. (2021, August 20). *TWELVE MINUTES Gameplay Walkthrough FULL GAME ALL ENDINGS [4K 60FPS PC] - No Commentary* [Video]. YouTube. https://www.youtube.com/watch?v=vFOowvQ7Nuw

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, *35*(4), 463–482. https://doi.org/10.1093/applin/amu024

Nagy, W., Herman, P., & Anderson, R. (1985). Learning Word Meanings from Context during Normal Reading. *American Educational Research Journal*, *24*(2), 237–270. https://doi.org/10.2307/1162893

Nation, I. (2006). How Large a Vocabulary is Needed For Reading and Listening? *Canadian Modern Language Review*, *63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, P. (2007). The Four Strands. *Innovation in Language Learning and Teaching*, *1*(1), 2–13. https://doi.org/10.2167/illt039.0

Nation, P. (2013). *Learning vocabulary in another language* (Second Edition.). Cambridge University Press.

Nation, I. (2014). How Much Input Do You Need to Learn the Most Frequent 9,000 Words? *Reading in a Foreign Language*, 26(2), 1-16.

Nation, I. (2017). The BNC/COCA Level 6 Word Family Lists (version 1.0.0) [datafile]

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., & Webb, S. (2018). Incidental vocabulary acquisition through viewing l2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Peterson, M. (2011). Digital gaming and second language development: Japanese learners' interactions in a MMORPG. *Digital Culture & Education*, 3(1), 56–73.

Plass, Mayer, R. E., & Homer, B. D. (2020). *Handbook of game-based learning*. The MIT Press.

Plonsky, L., & Oswald, F. L. (2014). How Big Is "Big"? Interpreting Effect Sizes in L2 Research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

QSR International Pty Ltd. (2020) NVivo (released in March 2020), https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Rankin, Gold, R., & Gooch, B. (2006). Playing for keeps: gaming as a language learning tool. *ACM SIGGRAPH 2006 Educators Program*, 44–es. https://doi.org/10.1145/1179295.1179340

Reinders, H., & Wattana, S. (2012). Talk to me! Games and students' willingness to communicate. In H. Reinders (Ed.), *Digital games in language learning and teaching* (pp. 156–188). Palgrave.

Reinhardt, J. (2019). *Gameful second and foreign Language teaching and learning theory, research, and practice*. Palgrave Macmillan.

Rodgers, M. (2018). The Images in Television Programs and the Potential for Learning Unknown Words. *ITL- international journal of applied linguistics*, 169(1), 191-211

Rodgers, M. & Heidt, J. (2020). Levelling Up Comprehensible Input and Vocabulary Learning: The Lexical Profile Of Videogames. W. Werner & F. Tegge (eds.) *pop culture in language: theory, research ,practice* (pp.213-225). Taylor & Francis Group.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, *21*(4), 589–619. https://doi.org/10.1017/S0272263199004039

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), Cognition and Second Language Instruction (pp. 3-32). Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524780.003

Schmitt, N. & Schmitt, D. (2020). *Vocabulary in language teaching* (Second edition.). Cambridge University Press.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and Exploring the Behaviour of Two New Versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88.

Sedig, K., Parsons, P., & Haworth, R. (2017). Player–game interaction and cognitive gameplay: A taxonomic framework for the core mechanic of videogames. *Informatics*, *4*(1). https://doi.org/10.3390/informatics4010004

Selen, K., & Zimmerman, E. (2003). *Rules of play: game design fundamentals*. MIT Press.

Shelton-Strong, S. (2012). Literature Circles in ELT. *ELT Journal*, *66*(2), 214–223. https://doi.org/10.1093/elt/ccr049

Skehan, P. & Foster, P. (2001) Cognition and tasks. In P. Robinson (Ed.), Cognition and Second Language Instruction (pp. 3-32). Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524780.003

Sundqvist, P., & Sylvén, L. K. (2014). Language-related computer use: Focus on young L2 English learners in Sweden. *ReCALL (Cambridge, England)*, *26*(1), 3–20. https://doi.org/10.1017/S0958344013000232

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/10125/44674

The jamovi project (2021). jamovi. (Version 2.2) [Computer Software]. Retrieved from https://www.jamovi.org

Tsai, Y., & Tsai, C. (2018). Digital game-based second-language vocabulary learning and conditions of research designs: A meta-analysis study. *Computers & Education*, *125*, 345–357. https://doi.org/10.1016/j.compedu.2018.06.020

Van Zeeland, H., & Schmitt, N. (2013). Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension? *Applied Linguistics*, 34(4), 457-479.

Waring, R. & Takaki, M. (2003). *At what rate do learners learn and retain new vocabulary from reading a graded reader?* http://nflrc.hawaii.edu/rfl

Waring, R. and Nation, I. (2004) Second Language Reading and Incidental Vocabulary Learning. *Angles on the English Speaking World*, 4, 97-110.

Webb, & Nation, I. S. P. (2017). *How Vocabulary Is Learned*. Oxford University Press.

Webb, S. & Rodgers, M. P. H. (2009a). Vocabulary Demands of Television Programs. *Language Learning*, *59*(2), 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

Webb, S. & Rodgers, M. P. H. (2009b). The Lexical Coverage of Movies. *Applied Linguistics*, *30*(3), 407–427. https://doi.org/10.1093/applin/amp010

Wijman, T. (2021, December 3). *Newzoo's trends to watch in 2021 (Games, esports and Mobile)*. Newzoo. Retrieved April 1, 2022, from https://newzoo.com/insights/articles/newzoos-trends-to-watch-in-2021games-esports-and-mobile/

Willis, D., & Willis, J. (2007). Doing Task-Based Teaching. Oxford University Press.

Yee, N., & Bailenson, J. (2007). The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human Communication Research*, *33*(3), 271–290. https://doi.org/10.1111/j.1468-2958.2007.00299.x

York, J., Thanyawatpokin, B., & Dehaan, J. (2019). “Kotoba Rollers” walkthrough: Board games, TBLT, and player progression in a university EFL classroom. In *Ludic Language Pedagogy* (Vol. 1). https://www.youtube.com/watch?v=XG-rUIYLd_c

York, J., Poole, F. J., & deHaan, J. W. (2021). Playing a new game—An argument for a teacher-focused field around games and play in language education. *Foreign Language Annals*, *54*(4), 1164–1188. https://doi.org/10.1111/flan.12585

Appendix A

Pre-gaming teacher intervention example

Game Research

READ A GAME GUIDE

There are game walkthrough instructions online. These include controls, chapter synopses and helpful strategies. Please read one chapter each. Students that are not reading should look up difficult words and write them in the table below

WATCH A WALKTHROUGH

There are some good walkthough videos on Youtube. Watch a segment of one with commentary online

Search YouTube with: <your game> walkthrough with commentary

and watch some native speakers play.

RECORD IMPORTANT VOCABULARY

Please write some important nouns and verbs that are used in the game guide and video walkthroughs.

An example for each is provided:

English (Verb)	First-language translation	English (Noun)	First-language translation	English (Adjective)	First-language translation
bluff	だます	*pawn*	コマ	*amazing*	すばらしい
1.		1.		1.	
2.		2.		2.	
3.		3.		3.	
4.		4.		4.	
5.		5.		5.	
6.		6.		6.	
7.		7.		7.	
8.		8.		8.	
9.		9.		9.	
10.		10.		10.	

What was the most difficult sentence in the rulebook? Write it out here and translate it into Japanese:

English sentence:

First-language translation:

Adapted from: York, J., Thanyawatpokin, B., & Dehaan, J. (2019). “Kotoba Rollers” walkthrough: Board games, TBLT, and player progression in a university EFL classroom. In *Ludic Language Pedagogy* (Vol. 1). https://www.youtube.com/watch?v=XG-rUIYLd_c

Appendix B

Post-gaming teacher intervention example

Summarizer

Gaming Circle Role Sheet

NAME: _____

DATE: _____

CIRCLE: _____

GAME: _____

YOUR ROLE

Your job is to prepare a brief summary of this week's play session. Your group discussion will start with your 5-minute statement that covers:

- the key points
- main highlights
- general idea of this session

SUMMARY

_____

KEY POINTS

- _____
- _____
- _____
- _____
- _____
- _____

CONNECTIONS

Did this session remind you of anything? Explain

_____

Figure B1. Summarizer role

Word Wizard

Gaming Circle Role Sheet

NAME: _____

DATE: _____

CIRCLE: _____

GAME: _____

YOUR ROLE

Your job is to be on the lookout for a few words that have special meaning in this week's game selection.

- Jot down puzzling or unfamiliar words while you are playing. Later, look up the definitions in either a dictionary or some other source.
- You may also run across words that stand out somehow in the game – words that are repeated a lot, used in an unusual way, or are crucial to the meaning of the text. Mark these special words, too, and be ready to share your ideas on their usage to the group.

WORD / DEFINITION

_____

STAND-OUT WORDS

- _____
- _____
- _____
- _____
- _____
- _____

STAND-OUT PHRASES

_____

Figure B2. Word Wizard role

Treasurer

Gaming Circle Role Sheet

NAME: _____

DATE: _____

CIRCLE: _____

GAME: _____

YOUR ROLE

Your job is to prepare keep a record of any useful items from thisthis week's play session. You should keep an eye out for:

- helpful Items
- plot items
- How do these items work?

INVENTORY

PLOT ITEMS

- _____
- _____
- _____
- _____
- _____
- _____

WHAT CAN YOU DO WITH THESE ITEMS?

Please explain how to use these items and record what they do.

Figure B3. Treasurer role

Investigator

Gaming Circle Role Sheet

NAME: _____

DATE: _____

CIRCLE: _____

GAME: _____

YOUR ROLE

Your job is to dig up some background information on any relevant topic related to your game. This might include

- The geography, weather, culture, or history of the game's setting
- Information about the time period portrayed in the game
- Information on any topics or events represented in the game
- Information on any topics or events that may have inspired the story
- Information about any character that is based on a historical person

SETTING

_____

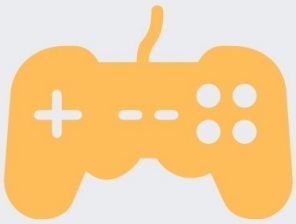
INSPIRATION

- _____
- _____
- _____
- _____
- _____
- _____

HISTORY

_____

Figure B4. Investigator role



Captain

Gaming Circle Role Sheet

NAME: _____

DATE: _____

CIRCLE: _____

GAME: _____

YOUR ROLE

Your job is to develop a list of questions that your group might want to discuss about this part of the game. Don't worry about the small details; your task is to help people talk over the big ideas in the game and share their reactions. Usually the best discussion questions come from your own thoughts, feelings, and concerns as you play.

You are also encouraged to ask your group about any strategies that they have used in play.

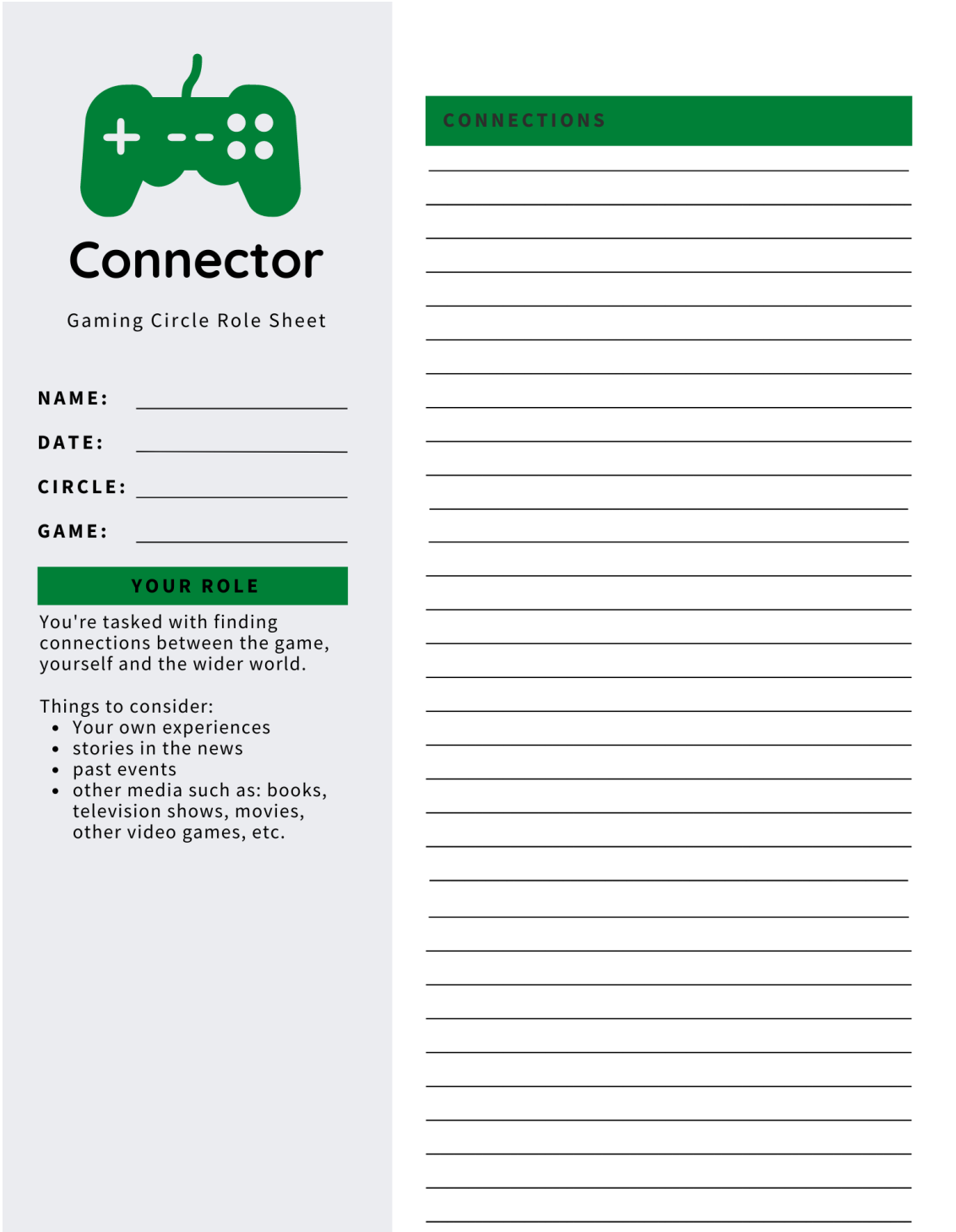
DISCUSSION QUESTIONS

- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____

STRATEGIES

_____
_____
_____
_____
_____
_____
_____
_____
_____

Figure B5. Captain role



Connector

Gaming Circle Role Sheet

NAME: _____

DATE: _____

CIRCLE: _____

GAME: _____

YOUR ROLE

You're tasked with finding connections between the game, yourself and the wider world.

Things to consider:
- Your own experiences
- stories in the news
- past events
- other media such as: books, television shows, movies, other video games, etc.

CONNECTIONS

Figure B6. Connector role

Adapted from: Daniels, H. (2002). *Literature circles : voice and choice in book clubs and reading groups* (2nd ed.). Stenhouse Publishers.