Computer-Assisted Language Learning Electronic Journal (CALL-EJ), 24(3), 107-132. 2023

Watching Gameplay or Playing Games: Measuring the Effects of Physical Interactivity on Language Learning

Shawn Andersson (andersson.hmt@osaka-u.ac.jp)
Osaka University, Graduate School of Humanities, Division of Foreign Studies, Japan

Abstract

Game-based language learning involves learners playing digital games for language-learning purposes. Such applications have predominantly involved users playing the games, as 'games' have been synonymous with 'play'. Nevertheless, watching gameplay online has gained significant popularity in which people elect to watch someone else play a game instead of playing it themselves. While watching gameplay has been addressed in the literature, its effectiveness for foreign language learning remains underexplored. In this regard, the main difference between playing and watching a game is the inclusion or exclusion of physical interactivity, or the utilization of a controller to manipulate the gameplay. This study includes a five-week experiment where thirty-two (n = 32) participants were assigned to player or watcher roles. A mixed-method approach was employed consisting of vocabulary tests, questionnaires, and interviews. The main results indicated positive vocabulary recall for both groups and slightly higher for the players. Nevertheless, no statistical difference was found. Additionally, the survey showed players expressing better attitudes toward pedagogy. Next, players recorded a higher mental effort and task difficulty, yet this did not hinder vocabulary recall or perceived effectiveness. The interviews suggest that, while some believed watching allows better learning opportunities, most participants felt watching games will hinder learning due to the loss of concentration.

Keywords: computer-assisted language learning (CALL), game-based language learning (GBLL), watching gameplay, physical interactivity, foreign language learning

Introduction

Game-based language learning (GBLL) is a research field exploring the facilitation of language learning through gameplay, drawing on the entertainment factors of games as a catalyst for motivation and learning. While the platform for games includes a wide range of avenues, the more specific field of digital game-based language learning (DGBLL) focuses solely on digital computer games. This creates overlaps and similarities to computer-assisted language learning (CALL) as both relate to technological applications for language learning.

Games have been shown to possess aspects favorable for facilitating learning, such as providing challenge, competition, purpose, and control (Admiraal et al., 2011; Nakamura & Csikszentmihalyi, 2009; Whitton, 2014). Games also promote learning by creating immersive (Johnsen et al., 2021) and safe (Jabbari & Eslami, 2019) environments, providing feedback (Calvo-Ferrer, 2021), and allowing for language negotiation to occur through communicative interaction and collaboration (Peterson, 2016). They also allow for ample scaffolding opportunities (Sun et al., 2021), giving educators the flexibility to individually address students' proficiency levels.

DGBLL research has predominantly involved learners playing games themselves through controllers because digital games were synonymous with playing games. Nevertheless,

in recent years, watching gameplay has become a popular phenomenon, especially amongst high school and university students, where nearly half (41%) of the popular game streaming website, Twitch.tv's audience is between the age of 16 and 24 (Iqbal, 2022).

Watching gameplay is the act of someone watching another person (streamer) play a digital game. These sessions can be either live or prerecorded through websites such as Twitch.tv and YouTube, the latter of which is particularly popular in Japan (Andersson, 2022; Andersson, 2023). Watching gameplay is "...a kind of real-time video social media that integrates traditional broadcasting and online gaming" (Li et al., 2020, p. 1). The streamers often provide commentary on their gameplay, and users are able to communicate with the streamer and fellow viewers through a chat box during live sessions. Watching gameplay has significantly increased in viewership and market value (Hamilton et al., 2014), often rivaling cable TV networks (Gilbert, 2018). This trend has increased during the COVID-19 pandemic (Clement, 2022) when people sought to alleviate associated negative emotions by turning to video games (Scerbakov et al., 2022). Notably, in various cases, more people watch someone else play a game than play it themselves (Andersson, 2022; Kaytoue et al., 2012; Orme, 2021). This trend is also taking hold in Japan, as illustrated in a 2021 survey of middle school students' ideal future careers. For the first time, "Game Streamer" was ranked top five for males and top ten for females (Sony Life Insurance Corporation, 2021).

Within game-watching sessions, there is even evidence that language learning is taking place. Here, the author was able to find a streamer who teaches viewers Japanese by playing Japanese role-playing games while explaining the in-game vocabulary. Additionally, various streamers advertise their channels as bilingual and alternate between two languages while playing games. While this is occurring, the viewers communicate in the chat box in both languages, drawing in diverse communities for potential language interactions, all of which are connected through their interest in watching games. Yet, despite its popularity and these instances of learning, we know little about the potential for language learning in terms of pedagogical effectiveness, learners' attitudes toward adopting this method, and their perceptions of its usefulness.

When examining watching gameplay for language-learning purposes, similarities can be observed regarding traditional media watching. Media watching is not new to education research, as studies have extensively looked at media applications such as movies and TV shows for language-learning applications (Parmawati & Inayah, 2019; Vanderplank, 2019). Yet, whereas media watching is considered mostly a passive experience, watching gameplay offers added opportunities for language learning through the abovementioned interactions between viewers and streamers, and communicative interaction has been shown to be conducive to language learning in game environments (Peterson, 2016).

While watching gameplay has been addressed in the literature, studies mostly focus on it as a social phenomenon, and its potential for language learning remains underexplored. Given the popularity of watching gameplay, it is important to thoroughly investigate its potential for language learning and compare it with traditional DGBLL applications.

Literature Review

Krashen (1981, p.1) defines first and second-language acquisition (SLA) as a process involving "... interaction in the target language–natural communication–in which speakers are concerned not with the form of their utterances but with the messages they are conveying and understanding." Thus, compared to foreign language learning, which focuses on explicit grammar instruction and error-checking, SLA takes a similar approach to how individuals learn their first language (Krashen, 1981). This notion has been significantly expanded upon and has become commonplace amongst the scholarly community. Nevertheless, Peterson (2013) points

out that there is limited research pertaining to SLA theory and CALL in general. Additionally, there is a tendency for educators to use the most recent technology applications with no regard for their usefulness (Peterson, 2013) or implement behavioristic practices where target learning is attained through repetition rather than meaning (Whitton, 2014). Therefore, Peterson (2013) argues for the importance of integrating theory with practice.

Within the realm of digital games, there have been multiple avenues of SLA theories considered, and the most prevalent attempts can be separated into either cognitive or social-linguistic constructs (Lightbown & Spada, 2013; Peterson, 2013; Whitton, 2014). Cognitive theories, which are the focus of this study, pertain to the mental processes occurring within learners themselves as they interact with a given system (Peterson, 2013). Socio-linguistic approaches, on the other hand, focus on how learners acquire a second language through mutual communicative interaction between others (Peterson, 2016).

Within the available studies on digital games and CALL, vocabulary recall is commonly used as a determiner to gauge if acquisition has occurred. This is because, while one's ability to remember vocabulary or phrases does not offer a complete representation of their language capabilities, words do make up the foundation of language and, thus, point to the importance for comprehension within SLA (Bowen & Marks, 2002). In line with this, Ebrahimzadeh (2017, p. 1) states that "lacking sufficient word knowledge results in not being able to produce a single utterance." Additionally, measuring vocabulary recall offers scholars a way to feasibly and objectively collect quantitative data through experimentation. Therefore, while the measurement of vocabulary has limitations, it can still offer useful information for a preliminary understanding of novel approaches that can be followed by more robust methods.

Nation (2001) argued that there are generally three processes that may allow for the retention of vocabulary, which include noticing, retrieval, and creative use. Here, noticing involves learners actively perceiving and attending to linguistic features in the input (Truscott, 1998), retrieval is the action of choosing particular vocabulary when producing second language (L2) output (Manchon et al., 2007), and creative use involves using vocabulary in various ways (Nation, 2001). Here, digital games are seen to offer multiple avenues for vocabulary learning. For example, players are given opportunities to notice the meaning of unknown vocabulary by deducing their meaning by how the game scenario is played out. Additionally, games also offer chances for creative use through peer communication, especially in multiplayer games.

The Watching-Gameplay Phenomenon

With the rise in watching gameplay, studies have addressed the trend but primarily focus on it as a social phenomenon unrelated to its potential for learning. Most studies can be categorized into the subjects of online social interaction (Churchill & Wen Xu, 2016; Diwanji et al., 2020; Hamilton et al., 2014), media consumption (Jang & Byon, 2019; Sjöblom et al., 2017), and the general motivational appeal for watching (Gros et al., 2017; Sjöblom & Hamari, 2017). Regarding the latter, observing motivational appeal could reveal language learning as a potential factor. Unfortunately, however, related studies utilize quantitative surveys with predetermined categories with only limited answering criteria such as entertainment, communicating with others in the online community, checking out a game before purchasing it, and learning game strategies. Importantly, 'language learning' is not a selectable option. Thus, despite the recent studies, there is minimal data addressing the relative linguistic effectiveness of watching a game as opposed to playing one.

When addressing the linguistic effectiveness of watching gameplay, it is important to note that the main difference between the two is the utilization or exclusion of game physical interactivity. The word 'interactivity' has different meanings in varying contents and can pertain

to communicative interactions between two or more people. Nevertheless, physical interactivity (hereafter referred to simply as 'interactivity') can be described as a player's physical interaction they have with the game using an electronic device (deHaan et al., 2010; Sims, 1997). It is described as the act of directly manipulating gameplay via a hand-operated controller, something only a player, not a watcher, can do and should not be confused with human-to-human communicative interactions. While studies on game interactivity are not directly related to the watching gameplay phenomenon, they are nonetheless important when examining the relative effectiveness of foreign language learning.

Researchers typically measure the effects of game interactivity on vocabulary acquisition, attention, mental effort and cognitive load, motivation, and flow experience (Cho et al., 2021; deHaan & Kono, 2010; deHaan et al., 2010; Ebrahimzadeh, 2017; Ebrahimzadeh & Alavi, 2016). Prior studies on interactivity often express conflicting evidence. On the one hand, it has been shown that, when learners are engaged in tasks that elicit a high degree of involvement, language acquisition, such as vocabulary learning, has been known to be facilitated (Peterson, 2021). Ali Mohsen (2016) provided support for this in a study on vocabulary learning for an educational game (later explained as a serious game) designed to teach knee surgery with results showing the players outperformed the watchers. Oppositely, interactivity has been negatively associated with causing split attention and overwhelming mental capacity to acquire target learning. This is because handling a physical controller "requires frequent input from the player and the input required can disrupt the player's involvement with the game space" (Taylor, 2002, p. 20). This draws similarities to prior studies showing instructional media applications causing split attention and extraneous cognitive load (Kalyuga et al., 1999), factors intrusive to learning. Cognitive load theory postulates that one's mental capacity is finite and can be overwhelmed with various mental tasks, which decreases opportunities for target learning (Sweller, 1994). In accordance with cognitive load theory, physical interactivity may increase mental effort (Pellouchoud et al., 1999) and hamper vocabulary learning (deHaan, 2005). Researchers have recommended that the level of cognitive load created by physical input should be addressed further (deHaan et al., 2010; Plass & Jones, 2005).

Other studies have inadvertently tested game interactivity by assessing the level of technological engagement (LTE) on learning in various education fields unrelated to foreign language learning. These studies test whether higher technology engagement leads to better measurable learning results and encompass various subjects such as software engineering (Gordillo et al., 2022; Hsu & Lin, 2016), biology (Chang et al., 2016), science (Chen et al., 2021), intercultural learning (Busse & Krause, 2016), history (Ijaz et al., 2017), and math (Kebritchi et al., 2010; Wang et al., 2018). Within these studies, a game-based application is often used as the highest level of LTE (i.e. high interactivity) since it involves users actively engaging with technologies. Then a video-based treatment is implemented at the middle level as a means of low engagement. Lastly, traditional classroom teaching constitutes no LTE.

Serious games and commercial off-the-shelf games

Within DGBLL applications, researchers employ one of two distinctly different game types. Commercial off-the-shelf (COTS) games, or non-serious games, are produced solely for entertainment purposes. Only after these games are completed and released to the public do researchers use them for learning in various fields, including foreign language learning (see deHaan et al., 2010; Ebrahimzadeh, 2017; Peterson, 2016). The second category of games is called serious games, and these are designed from the ground up primarily for educational purposes. This distinction is important because, while DGBLL research has grown considerably recently, most studies utilize serious games often created by the researchers

themselves (Boyle et al., 2016; Girard et al., 2013). In fact, the studies mentioned in the previous paragraph testing game interactivity via LTE are all considered serious games. Oppositely, the vast majority of online gameplay watching is done through COTS games.

While rare, there are studies that investigate interactivity with COTS games. deHaan (2005) tested interactivity for a baseball game on Japanese students with results indicating students had split attention due to them simultaneously controlling the gameplay while learning the language. Two follow-up studies were conducted, incorporating vocabulary achievement and subjective cognitive load for players and watchers for a reflex-based game (deHaan & Kono, 2010) and a music game (deHaan et al., 2010). Both studies showed watchers scoring higher in vocabulary achievement with less extraneous (adverse) cognitive load. This indicated that watching may have language-learning advantages over playing games since the watchers are free to focus on the target learning without the interruption of a controller. More recently, two similar studies by the same author (Ebrahimzadeh & Alavi, 2016; Ebrahimzadeh, 2017) looked at the effects of interactivity for a multiplayer online battle arena (MOBA) game. In the first study (Ebrahimzadeh & Alavi, 2016), 136 high school students played or watched the game for five weeks with measures including e-learning enjoyment and vocabulary recall. The results showed no significant difference between players and watchers. In the later study (Ebrahimzadeh, 2017), 241 male high school students were assigned to groups of readers, players, and watchers for five weeks. Here, players and watchers outperformed the readers. It is important to note that the focused learning of these studies included mainly short, text-only dialogues with minimal audio input. Additionally, other studies have looked at the effects of interactivity in COTS games on various factors unrelated to education. Examples of this include cognitive load and rape acceptance (Read et al., 2018), motivational processing and cognitive load (Huang, 2011), the responsibility and degree of character identification in violent games (Walter & Tsfati, 2016), and violence, perceived difficulty, and frustration (Polman et al., 2008).

Different games may be better suited for language learning (deHaan et al., 2010), and it is therefore essential to obtain data from various game genres to achieve a broad perspective. In this respect, real-time strategy (RTS) games are a popular subgenre within strategy games that, unlike turn-based strategy games, create a fast-paced playing experience with high player interactivity since all players move simultaneously. Thus, RTS games have been shown to generate good flow experiences (Sweetser & Wyeth, 2005), and measuring the effects of interactivity for this genre may produce interesting results.

Research Questions

In an effort to measure the relative effectiveness between playing and watching a game for language-learning purposes, this study tests the effect of the physical interactivity of a COTS RTS game on vocabulary recall and subjective invested mental effort. Additionally, this study addresses the feasibility of practical applications by observing learner attitudes toward playing and watching games for language-learning purposes and their perceptions of its effectiveness (see Bolliger et al., 2015). The research questions are as follows.

Q1 What measurable effect does physical interactivity (playing versus watching) have on immediate vocabulary achievement for a real-time strategy game?

Q2 What measurable effect does physical interactivity have on delayed vocabulary achievement for a real-time strategy game?

Q3 What are learners' motivation and attitudes toward playing or watching games for language-learning purposes, and what are their perceptions about the effectiveness of language learning before and after the treatment?

Q4 What effect does physical interactivity have on subjective invested mental effort (of the assigned task and learning English) and the perceived difficulty (of the assigned task and the game's language)?

Method

Experiment Design

This experiment follows a previous preliminary study (Andersson, 2022) and includes more participants and a longer experiment length of five weeks as opposed to a one-time treatment. The experiment was designed to test the effects of game interactivity on vocabulary recall, attitudes and perceptions for pedagogy, and subjective mental effort.

Interactivity, or the ability to control the gameplay, was assigned the independent variable and was controlled by assigning participants into one of two groups: players who played the game and watchers who only watched. Dependent variables included vocabulary achievement, attitudes and perceptions, and perceived mental effort and difficulty. To eliminate communicative interaction influences, participants were not permitted to talk with each other during the treatments. A mixed-methods data collection approach was employed consisting of vocabulary tests (immediate and delayed), questionnaires, and post-experiment interviews, as recommended by (deHaan et al., 2010).

Participants

The participants were recruited from one campus at a university in Japan via convenience sampling due to necessity. All participants were first-year undergraduate students taking similar compulsory English communication classes and were Japanese native speakers. Initially, 50 participants completed a survey expressing their interest in participating and inputted their availability. Of these, 42 participants took part in the first session, and over the course of the experiment, several were dropped after missed sessions. In all, 32 (n=32) students completed all five sessions and were paid for their participation.

A summary explanation of the participants is given as follows. Eleven were males and twenty-one were females, of which 22 majored in Economics and 10 in Management of Gastronomy. The participants were asked to volunteer either their English proficiency test scores or self-rated English proficiency levels. Six participants inputted their test scores, including TOEIC (550, 650), TOEFL (490), IELTS (5.0), and Eiken (level 1, level 2). Self-rated levels included 14 at the beginner level, one between beginner and intermediate, eight at the intermediate level, and two at the upper-intermediate level. Twenty-eight participants have no study abroad experience while four have studied abroad for an average of 3.46 months. The most popular method of studying English was movies/videos (21 occurrences), followed by playing digital games (17) and smartphone apps (17), and university courses (15).

Regarding playing games habits, eighteen participants play digital games while fourteen do not. The hours of weekly gameplay were diverse in that six participants play for less than one hour, five for between two to four hours, three for between five to seven hours, and four for more than eight hours a week. Most play games on gaming consoles (11 occurrences) or smartphones (11), and only three play games on PCs. Role-playing games (RPGs) were the most popular (12 occurrences), followed by strategy (7), and sports (7). Of

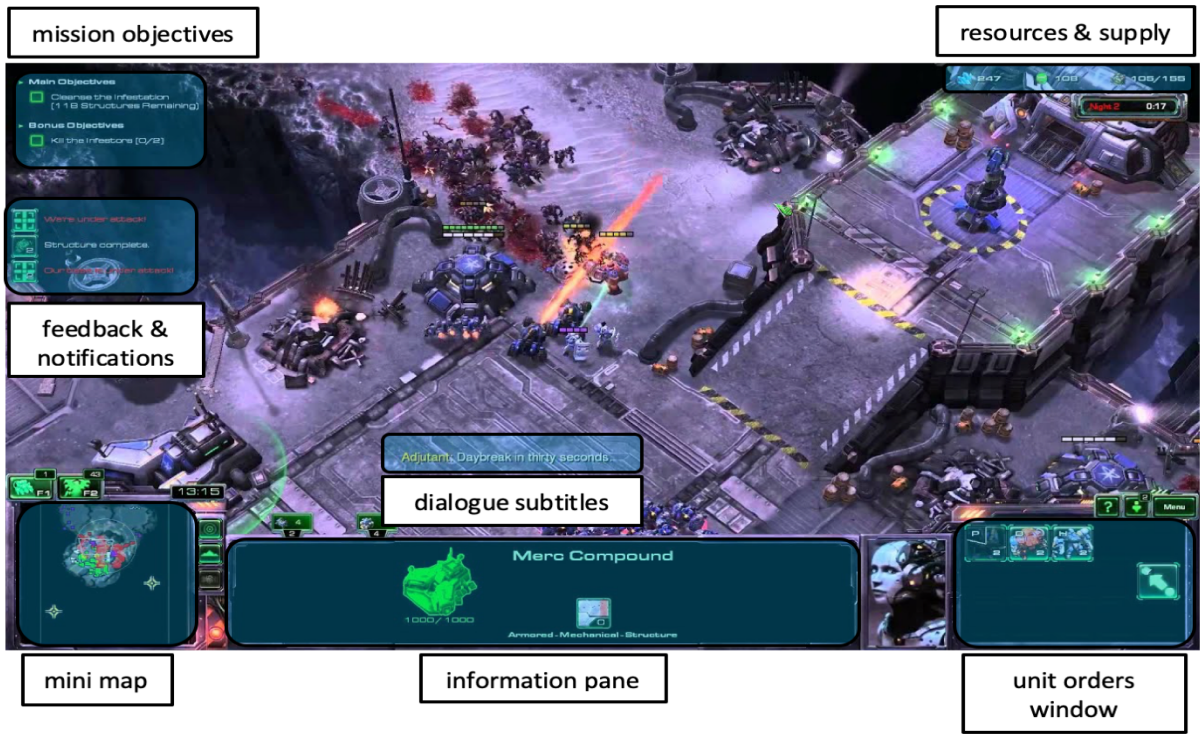
the eighteen who play digital games, nine have no experience playing games requiring the use of a mouse and keyboard such as the game used in this current experiment. Twelve have never played an RTS game before, two play them occasionally, and four often play them. None of the participants in this study have played this experiment's game. Finally, seven participants have tried learning English while playing a digital game.

Next, more than half of the participants (17/32) watch gameplay. Six of these participants watch for one hour or less a week, eight between two to four hours, and three between five to seven hours. The most popular genres watched were RPG (10 occurrences) and FPS (10), followed by Puzzle (5), Strategy (4), and Sports (4). The main reason for watching gameplay was for entertainment (17 occurrences), followed by learning new playing strategies (6) and checking out a game before purchasing it (5). There was only one occurrence of language learning being a reason to watch. All participants who watch gameplay utilize YouTube, and three participants also use Twitch.tv. Finally, five have tried learning English by watching gameplay.

Software Resources

The game selection criteria included a COTS RTS game with a single-player, mission-based campaign for better control. Necessary features included ample audio and player interactivity during dialogue sequences instead of movie-like cinematic cutscenes. Based on these criteria, StarCraft II: Wings of Liberty was chosen. StarCraft II was produced by Blizzard Entertainment in 2010 and continues to have a significant presence aside from its single-player campaign in global competitive championships. The game entails players controlling army units, constructing buildings, and pursuing mission objectives in real-time (see Figure 1). While the game can be fully operated using only a mouse, optimal gameplay requires the use of both a mouse and keyboard.

Figure 1
Game interface

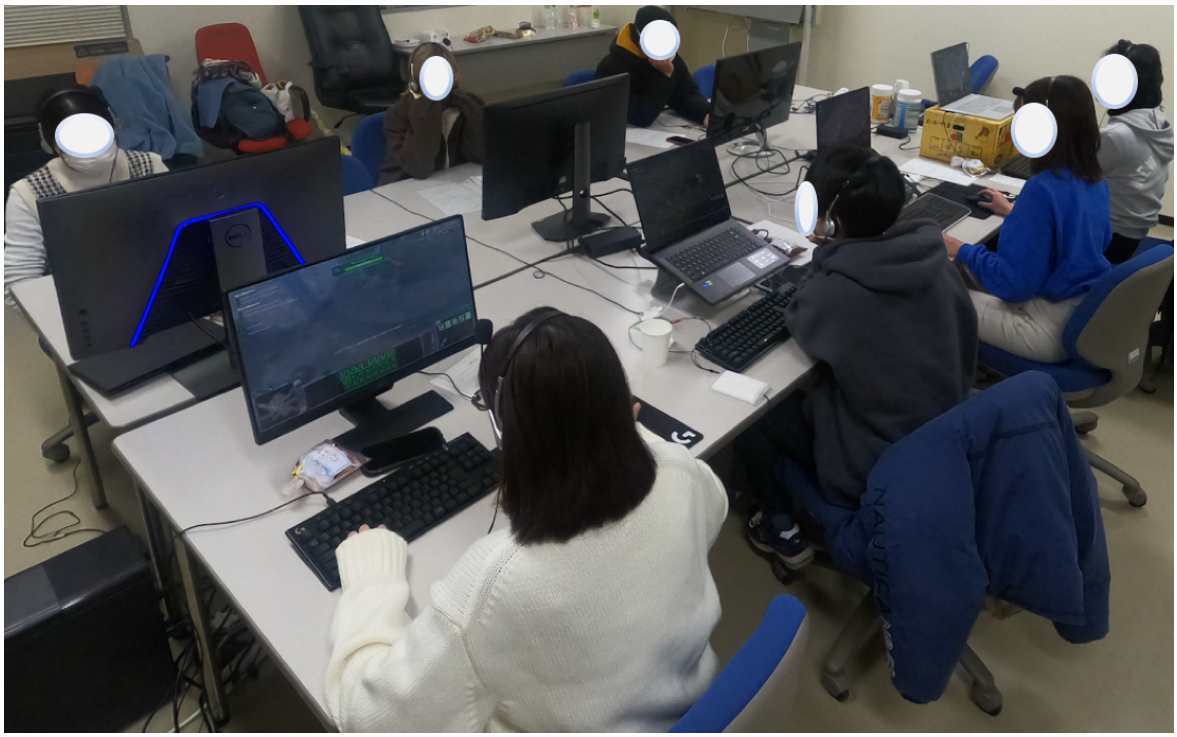


The author selected five missions by first identifying several potential missions, then writing out the entire mission dialogue scripts, and finally assessing them based on applicable vocabulary, approximate time length, and game difficulty. The difficulty was set to the lowest level of 'casual,' which the game states is recommended for players with little to no experience playing RTS games. The graphics were equally set to 'high' after ensuring each computer could run at this setting above 60 frames per second.

Physical Setup and Resources

The players and watchers sat across from each other as can be seen in Figure 2. The players were assigned a computer, mouse, keyboard, and stereo headphones. The experiment computers consisted of three laptops utilizing their built-in 19-inch screens and one desktop computer using a 22-inch monitor. All computers were equipped with Intel i7 or i9 processors, Nvidia GeForce RTX 3080 or GTX1060 graphics cards, and at least 16GB of RAM, which exceed the manufacturer's recommended hardware specifications. A video and sound splitter were used to run the simultaneous video and audio feed to the watchers. The watchers viewed the gameplay on 27-inch monitors, and groups with more than one watcher shared the same screen.

Figure 2
Experiment physical setup



Procedures

Grouping

Prior to the experiment, the researcher conducted an initial survey to solicit volunteers of which 50 participants completed and inputted their availability. This information was used

to group participants in pairs (one player and one watcher) with similar self-rated English proficiency levels, international experience, and game experience (deHaan et al., 2010). Nevertheless, participants' availability also needed to be considered. Next, 41 participants completed the pre-work and attended the first session and were placed into groups and randomly assigned to the player or watcher role. Since this was a longitudinal study, some players dropped from the experiment and left watchers without players. In these cases, the watchers were integrated into preexisting groups. In the end, 32 (n=32) participants successfully completed all 5 sessions (15 players and 17 watchers). These participants were in 15 groups, each with one player and mostly one or two watchers; four groups had only one player remaining, and one group had a player and four watchers.

Pre-work

A week before the experiment, the participants were given tasks to complete at their leisure. They were first instructed to read an explanation in Japanese outlining the experiment scope and data collection and give their informed consent. Next, they completed a vocabulary pre-check and questionnaire. Finally, they watched a 15-minute video in Japanese on YouTube explaining the core game mechanics and how to play the Terran (human) race, which the game's single-player campaign focuses on. This was done ahead of time to prevent participants from becoming bored by reducing the in-person treatment time (deHaan et al. 2010) while also ensuring that the participants were given a sufficient understanding of the complex game mechanics.

Sessions

The sessions occurred once a week on the same day and time, and each week introduced a new mission. The participants were assigned to their stations, which had a handout explaining each step in the session. They were instructed to either play or watch the mission and try to learn the language. During this time, no communicative interaction with other participants, notetaking, or word searching were permitted. The players were not permitted to pause the game unless necessary. And in the event of mission failure or success, they were asked to await further instructions from the researcher regarding repeating the mission or finishing the experiment. The researcher then explained these steps verbally, and the participants were given the opportunity to ask questions.

For the first week only, the session started with the game's tutorial explaining the game mechanics, which was conducted in English since there is no Japanese language option. All five sessions included a warm-up stage lasting 15 minutes on a skirmish map against a computer opponent (Terran vs. Terran) on the lowest difficulty. During this time, the players were free to play however they liked, and the researcher periodically gave them playing advice. The participants were informed that they did not need to learn the English during this time and could focus on learning the game. During these stages, the watchers also watched to ensure a consistent experience.

The treatment then began, which, after excluding the warm-up, post-test, and survey, averaged 31 minutes per session and two hours and 25 minutes for all five sessions. When necessary, the researcher acted as an active participant in cases such as participants asking questions or encountering mission failure. This was deemed necessary to ensure a smooth treatment, as observed in the preliminary study (Andersson, 2023).

Data Collection and Analysis

Vocabulary Tests

The following steps were taken for selecting the vocabulary items. First, the entire English dialogues of the five missions were written out. Next, both individual words and compound words such as phrasal verbs and compound nouns were selected. Colloquial words or idioms were excluded as they require cultural context to understand. To ensure a consistent experience, vocabulary items that were only experienced by players completing optional objectives were not used.

Prioritization was given to (1) words relevant to the mission scenario and means of accomplishing the mission, (2) observable dialogue in that the prompt was given, followed by the participant immediately seeing its action played out on the screen, allowing them to deduce the meaning even if they don't know the word, and (3) low-frequency words most likely unknown to the participants. In total, ten items were selected to be introduced weekly, equaling 50 altogether. The frequency of these items was dictated by the player, and no complete control could be implemented on the number of times each word was witnessed. Nevertheless, grouping the participants ensured a consistent experience within each group.

Prior to the experiment, a vocabulary pre-check was implemented instead of a pre-test to avoid priming the participants where the participants were asked to indicate (yes or no) if they knew the vocabulary item. An additional eight unrelated distracter items were also created for a total of 58 items.

All vocabulary tests items were arranged in random order, and the participants were instructed not to use dictionaries or Internet searches to find the answers. The post-tests (immediate and delayed) consisted of multiple-choice questions with four selectable answers. A sample sentence from the mission was displayed with the target items highlighted. The participants were required to select the equivalent Japanese translation of the item. The 2-week delayed post consisted of the same 50 questions.

Questionnaire

The survey consisted of 22 Likert questions and four categories incorporated from prior studies and used in the preliminary study (Andersson, 2023). All questions were translated from English to Japanese by a native Japanese speaker with English fluency.

First, criterion measures (5 questions) were based on Dörnyei and Taguchi (2010) and observed participants' current efforts, attitudes, and interest in studying English. A 4-point scale ranging from 1 (strongly disagree) to 4 (strongly agree) was used to prevent the participants from choosing a neutral option, as has been an issue in Japan (Wang et al., 2008).

The next part (13 questions, 4-point Likert scale) collected participants' attitudes towards playing games or watching gameplay to learn English and their perceptions of its effectiveness. The survey was derived from Bolliger et al. (2015) (α = .72) and Bourgonjon et al. (2010) (α > .70) whose questions were based on the Technology Acceptance Model (TAM) by Davis (1989). The questions included the four criteria of *ease of use*, *learning opportunities*, *experience*, and *preference* and were slightly modified to incorporate both playing digital games and watching gameplay.

The last two categories include two questions each pertaining to participants' subjective invested mental effort and perceived difficulty of their task. These questions were based on deHaan et al. (2010) (mental effort α = .551, material difficulty α = .565), who used the prior surveys of Kalyuga et al. (1998) (α = .4583) and Paas (1992) (α > .85). The question items are listed as follows:

Q1. How much mental effort did you invest in your assigned task (playing/watching)?

[9-point scale ranging from extremely low to extremely high mental effort]
Q2. How much mental effort did you put into learning English through your assigned task (playing/watching)?
[9-point scale ranging from extremely low to extremely high mental effort]
Q3. How difficult was your assigned task (playing/watching)?
[7-point scale ranging from extremely easy to extremely difficult]
Q4. How difficult was it to understand the English in the game?
[7-point Likert ranging from extremely easy to extremely difficult]

All categories except invested mental effort and perceived difficulty were administered prior to the experiment. Then the entire survey was conducted immediately after each of the five weeks of treatments. The resulting Cronbach's alphas are displayed in Table 1. As an overview, criterion measures' alpha for the combined immediate post-surveys was α =.79, attitudes and perceptions α =.87, mental effort α =.88, and perceived difficulty α =.74.

Table 1
Questionnaire reliability

	Cronbach's Alpha			
Session	Criterion Measures	Attitudes and Perceptions	Mental Effort	Perceived Difficulty
1	.71	.88	.86	.76
2	.78	.85	.92	.57
3	.82	.89	.88	.82
4	.85	.87	.87	.83
5	.78	.86	.80	.62
Total sessions	.79	.87	.88	.74

Interviews

The post-treatment interviews were conducted after the final session in either English or Japanese, and a native Japanese speaker with English fluency also participated to assist with translating. The interviews were structured on the questionnaire (see Appendix B for the guide used), which ensured the dialogue stayed relevant while also permitting participants to deviate and elaborate on their ideas and the researcher to pursue other lines of inquiry. The conversations were recorded with permission and transcribed and translated into English by a native Japanese speaker with English fluency. The transcripts were then coded by the author following open coding, where categories were initially created, followed by assigning titles, and finally arranging them into categories.

Results

Q1. What measurable effect does physical interactivity (playing versus watching) have on immediate vocabulary achievement for a real-time strategy game?

Table 2 shows the results of the vocabulary pre-check. The players and watchers had similar minimum and maximum scores and means, indicating that both groups shared similarities in their knowledge of the tested vocabulary.

Table 2
Descriptive statistics of the vocabulary pre-check

	N	*Min Score	*Max Score	M	SD
Players	15	15	41	27.40	7.39
Watchers	17	18	42	27.06	6.38

*participants responded with "yes" or "no" regarding knowing the vocabulary; the highest possible score = 50 after distracter words were omitted

The descriptive statistics for all combined immediate post-tests are displayed in Table 3. Overall, both groups experienced increases in their minimal and maximum scores and means. The standard deviations also decreased, indicating more consistent answers amongst each other.

When comparing the two groups, we see that the means were very similar, with players being slightly higher (M = 34.93, SD = 5.27) than the watchers (M = 34.18, SD = 3.25) but with a higher standard deviation. Players had a higher maximum score while watchers had a higher minimum score. Nevertheless, a paired sample t-test showed no statistical difference between the two groups [$t(14) = .496$, $p > 0.05$].

Table 3
Descriptive statistics of combined vocabulary post-tests

	N	Min Score	Max Score	M	SD
Players	15	28	46	34.93	5.27
Watchers	17	31	42	34.18	3.23

Accumulation of all five immediate post-tests; highest possible score = 50

Q2. What measurable effect does physical interactivity have on delayed vocabulary achievement for a real-time strategy game?

The descriptive information for the two-week delayed post-test is shown in Table 4, and this was compared to the accumulated results of the five prior post-tests (Table 3). The means for both the players and watches were higher on the 2-week delayed post-tests. This can be attributed to the participants possibly being exposed to the same vocabulary items throughout the five-week experiment after they were initially tested on it and thus could have learned the items. The players group increased their score while the watcher's score remained nearly unchanged. The means increased by 1.27 for the players and 0.26 for the watchers.

Additionally, the players' lowest score went up by three points, while the highest score went down by two points. Likewise, the watchers' lowest score went down by three points while the highest score went up by one point. Nevertheless, a paired-samples t-test showed no statistical difference between the groups [$t(14) = 1.077$, $p > 0.05$].

Table 4
Descriptive statistics of vocabulary 2-week delayed post-test

	N	Min Score	Max Score	M	SD
Players	15	31	44	36.2	4.43
Watchers	16	28	43	34.44	4.56

*highest possible score = 50

Q3. What are learners' motivation and attitudes toward playing or watching games for language-learning purposes, and what are their perceptions about the effectiveness for language learning before and after the treatment?

Table 5 displays the survey data for prior to the experiment and post final session, and the entire dataset for all sessions can be found in Appendix A. Players (M= 3.13 out of 4) and watchers (3.12) had similar pre-experiment Criterion Measures, indicating medium initial effort and motivation to study English. Both players and watchers' Criterion Measures stayed mostly the same throughout the experiment and returned to the original means by the fifth session, indicating no effect on participants' current effort and motivation to study English. A paired-samples t-test for Criterion Measures showed no statistical difference between players and watchers [$t(14) = .243$, $p > 0.05$].

Table 5
Descriptive statistics for surveys: prior to the experiment and final session

	Session	**Players**		**Watchers**	
		*M	SD	*M	SD
Criterion Measures	pre	3.13	0.74	3.12	0.84
	last (5th)	3.13	0.79	3.11	0.83
		0.00	0.05	-0.01	-0.01
Ease of Use	pre	2.57	1.14	2.74	0.90
	last (5th)	2.83	0.99	2.88	0.81
		0.27	-0.15	0.15	-0.09
Learning Opportunities	pre	2.96	0.83	2.96	0.79
	last (5th)	3.08	0.82	2.87	0.61
		0.12	-0.01	-0.09	-0.18

Experience	pre	2.03	1.07	1.91	0.75
	last (5th)	2.57	1.04	2.00	0.92
		0.53	-0.03	0.09	0.17
Preference	pre	2.78	0.85	2.76	0.77
	last (5th)	3.07	0.86	2.74	0.68
		0.28	0.02	-0.03	-0.09

**Max = 4*

note: the questions are worded to solicit answers specific to the assigned role (player or watcher).

Regarding the attitudes and perceptions items, first, *ease of use* increased for both the players (M = 2.83) and watchers (M = 2.88), although the watchers were higher and experienced a higher increase, albeit marginal. Additionally, a paired-samples t-test for *ease of use* showed no statistical difference [$t(14) = .520$, $p > 0.05$].

Second, players and watchers initially rated *learning opportunities* exactly even (M = 2.96) prior to the experiment. But by the final session, the players had a comparatively higher score (M = 3.08). A paired-samples t-test for *learning opportunities* showed a statistical difference [$t(14) = 3.118$, $p = 0.008$] between players and watchers, with players having a higher rating.

Next, *experience* started higher with players prior to the experiment (M = 2.03), went up for both groups over the course of the experiment, and ended higher for players (M= 2.57). A paired-samples t-test showed no statistical difference between the groups [$t(14) = 1.316$, $p > 0.05$].

Finally, *preference* started higher for players before the experiment (M = 2.78) and ended higher by the final session (M = 3.07). Furthermore, *preference* was shown to have a statistical difference [$t(14) = 2.314$, $p = 0.036$] between groups, with players rating higher than watchers.

Overall, besides *ease of use*, players expressed comparatively better overall attitudes and perceptions, especially with *learning opportunities* and *preference* where a statistical difference was established.

Q4 What effect does physical interactivity have on subjective invested mental effort (of the assigned task and learning English) and the perceived difficulty (of the assigned task and the game's language)?

The results of the invested mental effort and difficulty are displayed in Table 6 and Table 7. In all cases and for both players and watchers, the first session had the highest mental effort for the task and learning English, as well as task difficulty and English difficulty. When averaging all sessions together, the players (M = 3.33) had an overall higher mental effort in the task than the watchers (M = 3.28), while the watchers (M = 3.32) had a higher mental effort in learning the game's English compared to the players (M = 3.08). Paired-samples t-tests showed no statistical difference between either the task mental effort [$t(14) = .063$, $p > 0.05$] or English mental effort [$t(14) = .530$, $p > 0.05$].

Table 6
Descriptive statistics for invested mental effort

	Session	Players		Watchers	
		*M	SD	*M	SD
Task mental effort	1	4.33	1.95	3.94	1.56
	2	2.87	1.77	2.76	1.79
	3	3.20	1.97	2.88	2.15
	4	3.00	1.36	3.65	1.41
	5	3.27	1.28	3.18	1.51
English mental effort	1	4.27	1.98	4.12	1.62
	2	2.87	1.55	3.12	1.73
	3	2.93	1.69	2.82	1.67
	4	2.93	1.44	3.53	1.66
	5	2.40	1.12	3.00	1.27

**Max = 9*

Table 7
Descriptive statistics for task and language difficulty

	Session	Players		Watchers	
		*M	SD	*M	SD
Task difficulty	1	3.60	1.50	3.06	1.43
	2	2.93	1.39	2.47	0.94
	3	3.00	1.15	2.53	1.12
	4	2.87	0.99	3.00	1.22
	5	3.47	1.36	2.94	1.20
Language difficulty	1	4.40	1.55	4.06	1.14
	2	2.73	1.10	3.29	1.10
	3	3.00	1.37	2.76	1.15
	4	2.87	1.06	3.24	1.44
	5	3.13	1.13	3.12	1.05

**Max = 7*

Next, the players (M = 3.17) had a higher rated difficulty of the task than the watchers (M = 2.8). The difficulty of the game's English was almost equal, with watchers slightly higher (Players = 3.23, Watchers = 3.29). These results indicate that the players experienced more difficulty with playing the game, but the difficulty of the target learning was nearly equal. Nevertheless, paired-samples t-tests showed no statistical difference for the *task difficulty* [$t(14) = 1.058$, $p > 0.05$] or *language difficulty* [$t(14) = .317$, $p > 0.05$].

Interviews

The interviews and researcher observations provided additional explanation to the above quantitative data. Thirty of the thirty-two participants took part in the post-interview, which averaged approximately 17 minutes and 38 seconds and yielded the following trends.

First, most participants in both groups felt that their assigned task was at least somewhat effective for learning English. These opinions started out positive prior to the experiment and slightly increased by the end. Additionally, prior to the experiment, participants' motivation to learn English through their assigned task ranged from very motivated to somewhat motivated. This increased for most participants at the end of the experiment, while some were unchanged. One watcher's motivation went down after not feeling like there were learning gains.

Second, most participants agreed that physical interactivity (playing the game) required more mental effort and had a higher task difficulty. Nevertheless, several participants felt this helped with learning the target language since they were forced players to pay attention while watchers had no such obligation. Because of this, almost all interviewed players (12 out of 14) felt playing is more effective for language learning, stating comments such as, "If you are on the playing side, you can concentrate, and if you don't understand it, you can't move on. Even if the viewer doesn't understand it, once the player understands it, they can move on," and "Watchers don't give much thought to the English." The two players who felt watching is better both said that they were too busy concentrating on playing the game to learn the English and that watching would have allowed them to focus more on learning. The watchers' answers varied as some felt that they would have concentrated more if playing the game, while others felt that the game was too difficult and wouldn't have allowed them the ability to concentrate on learning English. One stated, "Up until the first or second week, it would have been nice to play. But it became fun to watch at the end. The game was getting harder and harder, so it was hard just to watch. So, I thought it would be even harder to play." Others commented that watching " allows me to focus on the English words without getting too involved in the game" and "If you are a watcher, you can concentrate on English, so I think it is effective for learning English." Therefore, participants expressed the benefits and disadvantages of both having physical interactivity as well as removing it, but there was more of a consensus for the players and playing over watching.

Next, there is an indication that the skill of the players may affect the overall experience and opportunities to retain the target learning of both the players and watchers, a trend also discovered in the preliminary study. Having low player skill may simultaneously force players' attention from the target learning to operate the game while generating watcher frustration through repetition and a slow pace. A watcher commented, "If the game itself doesn't progress at all, I get a little sleepy, so I think skill has an effect." Another stated, "I think you can have fun and concentrate by watching someone who is good at playing."

Lastly, some watchers expressed their disappointment in that they initially thought their role was going to more closely mimic the typical online watching experience rather than simply watching another person play. One watcher who often watches online competitive games explained, "In this experiment, ... the players were silent, and I learned only by looking at the English that appeared in the explanations in the game. But with YouTube, there are streamers who speak in English, and some give live commentary... I think it is easier to understand that

way." Another watcher who also watches gameplay online stated, "Before the experiment, I thought I would see something like online gameplay. I expected an English speaker to give commentary in English while I watched it. So, I thought I would learn a lot. But in actuality, I was just watching the screen of the other college student playing, so I thought it would be more effective if I just played it myself." Typically, online gameplay watching includes multiple features such as having a skillful streamer who entertains their audience and comments on the gameplay, as well as allowing live communicative interaction between the streamer and viewers in live broadcasts or the ability to pause and play back parts in prerecorded videos. This experiment eliminated such factors to control for physical interactivity, and adding them in may provide an advantage over playing games.

Discussion

When observing the results of the players and watchers together, we see that there were vocabulary gains (lower minimum and higher max scores and higher median scores) for both groups on the delayed post-test, providing support for the findings of Ebrahimzadeh (2017). Additionally, most participants reported feeling like their vocabulary at least slightly increased, and that they enjoyed participating in the experiment. Nevertheless, both groups rated their current effort in learning English and motivation to do so as mostly unchanged, indicating no effect in this regard. Those who expressed no interest in continuing learning English through their assigned role usually cited their lack of interest in games in general or felt that other means of learning were more effective.

Next, when comparing the two groups, the players had a slightly higher vocabulary score mean but a higher standard deviation. Additionally, the watchers maintained their vocabulary score average on the delayed post-tests while players were able to retain slightly more vocabulary, indicating that players performed marginally better on both the immediate and delayed post-tests. While this may provide some support for studies showing favorable results for physical interactivity and vocabulary recall in serious games (Ali Mohsen, 2016), no statistical difference could be confirmed like what was found in OTS music games (deHaan et al., 2010) and reflex games (deHaan & Kona, 2010), both of which have shown a statistical advantage for watching.

Additionally, this study has conflicting results from the preliminary study in that the watchers scored slightly higher on the post-test and had higher scores for attitudes and perceptions. This may be explained in that the preliminary study was a one-time treatment where the tutorial and practice time appeared insufficient for the players, as some struggled to grasp the core mechanics of the game. In turn, this could have influenced the score and experience of the players. Therefore, this 5-week experiment may have balanced this out for a more realistic representation. Despite the prework and several weeks of playing the game, the researcher noted some players still showed signs of low levels of game proficiency.

Several players also felt that with more hours of gameplay practice, their ability to focus on the target language learning would increase. And some watchers complained about the repetition of the gameplay due to the player's lack of skill to pursue objectives. Therefore, the skill of the player may affect the entertainment and learning opportunities of both the player and watcher. Nevertheless, most studies on physical interactivity do not address this.

In typical online game-watching scenarios unrelated to learning, viewers often watch streamers who can attract viewers by demonstrating exceptional gaming skills. Therefore, studies incorporating one-time treatments and games with higher learning curves may especially be affected by player skill.

For attitudes and perceptions, *ease of use* was rated higher for watchers, showing fewer boundaries to entry if participants wanted to learn English through watching gameplay. This is generally the case as other studies have indicated (Andersson, 2022; Andersson, 2023; Bolliger

et al., 2015). Nevertheless, the difference was not shown to be significant. On the other side, playing games was seen to have higher *learning opportunities* despite starting equal prior to the experiment, and a statistical difference was found. Thus, playing the game lead to better perceptions of opportunities for English learning than it did for watching the game. Additionally, *preference* was higher for players both initially and at the end of the experiment, and a statistical difference was seen in favor of the players. Therefore, watching gameplay was considered easier for independent learning, but playing had greater positivity for volition and language-learning potential.

Next, the results of invested mental effort and task difficulty showed a higher score for players, and most participants in both groups reiterated this in the interviews. Nevertheless, there was no statistically-significant difference between the groups. The score was greater in the first sessions as participants were probably getting used to playing the game, and thus, invested a relatively higher amount of mental effort and felt that the task was difficult. Yet, this did not appear to equate to adverse vocabulary results as shown in prior studies (deHaan et al., 2010; deHaan & Kona, 2010) or negative sentiment in the participant's perceived effectiveness. The participants elaborated on their feelings in this regard, indicating that despite watchers not having to worry about controlling the gameplay, only watching the game will make the watcher tune out and lose concentration on the target learning because they lack agency and are not obligated to progress the game. In regards to the effect of physical interactivity, two explanations can be offered. The first is that having a high mental effort and task difficulty did not negatively affect the target learning. The second explanation is that interactivity negatively impacted the target learning, but at the same time, watchers performed similarly because taking away physical interactivity also caused adverse effects. Evidence for the latter explanation can be found in the interviews where participants voiced both concerns. Adding a third group acting as a control in future studies may help to provide a more thorough answer.

The results of this study suggest that, when controlling for interactivity, simply watching a game can lead to difficulties in concentrating on the target learning. Nevertheless, as previously mentioned, online gameplay watching typically involves other features that may be inducive to learning and can offset the negative aspects of watching. Namely, online gameplay watching includes having a skilled player who comments on the gameplay and entertains and interacts with the streamers in live sessions. And in prerecorded sessions, viewers can easily pause the content to look up words as well as rewind parts for repeated listening, both of which were requested by participants of this study. These features can be investigated further regarding their merit for scaffold learning and motivation.

Finally, the genre of games may lead to different results when testing interactivity (deHaan et al., 2010). Games possess a wide range of levels of interactivity from simplistic, slow, and repetitive button-clicking to intense and complicated commands. It also may be possible to find an optimal genre or specific game suitable for playing or watching gameplay for language learning purposes. The number of OTS game genres tested for interactivity is limited and can be expanded upon in the future.

Conclusion

DGBLL has been investigated regarding the merits of playing games for language learning, and controlling for physical interactivity is a method to do so and has been done in a limited capacity and mostly for serious games. Due to the growth in popularity of online gameplay watching, the definition of what constitutes DGBLL applications may be expanded to also encompass gameplay watching. Then, it is also important to solicit evidence of COTS games for potential language-learning opportunities. This study demonstrates that playing games and watching gameplay may both possess merits and disadvantages for vocabulary

recall and attitudes and perceptions, and we cannot definitely say which is better. Further studies can provide additional evidence.

This study has several limitations that prevent it from making sweeping claims or broad generalizations. Particularly, as this was a longitudinal study, the remaining sample size for each group was not substantial after members were dropped. In turn, this may prohibit us from determining statistical significance with a high level of confidence. Therefore, future studies can include the incorporation of a larger sample size and attempt to employ methods of participant retention.

References

Admiraal, W., Huizenga, J., Akkerman, S., & Ten Dam, G. (2011). The concept of flow in collaborative game-based learning. *Computers in Human Behavior*, *27*(3), 1185–1194. https://doi.org/10.1016/j.chb.2010.12.013

Ali Mohsen, M. (2016). The use of computer-based simulation to aid comprehension and incidental vocabulary learning. *Journal of Educational Computing Research*, *54*(6), 863–884. https://doi.org/10.1177/0735633116639954

Andersson, S. (2022). Japan University EFL Students' Experience, Attitudes, and Perceived Effectiveness of Watching Gameplay for Language-learning Purposes. *JALT CALL Journal*. Volume 18(3), 412–443. https://doi.org/10.29140/jaltcall.v18n3.764

Andersson, S. (2023). Interactivity's Effect on Playing Versus Watching a Real-time Strategy Game: A Preliminary Experiment. *JALTCALL PCP*. 1–25. https://doi.org/10.37546/JALTSIG.CALL.PCP2022-01

Bolliger, D. U., Mills, D., White, J., & Kohyama, M. (2015). Japanese students' perceptions of digital game use for English-language learning in higher education. *Journal of Educational Computing Research*, *53*(3), 384–408. https://doi.org/10.1177/0735633115600806

Bourgonjon, J., Valcke, M., Soetaert, R., & Schellens, T. (2010). Students' perceptions about the use of video games in the classroom. *Computers & Education*, *54*(4), 1145–1156. https://doi.org/10.1016/j.compedu.2009.10.022

Bowen, T., & Marks, J. (2002). Inside teaching: Options for English language teachers. Oxford: Macmillan Heinemann.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, 94, 178–192. https://doi.org/10.1016/j.compedu.2015.11.003

Busse, V., & Krause, U. M. (2016). Instructional methods and languages in class: A comparison of two teaching approaches and two teaching languages in the field of intercultural learning. *Learning and Instruction*, *42*, 83–94. https://doi.org/10.1016/j.learninstruc.2016.01.006

Calvo-Ferrer, J. R. (2021). Effectiveness of Type of Feedback and Frequency on Digital Game-Based L2 Vocabulary Acquisition. *International Journal of Game-Based Learning (IJGBL)*, *11*(3), 38–55. https://rua.ua.es/dspace/bitstream/10045/116614/1/Calvo-Ferrer_2021_IntJGame-BasedLearn_final.pdf

Chang, C., Chang, C. K., & Shih, J. L. (2016). Motivational strategies in a mobile inquiry-based language learning setting. *System*, *59*, 100–115. https://doi.org/10.1016/j.system.2016.04.013

Chen, Y. C., Lu, Y. L., & Lien, C. J. (2021). Learning environments with different levels of technological engagement: a comparison of game-based, video-based, and traditional instruction on students' learning. *Interactive Learning Environments*, *29*(8), 1363–1379. https://doi.org/10.1080/10494820.2019.1628781

Cho, J. Y., Andersen, E., & Kizilcec, R. F. (2021). Delivery Ghost: Effects of Language Immersion and Interactivity in a Language Learning Game. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–7. https://dl.acm.org/doi/pdf/10.1145/3411763.3451767?casa_token=9q0bSxZ65MwAAAAA:FHmX0tMSeaBuxHQ82gh-QysAqdl5UI2rMsGcDcSI2wtBSsWoKJFrDSDjOFgSOYUcJ9mo2xX3epE8RA

Churchill, B. C., & Xu, W. (2016). The modem nation: A first study on Twitch. tv social structure and player/game relationships. In 2016 *IEEE international conferences on big*

data and cloud computing (BDCloud), social computing and networking (SocialCom), sustainable computing and communications (SustainCom)(BDCloud-SocialCom-SustainCom), 223–228. https://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.43

Clement, J. (2022, November 16). *Average number of concurrent viewers on Twitch from 2nd quarter 2018 to 3nd quarter 2022*. Statista. https://www.statista.com/statistics/761122/average-number-viewers-on-youtube-gaming-live-and-twitch/

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/249008

deHaan, J. (2005). Acquisition of Japanese as a foreign language through a baseball video game. *Foreign Language Annals*, *38*(2), 278–282. https://doi.org/10.1111/j.1944-9720.2005.tb02492.x

deHaan, J., & Kono, F. (2010). The effect of interactivity with WarioWare Minigames on second language vocabulary learning. *Journal of Digital Games Research*, *4*(2), 47–59.

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. https://www.lltjournal.org/item/543/

Diwanji, V., Reed, A., Ferchaud, A., Seibert, J., Weinbrecht, V., & Sellers, N. (2020). Don't just watch, join in: Exploring information behavior and copresence on Twitch. *Computers in Human Behavior*, *105*, 1–11. https://doi.org/10.1016/j.chb.2019.106221

Ebrahimzadeh, M. (2017). Readers, players, and watchers: EFL students' vocabulary acquisition through digital video games. *English Language Teaching*, *10*(2), 1–18. https://eric.ed.gov/?id=EJ1125224

Ebrahimzadeh, M., & Alavi, S. (2016). Motivating EFL students: E-learning enjoyment as a predictor of vocabulary learning through digital video games. *Cogent Education*, *3*(1), 1–14. https://doi.org/10.1080/2331186X.2016.1255400

Gilbert, B. (2018, February 14). *Amazon's streaming service Twitch is pulling in as many viewers as CNN and MSNBC*. Business insider. https://www.businessinsider.com/twitch-is-bigger-than-cnn-msnbc-2018-2

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: How effective are they? a meta-analysis of recent studies. *Journal of Computer Assisted Learning*, *29*(3), 207–219. 10.1111/j.1365- 2729.2012.00489.x

Gordillo, A., López-Fernández, D., & Tovar, E. (2022). Comparing the effectiveness of video-based learning and game-based learning using teacher-authored video games for online software engineering education. *IEEE Transactions on Education*, *65*(4), 524–532. 10.1109/TE.2022.3142688

Gros D., Wanner B., Hackenholt A., Zawadzki P., & Knautz K. (2017). World of streaming. Motivation and gratification on Twitch. In G. Meiselwitz (Ed.), *Proceddings of SCSM 2017: Social computing and social media. Human behavior*. (pp. 44-57). Springer. https://doi.org/10.1007/978-3-319-58559-8_5

Hamilton, W. A., Garretson, O., & Kerne, A. (2014). Streaming on Twitch: Fostering participatory communities of play within live mixed media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1315–1324). Association of Computing Machinery. https://doi.org/10.1145/2556288.2557048

Hsu, W. C., & Lin, H. C. K. (2016). Impact of applying WebGL technology to develop a web digital game-based learning system for computer programming course in flipped classroom. *Proceedings of the 2016 International Conference on Educational Innovation through Technology (EITT)* (pp. 64–69). IEEE. 10.1109/EITT.2016.20

Ijaz, K., Bogdanovych, A., & Trescak, T. (2017). Virtual worlds vs books and videos in history education. *Interactive Learning Environments*, *25*(7), 904–929. 10.1080/10494820.2016.1225099

Iqbal, Mansoor (2022, September 6). *Twitch Revenue and Usage Statistics (2022)*. Business of Apps. https://www.businessofapps.com/data/twitch-statistics/#:~:text=to%202021 %20(bn)-,Twitch%20age%20demographics,are%20aged%2055%20or%20above

Jabbari, N., & Eslami, Z. R. (2019). Second language learning in the context of massively multiplayer online games: A scoping review. *ReCALL*, *31*(1), 92–113. https://doi.org/10.1017/S0958344018000058

Jang, W. W., & Byon, K. K. (2019). Antecedents and consequence associated with esports gameplay. *International Journal of Sports Marketing and Sponsorship*, *21*(1), 1–22. https://doi.org/10.1108/IJSMS-01-2019-0013

Johnsen, H. M., Briseid, H. S., Brodtkorb, K., Slettebø, Å., & Fossum, M. (2021). Nursing students' perceptions of combining hands-on simulation with simulated patients and a serious game in preparing for clinical placement in home healthcare: a qualitative study. *Nurse Education Today*, *97*, 104675. https://doi.org/10.1016/j.nedt.2020.104675

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human factors*, *40*(1), 1-17. https://doi.org/10.1518/001872098779480587

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, *13*(4), 351–372. https://doi.org/10.1002/(SICI)1099-0720(199908)13:4<351::AID-ACP589>3.0.CO;2-6

Kaytoue, M., Silva, A., Cerf, L., Meira, Jr., W., & Raïssi, C. (2012). Watch me playing, I am a professional: a first study on video game live streaming. *Proceedings of the 21st International Conference on World Wide Web* (pp. 1181–1188). Association for Computing Machinery. https://doi.org/10.1145/2187980.2188259

Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education*, *55*(2), 427–443. https://doi.org/10.1016/j.compedu.2010.02.007

Li, Y., Wang, C., & Liu, J. (2020). A systematic review of literature on user behavior in video game live streaming. *International Journal of Environmental Research and Public Health*, *17*(9), 1168–1177. https://doi.org/10.3390/ijerph17093328

Krashen, S. (1981). Second language acquisition. *Second Language Learning*, *3*(7), 19–39.

Lightbown, P. M., & Spada, N. (2013). *How languages are learned 4th edition-Oxford Handbooks for Language Teachers*. Oxford university press.

Manchon, R. M., Murphy, L., & de Larios, J. R. (2007). Lexical retrieval processes and strategies in second language writing: A synthesis of empirical research. *International Journal of English Studies*, *7*(2), 149–174.

Nakamura, J., & Csikszentmihalyi, M. (2009). Flow theory and research. In C. R. Snyder & S. J. Lopez (Eds.), *Oxford handbook of positive psychology* (2nd ed., 195–206). Oxford University Press.

Nation, I. S. (2001). *Learning vocabulary in another language* (Vol. 10). Cambridge: Cambridge university press.

Orme, S. (2021). "Just watching": A qualitative analysis of non-players' motivations for video game spectatorship. *New Media & Society*, *24*(10), 2252–2269 https://doi.org/10.1177/1461444821989350

Parmawati, A., & Inayah, R. (2019). Improving students' speaking skill through English movie in scope of speaking for general communication. *Eltin Journal: Journal of English Language Teaching in Indonesia*, *7*(2), 43–53.

Pellouchoud, E., Smith, M. E., McEvoy, L., & Gevins, A. (1999). Mental effort related EEG modulation during video game play: Comparison between juvenile epileptic and normal control subjects. *Epilepsia*, *40*(4), 38–43. https://doi.org/10.1111/j.1528-1157.1999.tb00905.x

Peterson, M. (2013). Computer games and language learning. Palgrave Macmillan.

Peterson, M. (2016). The use of massively multiplayer online role-playing games in CALL: An analysis of research. *Computer Assisted Language Learning*, *29*(7), 1181–1194. https://doi.org/10.1080/09588221.2016.1197949

Peterson, M. (2021). Digital simulation games in CALL: a research review. *Computer Assisted Language Learning*, 1–24. https://doi.org/10.1080/09588221.2021.1954954

Plass, J. L., & Jones, L. C. (2005). Multimedia learning in second language acquisition. In R. E. Mayer (Ed.), *Cambridge handbook on multimedia learning* (pp. 467–488). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.030

Polman, H., De Castro, B. O., & van Aken, M. A. (2008). Experimental study of the differential effects of playing versus watching violent video games on children's aggressive behavior. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *34*(3), 256–264. https://doi.org/10.1002/ab.20245

Read, G. L., Lynch, T., & Matthews, N. L. (2018). Increased cognitive load during video game play reduces rape myth acceptance and hostile sexism after exposure to sexualized female avatars. *Sex Roles*, *79*(11), 683–698. https://doi.org/10.1007/s11199-018-0905-9

Scerbakov, A., Pirker, J., & Kappe, F. (2022). When a pandemic enters the game: The initial and prolonged impact of the COVID-19 pandemic on live-stream broadcasters on Twitch. *Proceedings of the Hawaii International Conference on System Sciences HICSS 2022* (pp. 1–10). University of Hawaii. http://jpirker.com/wp-content/uploads/2022/01/paper_final_manuscript.pdf

Sims, R. (1997). Interactivity: A forgotten art? *Computers in Human Behavior*, *13*, 157–180. https://doi.org/10.1016/S0747-5632(97)00004-6

Sjöblom, M., Törhönen, M., Hamari, J., & Macey, J. (2017). Content structure is king: An empirical study on gratifications, game genres and content type on Twitch. *Computers in Human Behavior*, *73*, 161–171. https://doi.org/10.1016/j.chb.2017.03.036

Sony Life Insurance Corporation. (2021, July 29). *Chūkōsei ga omoiegaku shōrai ni tsuite no ishiki chōsa [Awareness survey about the future envisioned by junior and senior high school students]*. *Sony Life Insurance Corporation*. https://www.sonylife.co.jp/company/news/2021/nr_210729.html

Sun, L., Ruokamo, H., Siklander, P., Li, B., & Devlin, K. (2021). Primary school students' perceptions of scaffolding in digital game-based learning in mathematics. *Learning, Culture and Social Interaction*, *28*, 100457. https://doi.org/10.1016/j.lcsi.2020.100457

Sweetser, P., & Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE)*, 3(3), 1–24. https://doi.org/10.1145/1077246.1077253

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Taylor, L. (2002). *Video games: Perspective, point-of-view, and immersion* [Master's thesis, University of Florida, USA]. Semantic Scholar. http://purl.fcla.edu/fcla/etd/UFE1000166

Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second language research*, 14(2), 103-135.

Vanderplank, R. (2019). 'Gist watching can only take you so far': Attitudes, strategies and changes in behaviour in watching films with captions. *The Language Learning Journal*, *47*(4), 407–423. https://doi.org/10.1080/09571736.2019.1610033

Walter, N., & Tsfati, Y. (2016). Interactive experience and identification as predictors of attributing responsibility in video games. *Journal of Media Psychology*, *30*(1), 1–32. https://doi.org/10.1027/1864-1105/a000168

Wang, S. Y., Chang, S. C., Hwang, G. J., & Chen, P. Y. (2018). A microworld-based role-playing game development approach to engaging students in interactive, enjoyable, and effective mathematics learning. *Interactive Learning Environments*, *26*(3), 411–423. https://doi.org/10.1080/10494820.2017.1337038

Wang, R., Hempton, B., Dugan, J. P., & Komives, S. R. (2008). Cultural differences: Why do Asians avoid extreme responses? *Survey Practice*, *1*(3), 1–8. https://doi.org/10.29115/SP-2008-0011

Whitton, N. (2014). *Digital games and learning: Research and theory* (1st ed.). Routledge.

Appendix A

Descriptive statistics for surveys (all sessions)

	Session	**Players**		**Watchers**	
		*M	SD	*M	SD
Criterion Measures	pre	3.13	0.74	3.12	0.84
	1	3.04	0.76	3.00	0.77
	2	3.05	0.73	3.14	0.80
	3	3.00	0.82	3.22	0.79
	4	3.13	0.78	3.09	0.84
	5	3.13	0.79	3.11	0.83
Ease of Use	pre	2.57	1.14	2.74	0.90
	1	2.53	0.82	2.76	0.85
	2	2.87	0.82	2.91	0.71
	3	2.80	0.92	2.94	0.89
	4	2.67	0.96	3.03	0.87
	5	2.83	0.99	2.88	0.81
Learning Opportunities	pre	2.96	0.83	2.96	0.79
	1	3.13	0.78	2.67	0.76
	2	3.25	0.66	2.75	0.67
	3	3.20	0.77	2.99	0.78
	4	3.20	0.81	2.92	0.73
	5	3.08	0.82	2.87	0.61
Experience	pre	2.03	1.07	1.91	0.75
	1	2.20	1.00	1.94	0.85
	2	2.43	1.07	2.06	0.81
	3	2.30	1.02	2.09	0.97
	4	2.33	1.09	2.15	1.05
	5	2.57	1.04	2.00	0.92
Preference	pre	2.78	0.85	2.76	0.77
	1	3.12	0.78	2.54	0.82
	2	3.17	0.81	2.69	0.58
	3	3.18	0.77	2.81	0.80
	4	3.07	0.84	2.69	0.72
	5	3.07	0.86	2.74	0.68

**Max = 4*

note: the questions are worded to solicit answers specific to the assigned role (player or watcher).

Appendix B

Interview questions template

Overall experience and preference 感想、印象

- Please describe your overall experience. Did your experience change over the 5 sessions?
- Do you feel your English level improved?
- If given a choice, which task (play/watch) would you have liked to have been assigned? Why?
- Do you play video games or watch gameplay?
- What is your usual way of studying English?

Learning opportunities (perceived effectiveness) 学習の機会

- Before the experiment, how effective did you think your assigned task (playing/watching) was for learning English? Has this changed?
- Which (play/watch) do you think is more effective for learning English?

Ease of use アクセシビリティ

- If you were to continue doing this method (play/watch) by yourself to learn English, how difficult would it be (e.g. acquiring equipment, setting it up)?

Motivation and attitudes 動機と態度

- Before the experiment, how motivated were you to try your assigned method (playing/watching) to learn English? Did your opinion change?
- How motivated are you to continue learning English this way?
- Do you plan on continuing to study English this way?

Perceived difficulty and mental effort 知覚された困難と精神的努力

- How difficult did the game itself seem (unrelated to language learning)?
- How difficult was your assigned task (watch/play)?
- How difficult was it to learn English while doing your task?
- How much of your mental concentration was used on performing your task (watch/play) versus learning English?
- Which task (play/watch) do you think would generally demand more of your mental effort?
- For watchers: How was the skill of the player? Did this affect your experience?
- For players: How was your perceived level of skill in the beginning and after 5 sessions? Did this change your learning experience?

Do you have any other comments? ほかにコメントはありますか。