

Classroom Intervention for Integrating Simulation Games into Language Classrooms: An Exploratory Study with the SIMs 4

Judy (Qiao) Wang (judywang.jp@gmail.com)
Kyoto University, Japan

Abstract

This study explored three forms of classroom intervention: teacher instruction, peer interaction and in-class activities, for the purpose of integrating simulation games into a vocabulary-focused English classroom. The aim was to establish which intervention is most effective, as well as what improvements should be made for future application. The study took the form of a controlled experiment and evaluation of the interventions was based on concurrently collected quantitative and qualitative data. The researcher concluded that while quantitative data failed to confirm any statistical significance between the two groups, qualitative data suggested two forms of intervention, teacher instruction and in-class activities, were effective. Peer interaction, however, did little to promote vocabulary acquisition. The researcher proposes implementing more diversified in-class activities and game quests relating to curriculum goals in existing classroom interventions. The discussion concludes by highlighting promising areas for future research.

Keywords: game-based learning; language classroom; computer simulation games; classroom intervention; vocabulary acquisition; *Sims 4*

Introduction

In his book *What Video Games Have to Teach Us about Learning and Literacy*, Gee (2007), one of the best recognized researchers in video games, presented 36 principles of learning applied in video games and discussed how cognitive science-supported games enhanced learning. Current studies on computer game-based learning (GBL) in language acquisition, an emerging stratum in computer-assisted language learning, also base themselves on learning theories and draw on psycholinguistic, cognitive, and sociocultural rationales. Researchers have concluded that GBL increases learner motivation (Dickey, 2007; Papastergiou, 2009), reduces learning anxiety (Kiili, 2005; Hwang, Hsu, Lai, & Hsueh, 2017), exposes learners to rich sources of target language input (Berns, Gonzalez-Pardo, & Camacho, 2013), and provides access to conditions conducive to peer learning (Sylvén & Sundqvist, 2012). In an expanding body of GBL research, simulation games have been a major subject of investigation (Peterson, 2010). These environments offer highly engaging virtual reality-based simulations where players are presented with language learning opportunities through their exposure to game content and, if the game is network-based, through online socialization (Peterson, 2012). However, as commercial off-the-shelf (COTS) games are not designed specifically for educational purposes (Yudintseva, 2015), their use is mostly limited to incidental learning.

Therefore, there exists one major issue in GBL: pedagogical relevance. To address this issue, research on the integration of simulation games into formal language classrooms is necessary. In particular, investigation of different classroom interventions in GBL represents an area of major interest.

Background

Computer-based simulation games and language classrooms

The application of computer-based games in language classrooms has been widely investigated. In a study conducted in Taiwan, Wu, Chen, and Huang (2014) utilized a digital board game in a language classroom and discovered that the game enhanced performance and contributed to an immersive environment that fostered communication. Similarly, Ebrahimzadeh and Sepideh (2017) found that video games significantly enhanced the language learning motivations of high school students in Iran. There is substantial evidence in the literature supporting the application of computer-based games as a means to enhance aspects of language learning (see for example Reinders & Wattana, 2015; Alyaz, Spaniel-Weise, & Gursay, 2017).

Computer-based simulation games (henceforth simulation games) are a subcategory of games that have attracted the attention of language classroom researchers. Fletcher, Tobias and Wisher (2007) defined a computer-based simulation game as a piece of entertainment computer software that is reality-based, goal-focused and interactive. They further claim that in this type of game decisions must be made and reactions to decisions are discernable. Such games are believed to provide content for language learning that is “naturally rich in associations” through exposure to meaningful contexts (Purushotma, 2005, p.84). In an article summarizing CALL software, Healey (1999) proposed the use of the simulation game *SimCity* to teach students reading in an English classroom. In support of Healey’s work, Schwienhorst (2002) discussed autonomous and interactive learning opportunities offered by simulation games. Furthermore, it was suggested that vocabulary acquisition was facilitated by the video game’s repetitive, highly contextualized, and simultaneously presented textual language (Calvo-Ferrer, 2017). In particular, these input-rich environments facilitate incidental vocabulary learning, where learners guess the meaning of a word through contextual clues, albeit at a rather slow rate (Coady, 1993). In a study on the effect of digital games on Iranian children’s English vocabulary learning, Aghlara and Tamjid (2011) concluded that the mean score of children who learned vocabulary through a digital game was significantly higher than those who learnt without the game. DeHaan (2005) conducted a one-month study to investigate how one intermediate Japanese-as-a-foreign-language student improved his listening and vocabulary in terms of *kanji* character recognition by playing a baseball video game. Miller and Hegelheimer (2006) incorporated an authentic simulation game, *the Sims*, into their language classroom and found that the use of supplementary materials enhanced vocabulary learning. Ranalli (2008) used the same game in his classroom and utilized Miller and Hegelheimer’s work by focusing on task appropriateness in order to enhance the pedagogical benefits of supplementary materials.

Although research on the use of simulation games has produced broadly positive results, researchers caution that the pedagogical benefits are not guaranteed (Anderson, Reynolds, Yeh, & Huang, 2008). Most studies on simulation games situate themselves in informal learning contexts and a recent research review by Peterson, Wang and Mirzaei (2018) revealed that only five case studies (Canto, de Graff, & Jauregi, 2014; Chen, 2016; Deutschmann & Panichi, 2009; Kruk, 2015; Wigham & Chanier, 2013) have investigated integration of game activities into regular courses. In a noteworthy dissertation Egenfeldt-Nielsen (2005), concluded that more classroom-based research is necessary in order to establish the most effective application of computer games in educational settings. Moreover, it is noted in the literature that when integrating games into formal educational contexts, teachers have a key role to play in securing beneficial outcomes (Peterson, 2016). In this context, Squire (2005) described several sessions using simulation games which highlighted the considerable effort required on the part of the teaching staff to use the game effectively in line with the curriculum. As this discussion shows, more research focusing on classroom intervention is required as such an effort generates evidence that can persuade teachers to implement simulation games in their classrooms.

Classroom intervention

Interventions are change strategies purposively implemented at the individual, family, group, organizational, community and societal levels (Fraser & Galinsky, 2010). In educational settings, classroom intervention is comprised of interventions with class participants as intervention agents and therefore should be subject to unique adaptations by individual teachers and students according to the exigencies of their own curricula, values, and beliefs (Randi & Corno, 1997). Such interventions in language classrooms can be theory-based or practice-based, ranging from teacher's corrective feedback to student discourse, from the arrangement of classroom activities to classroom management and disciplines. Nagahashi (2007) conducted an intervention study among freshman students in an English for academic purposes course in Japan. He found that classroom intervention by means of the teacher arranging cooperative learning activities was effective in reducing language anxiety by providing a non-threatening, supportive environment that helped develop language skills. Jean and Simard (2011) conducted a descriptive observational study aimed at exploring the form-focused instruction interventions used by four French and four English high school teachers in classrooms. They discovered that grammar-oriented interventions were rather frequent in both contexts while some differences in preferences existed. Under a circumstance where pronunciation teaching in Japan depends heavily on decontextualized practice such as mechanical drills and repetition, Saito (2012) investigated the intervention of teacher instruction in students' pronunciation improvement and found significant improvement resulting from instruction. In intervention research conducted in an EFL classroom in Turkey, Bush (2015) concluded that games were an effective intervention in learner motivation although statistical results showed no significant difference between the experiment group and the control group.

There are also several studies in GBL that highlight measures resembling classroom interventions, such as teacher-produced scaffolds in the form of supplementary materials based on game content (Coleman, 2002; Ranalli, 2008), but there is still ample room to systematically investigate classroom interventions that make simulation games

suitable for formal classroom instruction. In light of this, this study aims to explore interventions that have the potential to increase the pedagogical relevance of simulation games. This study is exploratory in nature and serves as part of a larger project to identify effective and generalizable classroom interventions which teachers can apply to their classrooms with varying local needs.

As such, the researcher proposes the following two research questions:

1. Do the interventions selected for this study promote the integration of the simulation game into the language classroom?
2. Are there any improvements to the classroom interventions?

Methodology

The study took place in an English communication class at freshman level in a premier Japanese university and the curriculum goal was the acquisition of daily vocabulary as part of cognitive competence in English communication. The researcher blended and evaluated existing GBL studies that are related to, but not necessarily focused on, integration of games into language curriculum (see for example Miller and Hegelheimer, 2006; Ranalli, 2008), summarized the measures, and profiled these measures against the configuration of this study. The result was three forms of classroom intervention: teacher instruction, peer interaction and in-class activities. Teacher instruction was in English and consisted of two components: group-oriented instruction before gameplay where the teacher instructed the whole class in game vocabulary and gaming skills, and teacher-student interaction during gameplay where the teacher walked around and resolved specific questions from each pair. Peer interaction included communication in both first and second languages, to ensure that the students would feel comfortable in interaction, but of course, the latter was strongly encouraged. Regarding in-class activities, participants were given game quests or tasks they need to complete in the game. They were also required to record interesting gameplay episodes for presentations later. The presentations were expected to solicit the use of game vocabulary in a meaningful context and encourage participants to interact with the audience in Q&A. To evaluate these interventions both qualitatively and quantitatively, the researcher decided to conduct a controlled experiment where another group of students would play the game outside the classroom without the interventions in the control group. In the experimental group which was situated in the class, the researcher would also assume the role of the teacher.

Participants

This study consisted of three groups of participants: a baseline group, an experimental group, and a control group. Participants of the baseline group were all native Chinese first-year students randomly selected from a tier 2 university (ranking 100-200 nationwide) in China. They regularly took compulsory intensive and extensive English reading classes twice a week. Before the controlled experiment started, the researcher asked the baseline group, consisting of 10 participants, to take vocabulary quizzes

designed for this study online without playing the game. The purpose of the baseline group was to give a general idea of how random university students would perform on the quizzes. Apart from that, the baseline group was not involved in any part of the controlled experiment. The experimental group included 12 first-year education majors who were taking the English communication course once a week in a Japanese university as part of the requirement for obtaining a teacher's license in Japan. The students in the experimental group were of different English learning backgrounds and varying levels of English proficiency: one student came from Norway speaking fluent English and three Japanese students had overseas experience in English speaking countries. The Japanese natives were in the B1 to B2 (intermediate) competence range under CEFR (Common European Framework of Reference for Languages: Learning, Teaching, Assessment, n.d.) according to English test scores they reported in the pre-questionnaire (TOEFL, IELTS, TOEIC or EIKEN, a Japanese national English test). Lastly, the control group was situated in an out-of-class environment in China and included 12 Chinese first-year students in the same university as the baseline group. They were selected volunteer students who reported intermediate competence range under CEFR according to scores from their national college entrance exam. None of them had overseas experience. For both the experimental and control groups, English was a compulsory course for general education and none of the participants were majoring in English for academic purposes. In terms of gaming experience, the students in the control group were more familiar with gaming environments, with all reporting having played games on PCs and mobile phones. Five of them were frequent gamers but none of them played the game used in this study before. In the experimental group, only one student said he frequently played games in spare time. However, two students said they had previous experience playing earlier versions of the game used in the study. The most preferred gaming environment of the experimental group was the mobile phone, while the PC the most used platform in the control group.

Game selection

When selecting the simulation game for this study, the researcher first referred to the six criteria for CALL software selection proposed by Chappelle (2001), namely language learning potential, learner fit, meaning focus, authenticity, positive impact, and practicality. Naturally, simulation games provide authentic learning environments and have been proven to have a positive impact on students' learning motivation (see for example Garris, Ahlers, & Driskell, 2002; Sitzmann, 2011). The course of the experimental group was designed to improve students' daily communication from the aspect of vocabulary, so an ideal game would be one that has vocabulary acquisition potential, fits learners in this study, focuses on meaning instead of forms and runs on the laptops in the research lab. Although research on network-based simulation worlds is becoming increasingly popular (Peterson, 2017; Sadler, 2012), the unpredictability of target language (L2) exposure from socialization in such games also renders peer interaction uncontrollable and thus unsuitable for this vocabulary-focused research. Thus, games with more predictable and controllable contents that nevertheless do not compromise literature-butressed advantages of higher motivation should be explored. A stand-alone simulation, for instance, is a good choice given its engaging experience and

definitive L2 input from the game itself. From there, the researcher searched online a list of most rated standalone simulation games ("The Best Life Simulation Games of All Time", n.d.), and the game *Sims 4* (n.d.) by Electronic Arts emerged from a list of candidates. The *Sims 4* is a life simulation game, the fourth major title in the series *The Sims*. Players create characters called "Sims" and assign them personality traits that influence their daily "whims", wishes or goals the characters attempt to achieve. They control the lives of Sims to explore different possibilities in a simulated world, where everything works just as in reality. Players need to take care of their characters by having them eat, shower, sleep, socialize, use the bathroom and have fun. With various game expansion packs, features such as seasonal activities and celebrations, tourist resorts, club gathering, dining at restaurants and opening up businesses are also available. This game provides players exposure to vocabulary mainly through their characters' interaction with objects and other characters in the world. Figure 1 below shows an example of such interaction.



Figure 1. Exposure to vocabulary in the game

There are two studies on the previous series of *The Sims* (see Miller & Hegelheimer, 2006; Ranalli, 2008), both of which, however, were conducted more than 10 years ago. With the development of games and language, it is worth reinvestigating the potential of this game in language classrooms.

Table 1 shows the evaluation of *Sims 4* against the six criteria proposed by Chapelle.

Table 1.
Evaluation of *SIMS 4* against Chapelle's criteria

Criteria	Classroom features	Sims 4 features
----------	--------------------	-----------------

Language learning potential	Vocabulary acquisition as the curriculum goal	Rich daily vocabulary
Learner fit	Twelve female and male students of varying levels of English	Suitable for both genders; intuitive user interface; non-academic; inter-mediate
Meaning focus	Focus on meaning	Focus on meaning
Authenticity	Daily communication	Daily life scenarios
Positive impact	Vocabulary acquisition; positive student attitudes	Game-based vocabulary acquisition; higher motivation
Practicality	Laptops of medium-level hardware	Compatible with average PCs

Research design

Before the experiment started, the researcher invited 10 university students from China, the baseline group, to complete all the vocabulary quizzes. Their results would provide a reference to the consistency among all quizzes so that the researcher would know whether in-group fluctuations of scores should be attributed to varied levels of test difficulty. For evaluating the interventions as treatment, the experimental group were divided into pairs for enhanced peer interaction, received instruction from the teacher before and during gameplay, and completed structured in-class activities; whereas the control group played the game alone outside of a classroom environment. Both the experimental and control groups were introduced to the basics of gameplay at the beginning and took regular vocabulary tests as an assessment of their learning results. In the experimental group, there were 11 weekly sessions (every 1.5 hours) that entailed 4 themed units of gameplay. These sessions consisted of teacher instruction (around 100 minutes), participants' autonomous gameplay (around 10 hours for each participant), presentations (around 120 minutes for the whole class, with individual presentations lasting 1-5 minutes) and vocabulary tests and surveys (around 40 minutes for each participant). Participants in the control group were asked weekly to submit automatically generated "save data" and to indicate from and to what time they were playing the game in the file name. Each of them submitted 11 files that added up to 16.5 hours, exactly the total duration of the 11 sessions in the experimental group. Refer to Appendix 1 for a general view of sessional themes and activities in both groups.

Data collection

In this research, data from the experimental group came from three sources: surveys, recorded gameplay sessions and vocabulary quizzes. Data from the control group consisted solely of vocabulary quizzes.

Vocabulary quizzes. The researcher administered 5 quizzes to evaluate students' vocabulary acquisition: a pre-test and four unit tests, each with 50 words. The vocabulary base for the tests was extracted from the game language pack and profiled against the most frequent 1000 word families, the second 1000 and the Academic Word List (Vocabularyprofile English, n.d.). Then the researcher reviewed the words one by one to finalize a pool that contained words mostly likely to be encountered in gameplay. The pre-test was conducted in the introductory session before students began to play the game. The test consisted of randomly selected words from the pool to provide a general idea of students' vocabulary level against the game environment. The four unit tests included theme-specific words the researcher selected from the pool. They were conducted one week later, instead of immediately after the completion of a certain theme. By doing so, delayed investigation of students' retention of words was allowed and thus there was no post-test.

Recordings. Recordings in the experimental group were screen recordings of the game that also captured environment sounds so that interaction during gameplay and in-class activities were also included in audio format (the students were reluctant to show their faces, so no videos of activities were recorded). The researcher used the screen recording software, Bandicam (n.d.), to record full-time sessions of gameplay by the students. They provided valuable qualitative data regarding how students acquired and used vocabulary through gameplay and how effective the three interventions were respectively. A total of 72 hours of recordings from 6 pairs during a span of 10 sessions were examined in this study.

Surveys. Surveys included a pre-questionnaire, a post-questionnaire, and interviews for the experimental group, and only a pre-questionnaire for the control group. The pre-questionnaire focused on students' English background and their attitudes towards GBL and was also the base for selecting similar-level students in the control group. The post-questionnaire focused on the perceived learning experience and the influence of GBL on learning outcomes of the experimental group. Follow-up interviews were administered based on the post-questionnaire to acquire more in-depth qualitative information. Refer to Appendix 2 and 3 for the two questionnaires.

Data Analysis

Vocabulary quizzes. In these quizzes, participants from the three groups reported whether they knew and how much they knew a certain word. Two points were credited for clearly knowing a word, including its meaning and usage; 1 for a vague impression of a word; and 0 for not knowing a word at all. In the first two scenarios, participants were also required to provide the meaning of the word with which the researcher double-checked their final scores to ensure the accuracy of self-reporting. To ensure integrity, all participants were told that the test results would only be used for research and would have no bearings on their credit or final grading. The following are examples of the three scenarios:

	I know this word/phrase clearly	I vaguely know this word/phrase	I don't know this word/phrase at all
career	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Meaning of the word	<input type="text" value="Work path"/>		
retail	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Please explain	<input type="text" value="Where to sell some goods"/>		
stunning	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Please explain	<input type="text"/>		

Figure 2. Three scenarios of students' responses

Recordings. The researcher used a closed coding scheme, or what is more popularly known as concept-driven analysis, to analyze the recordings. Working in a concept-driven way means basing the categories, or codes, on previous knowledge: a theory, prior research, every knowledge, logic, or an interview guide (Schreier, 2012). Also, according to Crabtree and Miller (1999), there are several approaches to coding. On one hand, researchers can rely on predefined or priori codes, generally based on understandings from prior research or theoretical considerations. On the other hand, the researcher can develop codes only after some initial exploration of the data has taken place. The latter is also referred to as the open coding scheme. The reason why the researcher chose a closed coding scheme is that she is a teacher-researcher who had observed the classroom before analyzing the recordings. Therefore, rather than exploring the recordings with the open coding scheme, she predetermined a set of codes based on previous knowledge: classroom observation and the research focus (see Table 2). Transcripts typical of certain audible codes were yielded as examples.

Table 2.
Coding schemes of recordings

Code	Description
PIJ	Peer interaction in Japanese
PIE	Peer interaction in English
TSIE	Teacher-student interaction in English
GD	Gameplay development

ICA	In-class activities
-----	---------------------

Surveys. The pre-questionnaire focused on the background of students and was analyzed in Participants. The post-questionnaire was quantitative-dominated and focused on students' perceptions of the class. From the questionnaire, the researcher identified salient patterns and peculiarities in students' responses, both of which formed the basis for the follow-up interview. The researcher then took field notes during the interview and categorized individual data under the corresponding salient patterns and peculiarities according to students' post-questionnaire responses.

Results

Quantitative results

The quantitative results of this research came from vocabulary quizzes. Figure 3 is a box plot showing the quiz results of three groups: the baseline group, the experimental group, and the control group.

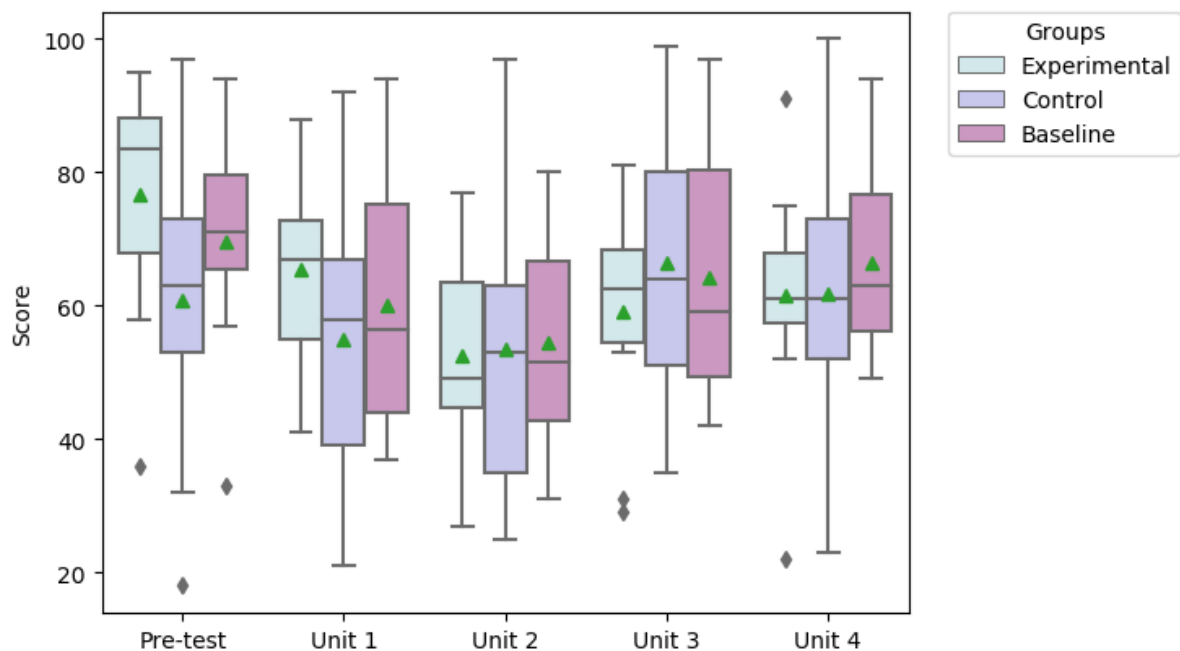


Figure 3. Quiz results of the experimental, control and baseline groups (full score: 100 points)

In the box plot, the green triangular markers represent mean scores; and horizontal bars in the middle of the boxes, median scores. The boxes along with their whiskers at both ends show the distribution of data in three parts. The bottom edge of the box represents the first quartile (Q_1), which is the median score of the lower half of the scores;

and the top edge of the box represents the third quartile (Q_3), which is the median score of the upper half of the scores. Gray diamond markers outside the boxes represent outliers, or extreme values, that lie more than one and a half times the length of a box from either end of the box.

It can be seen from the boxes of the baseline group that the quizzes were of varying difficulty levels, with the pre-test being obviously easier than other unit tests (see Appendix 4 for the pre-test). In addition, the dynamics of the boxes of the baseline group roughly matched those of the other two groups. Therefore, in-group fluctuation of scores should be attributed to varying levels of test difficulty, rather than to experiment treatment. In terms of outliers, the experimental group produced more outliers than the control group, especially those in the lower part of the box.

At a glance, it is apparent that the control group did not perform so well as the experimental group at the beginning, but soon caught up with and even exceeded the experimental group. This finding did not conform to the researcher's expectation that the experimental group would perform better all the way through the treatment. To better understand this phenomenon, the researcher also conducted a 2-tailed t-test to determine whether such differences were statistically significant. The sample sizes of the control group (N_1) and experimental group (N_2) are both 12, and the total sample size (N) is 24. As this sample size is less than 30, it means that the distribution of these samples is not normal, and therefore it is more appropriate to use the t-test instead of the z-test. In the t-test, there are two kinds of hypotheses, the null hypothesis (H_0) and the alternative hypothesis (H_1). The alternative hypothesis assumes that some differences exist between two groups in comparison, whereas the null hypothesis assumes that no difference exists. In this study the null hypothesis (H_0) is set to be that the experimental group and the control group achieve the same scores in the vocabulary quizzes; and the alternative hypothesis (H_1), that the experimental group achieves higher or lower scores than the control group in the vocabulary quizzes. Table 3 shows average scores, t-values, and p-values from the analysis. The mean of the t-values is -0.45, while the standard deviation of the t-values is 1.22

Table 3. P-values of the control and experimental groups

Quizzes	Average (Control)	Average (Experimental)	T-value	P-value
Pre-test	60.77	76.58	-2.01	0.06
Unit1	55.08	65.58	-1.38	0.18
Unit2	53.38	52.50	0.12	0.91
Unit3	66.38	59.17	0.99	0.33
Unit4	61.85	61.67	0.02	0.98

Of all the P values, none was less than 0.05, the cutoff level for statistical significance, thus the null hypothesis (H_0) could not be rejected. This means that, although the experimental group and control group exhibited differences in their test scores, the

differences were not statistically significant and therefore it cannot be concluded that experiment treatment, in any quantitative sense, influenced vocabulary acquisition in this study.

Qualitative results

Qualitative data consisted of gameplay recordings from the experimental group, a post-questionnaire, and interviews.

Recordings. An analysis of recordings showed three salient features in the experimental group. First, game development showed students' unanimous enjoyment of the game yet with differentiated preferences for game themes; second, peer interaction in L1 and L2 was less observed than teacher-student interaction; and third, students were able to use the vocabulary they had learned in previous sessions during in-class activities.

In relation to the first finding, the recordings lend support to previous research on students' attitudes and motivations in GBL (see for example Liu & Chu, 2010; Mahmoud & Tanni, 2014). The analysis showed that students were highly motivated in all sessions, trying to build their homes, socialize with neighbors, get promoted in their professions and start their own businesses in the simulated world. In this process, they constantly needed to deal with new words and phrases presented by the game and to make decisions. The game offered the students challenges and fulfillment that motivated them to learn more in order to succeed in the game. The coding on game development also showed that each group had its own preference for game themes, and though time devoted to each theme was evenly planned in the experiment design, students tended to spend more time on themes they liked, especially in free gameplay sessions.

In relation to the second finding, contrary to the researcher's expectation, peer interaction was not sufficiently observed and even if students did talk to each other, such discussion was usually in Japanese. The only exception was found in the pair composed of an overseas student and a Japanese student, as they primarily had to rely on English for communication. Meanwhile, teacher-student interaction, or teacher instruction during gameplay, as was observed when the teacher constantly circulated to check on the students, was much more frequent and efficient in solving students' puzzles. On average, interaction in the classroom can be summarized as peer interaction in L2 < peer interaction in L1 < teacher-student interaction in L2. The following are some transcripts that show typical examples of such interactions in the class. The letter S represents students and T, the teacher.

Transcript 1 (originally in Japanese)

S1: what should we do now?

S2: (pointing to a cross with the word "remove") *remove*, this probably means cancel the wall.

S1: (after clicking) ah~ yes.

Transcript one shows an example of peer interaction aimed at completing a task in the game that also provided an opportunity to learn the new word "remove". This kind of interaction was effective in acquiring new vocabulary, but was infrequent and mostly carried out in Japanese.

Transcript 2

S1: (choosing a walking style) hahaha this looks like me, nekoze.
S2: hahaha, how to say that in English, nekoze?
S1: nekoze...hmmm.. I don't really know. (turning to the teacher) Do you know nekoze in English?
T: uhm?
S1: Nekoze, you know like this guy, what' the word for having a round back, having a back structure like this?
T: hmmm. hunchback?
S1: Yes hunchback. Yes, I think it's hunchback.

Transcript 2 shows an example where students communicated in their L2 before seeking help from the teacher. Such peer interaction was limited to the pair with an international student.

Transcript 3

S: How can I complete this task?
T: Mop up a puddle? You first need to find a puddle.
S: Puddle?
T: Do you know what a puddle is?
S: This? (pointing to a pile of rubbish)
T: No, that's rubbish. A puddle is some water on the ground. For example, if someone just took a bath and he forgot to dry himself before he came out from the tub, there would be a puddle where he stood.
S: oh okay.

Transcript 4

T: ... and prioritize food means to ask the chef to cook her food first. So, prioritize, do you know the word?
S: Prioritize, hmmm
T: It means to put something on the top of your to-do list

(waiting for the student to prioritize food in the game and then explaining the next choice)

T: Comp food means to let them eat for free.
S: Com..comp, really, free?
T: Yeah
S: Com..
T: Comp (correcting pronunciation). It means to give away something for free. just like *sabisu* in Japanese
S: Oh I get it.

Transcript 3 and 4 show examples of teacher-student interaction in English, the most frequently occurring form of interaction found in the recordings. It should also be

noted that it was usually the teacher who initiated the dialogue when she saw puzzled faces from the students.

In the case of the third finding, analysis of classroom activities revealed that students were able to use the words they had encountered in the game in their presentations. However, while their word choices were suitable for the context, they made grammar mistakes from time to time. The following are transcripts of students' presentation in class and words in italics are game vocabulary.

Transcript 5

S1: Mikito Toida is the father of this *household*. (We combined the name of each of us.) He is surely handsome and has warm voice. He is actually *aspiring* to be a *public enemy*, related to *deviance*. He walks in a *goofy* way, which makes him more like a *criminal*. On everyday occasion, he wears a *jacket*, white *jeans*, and *sneakers*. I guess you can easily imagine that he would be arrested as soon as you see him!

S2: This girl is 20 years old. And she is *perky* and active. She get upset if she doesn't *muscle training*. She has an *aspiration* to be wealthy and to have successful career.

Transcript 5 contains excerpts from students' introduction to their households. It can be seen students were able to use the trait-related words to describe their characters. Words such as *deviance*, *goofy* and *perky* were new to the students and they understood these words with the help of pictures and animations in the game.

Transcript 6

S: Haruka took a bath 8:30 p.m. She became *playful* thanks to their *bubble bath*. Then her son Taro started sleeping in Atsuya's bed because his father had already slept in his bed. After that, Haruka was watching TV 10:30 p.m. because she *hoped entertainment*. Although she watched *comedy* at first, she changed the *channel* because she liked *romance*. She *delighted* in it.

Transcript 6 also shows an example of how a student used game vocabulary in her presentation. The researcher discovered that as game vocabulary most often appeared as individual words (see Figure 1), students often failed to put newly learned words into grammatical sentences. For example, in the underlined text the student said that her character “hoped entertainment” because the word “entertainment” was not accompanied by any collocations or sentences when she encountered it.

Survey results

Survey data showed that motivation-wise, the experimental group was greatly satisfied with the course, rating on average the course above 4 on a spectrum of 1 to 5. Words the students used most to describe the course were “entertaining” and “interesting”. Six students were “very happy”, four students were “somewhat happy” with the non-pressurized classroom and the role of the teacher as a facilitator, while another said he was neither happy nor unhappy in the class and one said he felt a bit unhappy (see Table 4). When the latter two students were asked why they did not feel happy, they expressed

their concerns that “I may end up just enjoying the game” and “although I enjoyed the game, I don't feel so much that my English skills have improved”.

Table 4. The extent to which students were happy with the class

Attitude	Not happy at all	A bit unhappy	Neither happy nor unhappy	Somewhat happy	Very happy
Number of students	0	1	1	4	6

Surveys further indicated that teacher instruction was the most effective form of classroom intervention in the experiment. The students reported greater likelihood to recall words encountered in both instruction and gameplay, compared to words they came across in gameplay alone. However, they did not think peer interaction helped in improving vocabulary acquisition and suggested that the teacher should always remind or even force them to switch to English in the classroom, though they felt awkward to talk to a Japanese peer in a language other than their mother tongue. This also supported results from the recordings where, in most sessions, some students only spoke English when they were sharing their stories in front of the whole class or when the teacher checked on them. Besides, pair discussion was often less efficient to help the students understand new words. As a student under the pseudonym Marco¹ argued in his interview, it was more efficient to simply ask the teacher, because the teacher was “more fluent and uses new words”.

In terms of in-class activities, the students considered presentations interesting and conducive to vocabulary consolidation. Three students also mentioned that such activities enhanced their writing as well, as they would usually prepare a written script. When asked what improvements should be made to current activities, two opinions caught the researcher's attention. There was one time when the teacher used a screen sharing software for the students to show their stories to the whole class, and a participant named Tak in his interview said there should be more of this activity. The second opinion was proposed by Mukko, indicating that the game entailed too many life choices and that therefore “it was probably better to ask all students to follow the same life path”.

Meanwhile, the experimental group unanimously agreed that they had learned a lot of words for daily communication, students were primarily concerned with two issues not efficiently addressed by the class: vocabulary retention and translation. A girl with the pseudonym Alison said in her interview that she was not sure if she could remember the words later, and Ami described her learning experience as “vague”, saying “I don't know [the corresponding] Japanese translation but know what it means in English”. Examples given by Ami were food names and action verbs. In addition, students had mixed opinions towards game quests and their effectiveness in facilitating vocabulary learning. Some students particularly stressed in their interviews that they needed more guidance on how to complete quests in games.

Discussion

Quantitative data

As the prior discussion shows, quantitative data in the experiment failed to prove the effectiveness of the experiment treatment in promoting vocabulary acquisition. This failure can be explained from three perspectives: the small sample size and number of tests, unequal gameplay time between the two groups and unclear criteria in evaluating learning outcomes.

The controlled experiment involved altogether 24 participants. The small sample size rendered the experiment data less relevant in a broader statistical context. As Peterson et.al. (2018) observes, large-scale research is very rare in game-based learning. This was partly due to the short-term nature of most studies and the difficulty in recruiting students in such a non-traditional teaching method. To address this issue, a great deal of qualitative data must be utilized to complement quantitative data. As such, mixed-methods would continue to be the most suitable way to conduct research in this field. In addition, the number of vocabulary tests is limited. The researcher could not generate a solid trend from the tests as the ups and downs in results may be an artifact of this particular project configuration. More tests should be added, preferably after each session. It was difficult to contact students after the course ended, and therefore the study did not include a post-test conducted after a prolonged period. For future studies where vocabulary retention is a major topic, the inclusion of a post-test should be a priority.

Another uncontrollable factor is the discrepancy in gameplay time between the two groups. Although the researcher asked the control group to play the exact same time as the experimental group and checked their playtime according to submitted save files, the out-of-class nature made it impossible to prohibit students of the control group from playing for more hours. As such, exposure to game vocabulary may be higher in the control group and thus produced higher scores in their unit vocabulary tests. Instead of limiting play hours, the researcher believes it is more appropriate to allow students in classrooms to play the game in their spare time as well. Further, ten students in the experimental group were absent for one or two sessions, and they were not able to make up for the missed sessions outside the classroom. Therefore, when such students took vocabulary tests, they performed lower than the control group who were able to play the game anytime anywhere. This may also account for the 3 outliers in the experimental group during unit 3 and 4 tests and the fewer peer interaction observed in the recordings.

Third, some students in the experimental group were not sure if they had fully acquired a word and as such their test may be under-scored. When asked about whether she has learned new words through the game, a student expressed her concern regarding a lack of corresponding Japanese translation of words. The following is an excerpt from her interview.

Transcript 7

T: You said in your questionnaire that this course helped with your vocabulary acquisition only “a little bit”. Could you talk more about this in detail?

S: Err... I mean, I learnt some words from the playing the game, but I'm not sure I can remember it.

T: You mean you are not sure whether you will be able to recall and use it in the future?

S: Yes. And some words... I don't know I learnt some words or not. I just know what it means, not its meaning.

T: You mean its Japanese meaning?

S: Yes. When I play the game, I know what it means in the game, like if I click a word, my character will do an action. I can describe the word, but I don't know its Japanese translation, so I'm not sure.

It is clear from the transcript that the student's perception of vocabulary acquisition complies with a grammar-translation method where she needs to know the exact corresponding Japanese translation of a certain word in order to fully acquire it, and she chose "vaguely know a word" for such words in the tests. This raises another issue as to how to determine whether a word is "acquired". Does such acquisition require an accurate mapping of words between L1 and L2, and does it require both receptive and productive recall? For future CALL research in vocabulary acquisition, the criteria for evaluating learning outcomes should be stipulated in detail.

Qualitative data

Teacher instruction. Teacher instruction was shown to have a positive influence on students' vocabulary acquisition in this study. Students mentioned in their interviews that the teacher not only provided them with useful information through instruction but also played an untraditional role of a facilitator, which gave them a feeling of closeness and involvement that were not experienced in other traditional English classes. However, as a major part of teacher instruction in this study included teacher interacting with an individual student or a pair, it means that the classroom size needs to be strictly controlled if every one of the students is to get amply involved in such interaction. The attention required from the teacher makes teacher-student interaction unsuitable for large-scale classes.

While they were satisfied overall with teacher instruction, students proposed three areas for future improvement. First, they suggested that as the game offered numerous life paths for players, vocabulary instruction before gameplay seemed slightly general as the teacher tried to cover all possible paths. Therefore, they believed it may have been better if the teacher had restricted gameplay to just one life path where all pairs would have the same job, open the same store, etc. This suggestion is worth considering especially in a formal instructional context where time is limited. The researcher further suggests additional out-of-class gameplay as a complement to GBL in classrooms. In such an approach, students will be able to focus on the restricted theme during class and can still explore other possibilities the game offers after class. Second, the students were concerned with their vocabulary tests and asked for feedback from the teacher. One student mentioned in her interview, "There were words that I didn't know in the test and I want to know them after the test". This was something the researcher did not expect because the tests were originally designed only as a tool to collect quantitative data for the research. In future experiments where vocabulary tests will be employed, feedback on the tests can be added.

Peer interaction.

Peer interaction was not so successful in promoting students' vocabulary learning in this study. A major reason was the awkwardness of students to talk in English, followed by discussion topics on gameplay techniques rather than vocabulary. Some students suggested "the teacher should force us into speaking English in class" while others argued "the teacher should often remind us to speak English". However, the researcher does not believe either method would work, as they would make the students feel uncomfortable and students are highly likely to return to Japanese when the teacher is gone. Insights from the pair where there was an overseas student and a Japanese native suggests that pairing up students of different L1s may be more effective. However, for most language classrooms around the world such an ideal arrangement is extremely difficult to achieve. Therefore, the researcher does not believe teachers should count on peer interaction in L2 with students of the same L1. However, teachers can surely add more vocabulary-focused game quests, such as learning the name of an item by finding or using it in the game, or learning an action verb by completing the action, so that the students will naturally discuss vocabulary-related questions. Such discussions, even in their L1, may promote vocabulary acquisition.

In-class activities

In-class activities in the research comprised of mostly presentations; they successfully solicited output from the students and promoted their productive knowledge of game vocabulary. Such activities can surely be integrated into GBL in classrooms. Furthermore, as students prepared scripts before their presentations, in future research and practice teachers can also ask students to submit their scripts and give them feedback, so that students will be able to know the correct grammar and usage of words. Similar to teacher interaction, a major issue with in-class activities is the time limit. If each pair is to present in class, the number of students must be small enough to allow time for teacher instruction and autonomous gameplay.

In addition, more activities should be explored to avoid repeated presentation activities that may lead to a loss of interest on the part of the students. For example, students can work in pairs to dub a short clip of their gameplay videos based on scripts they have written. They can also exchange computers from time to time to experience the gameplay of various styles created by peers. Moreover, as one accidental screen-sharing activity attracted their attention, students asked for more such activities in the future. "Really interesting, everybody was looking at the same thing". Similarly, in the study by Ebrahimzadeh and Sepideh (2017) which investigated short-and long-term vocabulary retention through a video game, the group of students who watched their peers playing the game outperformed students who learned words through intensive reading. Based on this finding, the researcher believes adding live game streaming into in-class activities may also prove conducive.

Conclusions and Limitations

This study concludes that among the three forms of classroom intervention, teacher instruction and in-class activities were proved by qualitative data to be effective in promoting students' vocabulary acquisition in a GBL classroom. In particular, teacher instruction not only helped students master gaming techniques but also pointed out key words to focus on so that students learned in a more efficiency manner. Teacher-student interaction during gameplay also offered students quick ways to solve their vocabulary puzzles and create opportunities to communicate in English. Peer interaction, however, failed to facilitate learning as students mostly utilized their L1, and such interaction seldom focused on vocabulary itself. In addition, quantitative data were not able to generate any statistical significance between the experimental and control groups. The findings also suggest that for teacher instruction and in-class activities to play a full part, the size of the classroom should be strictly controlled, so that there will be a balance in time allocation among instruction, activities, and gameplay. Improvements to classroom interventions include designing more vocabulary-centered game quests to promote useful peer interaction in both L1 and L2, adopting varied in-class activities and adding feedback on evaluative tests.

The limitation of the study lies mainly in the recruitment of participants and data collection where the self-report technique dominated. In the recruitment of participants, it would have been better to recruit Japanese participants for both the baseline and control group. However, as this study was not funded, the researcher could only ask an English teacher at the Chinese university to engage her students. Although the English competencies of the control and experimental groups were similar according to self-report data, the different cultural background and gaming experiences may have affected the outcome of the study. The small sample sizes are another issue in participant recruitment. The number of registered students in the experimental group determined the small size of the study, rendering it less relevant in larger classrooms. In such a case, more qualitative data must be collected to complement quantitative data. This gives rise to another limitation in the study, the reliability of participants' self-reporting in qualitative data collection. Maxwell and Lopus (1994) argue that students tend to overstate their academic accomplishments. Such systemic reporting biases may well have emerged in this study, particularly when the students were asked to comment on their learning experience. Similar to overstatement, understatement can undermine data reliability as well. For instance, the control and experimental groups showed signs of under-evaluation in vocabulary tests, marking words they actually knew as vague. Apart from limitations in study design, the generalizability of the study is also limited. The study was situated in an English classroom with daily vocabulary as the course goal. Therefore, its results best apply to classes that focus on English for general purposes, rather than for academic purposes. Likewise, the interventions need to be changed if the study is to be applied in courses focusing on writing, listening or other skills. The number of students is another limitation as too many students would render teacher-student interaction and individual presentations impossible.

The researcher, therefore, proposes the following directions for future research: conducting studies with larger sample sizes and diversifying quantitative data, complementing GBL in classrooms with out-of-class gameplay, and in a broader context, investigating the influence of live streams on language learning.

Note

1 All subsequent participant names used are pseudonyms.

References

- Aghlara, L., & Tamjid, N. H. (2011). The effect of digital games on Iranian children's vocabulary retention in foreign language acquisition. *Procedia-Social and Behavioral Sciences*, 29, 552-560.
- Alyaz, Y., Spaniel-Weise, D., & Gursoy, E. (2017). A study on using serious games in teaching German as a foreign language. *Journal of Education and Learning*, 6(3), 250-264.
- Anderson, T. A., Reynolds, B. L., Yeh, X. P., & Huang, G. Z. (2008). Video games in the English as a foreign language classroom. *Proceedings of the 2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning*, Banff, Canada, (pp. 188-192). IEEE.
- Bandicam. (n.d.). Retrieved from <https://www.bandicam.com/>
- Berns, A., Gonzalez-Pardo, A., & Camacho, D. (2013). Game-like language learning in 3-D virtual environments. *Computers & Education*, 60(1), 210-220.
- Bork, A. (1981). *Learning with computers*. Bedford, MA: Digital Equipment Corporation.
- Bredemeier, M. E., & Greenblat, C. S. (1981). The educational effectiveness of simulation games: A synthesis of findings. *Simulation & Games*, 12(3), 307-332.
- Bush, J. C. (2015). The impact of classroom games on the acquisition of second language grammar. *Language in Focus*, 1(2), 17-29.
- Calvo-Ferrer, J.R. (2017). Educational games as stand-alone learning tools and their motivational effect on L2 vocabulary acquisition and perceived learning gains. *British Journal of Educational Technology*, 48 (2), 264-278.
- Canto, S., de Graaff, R., & Jauregi, K. (2014). Collaborative tasks for negotiation of intercultural meaning in virtual worlds and video-web communication. In M. Gonzalez-Lloret, & L. Ortega (Eds.), *Technology and tasks: Exploring technology-mediated TBLT* (pp. 183-212). Washington, DC: Georgetown University Press.
- Carrier, M. (1991). Simulations in English language teaching: A cooperative approach. *Simulation & Gaming*, 22(2), 224-233.
- Chapelle, C. A. (2001). *Computer applications in second language acquisition*. Cambridge: Cambridge University Press.
- Chen, J. C. C. C. (2016). The crossroads of English language learners, task-based instruction, and 3D multi-user virtual learning in Second Life. *Computers & Education*, 102, 152-171.
- Coady, J. (1993). Research on ESL/EFL vocabulary acquisition: Putting it in context. In T. Huckin, M. Haynes, & J. Coady (Eds). *Second language reading and vocabulary learning* (pp.3-23). Norwood, NJ: Ablex Publishing.
- Coleman, D. W. (2002). On foot in SIM CITY: Using SIM COPTER as the basis for an ESL writing assignment. *Simulation & Gaming*, 33(2), 217-230.
- Common European Framework of Reference for Languages: Learning, Teaching, Assessment. (n.d.). Retrieved from <https://www.coe.int/en/web/common-european-framework-reference-languages>

- Crabtree, B., & Miller, W. (1999). Using codes and code manuals: A template organizing style of interpretation. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (2nd ed.) (pp. 163-177). Thousand Oaks, CA: Sage Publications.
- Crook, C. (1996) Schools of the future. In Gill, T. (Ed.). *Electronic Children: How children are responding to the information revolution* (pp. 75–88). London: National Children’s Bureau.
- DeHaan, J. W. (2005). Acquisition of Japanese as a foreign language through a baseball video game. *Foreign Language Annals*, 38(2), 278-282.
- Deutschmann, M., & Panichi, L. (2009). Talking into empty space? Signaling involvement in a virtual language classroom in Second Life. *Language Awareness*, 18(3-4), 310-328.
- Dickey, M. D. (2007). Game design and learning: A conjectural analysis of how massively multiple online role-playing games (MMORPGs) foster intrinsic motivation. *Educational Technology Research and Development*, 55(3), 253-273.
- Driskell, J. E., & Dwyer, D. J. (1984). Microcomputer videogame based training. *Educational technology*, 24(2), 11-16.
- Ebrahimzadeh, M., & Sepideh, A. (2017). The effect of the digital video games on ELF students’ language learning motivation. *Teaching English with Technology*, 17(2), 87-112.
- Egenfeldt-Nielsen, S. (2005). *Beyond edutainment: Exploring the educational potential of computer games*. Unpublished Ph.D. thesis, IT-University of Copenhagen, Copenhagen.
- Fletcher, J. D., Tobias, S., & Wisner, R. A. (2007). Learning anytime, anywhere: Advanced distributed learning and the changing face of education. *Educational Researcher*, 36(2), 96-102.
- Fraser, M. W., & Galinsky, M. J. (2010). Steps in intervention research: Designing and developing social programs. *Research on Social Work Practice*, 20(5), 459-466.
- Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, 33(4), 441-467.
- Gee, J. P. (2007). *What video games have to teach us about learning and literacy* (2nd ed.). New York: St. Martin’s.
- Goodyear, P. (1992). The provision of tutorial support for learning with computer-based simulations. In E. de Corte, M. Linn, H. Mandl, & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving* (pp. 391-409), Berlin: Springer Verlag.
- Goodyear, P. & Tait, K. (1992). Learning with computer-based simulations: tutoring and student modelling requirements for an intelligent learning advisor. In M. Carretero, M. Pope, R. Simons, & J. Pozo (Eds.), *Learning and instruction* (Vol. 3, pp. 463–481), Oxford: Pergamon Press.
- Gredler, M. E. (1996). 17. Educational games and simulations: A technology in search of a (research) paradigm. *Technology*, 39, 521-540.
- Healey, D. (1999) Classroom practice: Communicative skill-building tasks in CALL environments. In J. Egbert & E. Hanson-Smith (Eds.). *CALL environments: Research, practice, and critical issues* (pp. 116–136). Alexandria, VA: TESOL.
- Hwang, G. J., Hsu, T. C., Lai, C. L., & Hsueh, C. J. (2017). Interaction of problem-based gaming and learning anxiety in language students' English listening performance and progressive behavioral patterns. *Computers & Education*, 106, 26-42.

- Jean, G., & Simard, D. (2011). Grammar teaching and learning in L2: Necessary, but boring?. *Foreign Language Annals*, 44(3), 467-494.
- Jones, C. and Fortescue, S. (1987) *Using computers in the language classroom*. New York: Longman, Inc.
- Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education*, 8(1), 13-24.
- Klauer, K. J. (1985). Framework for a theory of teaching. *Teaching and Teacher Education*, 1(1), 5-17.
- Kruk, M. (2015). Practicing the English present simple tense in Active Worlds. *International Journal of Computer-Assisted Language Learning and Teaching*, 5(4), 52-65.
- Leutner, D. (1993). Guided discovery learning with computer-based simulation games: Effects of adaptive and non-adaptive instructional support. *Learning and Instruction*, 3(2), 113-132.
- Liu, T. Y., & Chu, Y. L. (2010). Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers & Education*, 55(2), 630-643.
- Malone, T. W. (1981). What makes things fun to learn? A study of intrinsically motivating computer games. *Pipeline*, 6(2), 50.
- Maxwell, N. L., & Lopus, J. S. (1994). The Lake Wobegon effect in student self-reported data. *The American Economic Review*, 84(2), 201-205.
- Miller, M., & Hegelheimer, V. (2006). The SIMs meet ESL Incorporating authentic computer simulation games into the language classroom. *Interactive Technology and Smart Education*, 3(4), 311-328.
- Mahmoud, A. A. A., & Tanni, Z. A. (2014). Using games to promote students' motivation towards learning English. *Journal of Al-Quds Open University for Educational and Psychological Research and Studies*, 336(2320), 1-21.
- Nagahashi, T. L. (2007). Techniques for reducing foreign language anxiety: Results of a successful intervention study. *Akita University Annual Report on Basic Education*, 9, 53-60.
- Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education*, 52(1), 1-12.
- Peterson, M. (2010). Computerized games and simulations in computer-assisted language learning: A meta-analysis of research. *Simulation & Gaming*, 41(1), 72-93.
- Peterson, M. (2012). Learner interaction in a massively multiplayer online role playing game (MMORPG): A sociocultural discourse analysis. *ReCALL*, 24(3), 361-380.
- Peterson, M. (2016). Virtual worlds and language learning. In F. Farr & L. Murray (Eds.), *Routledge handbook of language learning and technology* (pp.308-319). Oxford: Routledge.
- Peterson, M. (2017). Introduction. In M. Peterson (Ed.), *Digital language learning and teaching: Critical and primary sources, Vol. IV. New developments in computer assisted language learning* (pp.1-18). London: Bloomsbury.
- Peterson, M., Wang, Q., & Mirzaei, M.S. (2018). The use of network-based virtual worlds in second language education: A research review. In M. Kruk (Ed.), *Assessing the effectiveness of virtual technologies in foreign and second language instruction* (pp.1-17). Pennsylvania: IGI Global.

- Purushotma, R. (2005). Commentary: you're not studying, you're just.... *Language Learning & Technology*, 9(1), 80.
- Ranalli, J. (2008). Learning English with The Sims: exploiting authentic computer simulation games for L2 learning. *Computer Assisted Language Learning*, 21(5), 441-455.
- Randi, J., & Corno, L. (1997). Teachers as innovators. In B. Biddle, T. Good, & I. Goodson (Eds.) *International handbook of teachers and teaching* (pp. 1163-1221). Dordrecht: Springer.
- Reinders, H. & Wattana, S. (2015) Affect and willingness to communicate in digital game-based learning. *ReCALL*, 27(1), 38-57.
- Sadler, R. (2012). *Virtual worlds for language learning: From theory to practice*. Bern: Peter Lang.
- Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, 46(4), 842-854.
- Schreier, M. (2012). *Qualitative content analysis in practice*. Thousand Oaks, CA: Sage.
- Schwienhorst, K. (2002). The state of VR: A meta-analysis of virtual reality tools in second language acquisition. *Computer Assisted Language Learning*, 15(3), 221-239.
- Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*, 64(2), 489-528.
- Squire, K. (2005). Changing the game: What happens when video games enter the classroom?. *Innovate: Journal of Online Education*, 1(6).
- Squire, K., & Barab, S. A. (2004). *Replaying history: Learning world history through playing Civilization III*. Bloomington, IN: Indiana University.
- Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, 24(3), 302-321.
- The Best Life Simulation Games of All Time. (n.d.). Retrieved from <https://www.ranker.com/list/the-best-life-simulation-games/reference>
- The Sims 4 (n.d.). Retrieved from <https://www.ea.com/games/the-sims/the-sims-4>
- Thurman, R.A. (1993) Instructional simulation from a cognitive psychology viewpoint. *Educational Technology Research and Development*, 41(4), 75-89.
- Vocabularyprofile English. (n.d.). Retrieved from <https://www.lexutor.ca/vp/eng/>
- Wager, W., & Gagné, R. M. (1988). Designing computer-aided instruction. In D. H. Jonassen (Ed.), *Instructional designs for microcomputer courseware*, (pp. 35-60). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wigham, C. R., & Chanier, T. (2013). A study of verbal and nonverbal communication in second life—the ARCHI21 experience. *ReCALL*, 25(1), 63-84.
- Wu, C. J., Chen, G. D., & Huang, C. W. (2014). Using digital board games for genuine communication in EFL classrooms. *Educational Technology Research and Development*, 62(2), 209-226.
- Yudintseva, A. (2015). Game-enhanced second language vocabulary acquisition strategies: A systematic review. *Open Journal of Social Sciences*, 3(10), 101.

Appendix 1 Sessional themes and activities for both groups

Session	Themes	Gameplay quests	Tests and activities	
			<i>Experimental group</i>	<i>Control group</i>
1	Theme 1: my household	Getting to know the game and deciding household members in the game	Pre-questionnaire and pre-test of game vocabulary	Pre-questionnaire and pre-test of game vocabulary
2		Creating characters, or Sims	Introducing household members	NA
3		Building houses	Presenting family members in detail	NA
4	Theme 2: my life	Living in the simulated world	Vocabulary test 1: my household Presenting homes	Vocabulary test 1
5		Making friends and going on vacations	Storytelling what happened in the family	NA
6&7	Theme 3: my job	Finding a job and getting promoted	Vocabulary test 2: my life Sharing vocabulary	Vocabulary test 2
8	Theme 4: my business	Opening a retail store	Vocabulary test 3: my job Storytelling what happened in your work	Vocabulary test 3
9		Opening a restaurant	Presenting stores	NA
10	Free gameplay	Freely Exploring the world	Presenting restaurants	NA
11		Freely Exploring the world	Vocabulary test 4: my business	Vocabulary test 4

			Post-questionnaire and interview	
--	--	--	-------------------------------------	--

Appendix 2 Pre-questionnaire

1. Please write down your name and email address here.
2. Do you have a recent TOEFL, TOEIC, IELTS or EIKEN score? If so write it below.
3. Have you lived/travelled overseas? If so, write down where you have been and how long you stayed there.
4. How would you describe your daily English?
 - I can barely communicate in English.
 - I can somehow understand and make myself understood in English.
 - I can communicate relatively smoothly in English with occasional mistakes.
 - I can communicate freely in English.
 - Others:
5. How would you describe your computer experience?
 - I'm a very experienced computer user.
 - I often use computers.
 - I seldom use computers.
 - I almost never use computers.
 - Others:
6. How would you describe your gaming experience?
 - I am a game addict.
 - I often play games in my spare time.
 - I seldom play games.
 - I almost never play games.
 - Others:
7. On what kind of platform do you prefer to play games?
 - Mobile phones
 - PCs
 - Consoles (PS4, Xbox, etc.)
 - Handheld consoles (PSV, Switch, etc.)
 - Others:
8. Have you played the *Sims* before? If so write down how long you have played it in the box.
9. Do you think that the game-based class sessions will help improve your vocabulary and communication skills? Please explain why in the comment box.
 - Yes.
 - No.
 - I'm not sure.

Why do you think so?

10. Write down any questions or any comments you have for this course.

Appendix 3 Post-questionnaire

1. To what extent do you think you are happy with the class? Please briefly state your reasons as well.

Not happy at all A bit unhappy Not happy nor unhappy Somewhat happy Very happy

Others:

2. This is a communication class. To what extent do you think this class has improved your English communication skills? And in what ways?

Not at all A little bit Somewhat Very much

Others:

3. To what extent do you think this class has enriched your vocabulary? How did you learn these words?

Not at all A little bit Somewhat By a large margin

Others:

4. Are there any other skills you think this class has helped you improve? And why?
5. What do you think of this class when compared with other traditional textbook-based English classes? You can state both advantages and disadvantages.
6. Do you have any suggestions to this class? Anything is okay.
7. As we are using this data for research, please write down a nickname for yourself in case this research is going to be published. Thank you very much!

Appendix 4 List of words in the vocabulary pre-test

achieve	habit
advanced	historic
advocate	humble
aim	insult
amusement	journey
anticipate	manipulate
apology	material
automate	minimal
band	offer
bill	overlap
broadcast	predict
charm	programming
cheer	property
conceive	recover
confidence	regular
convenience	reject
curious	relieve
decorative	stress
design	sweat
domestic	tax
dull	trigger
employ	unique
extract	various
flexible	whisper
fundamental	wicked