Language Learning & Technology
ISSN 1094-3501

February 2019, Volume 23, Issue 1
pp. 87–113

ARTICLE

Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary

Pia Sundqvist, University of Oslo and Karlstad University

Abstract

The purposes of this study are to examine the relation between playing commercial-off-the-shelf (COTS) games in the wild and L2 English vocabulary and to offer comparisons with non-gamers' vocabulary. Data were collected from two samples of teenage L2 English learners in Sweden, Sample A (N = 1,069) and Sample B (N = 16). Questionnaires and English grades were collected from A and B, productive and receptive vocabulary tests from A, and interviews and essays from B. A quantitative-dominant mixed-methods approach was adopted. Results showed a significant positive correlation between time played and test scores. They also showed that time played was related to types of games played. Multiple regression analysis including time played and types of games as predictor variables and L2 vocabulary as the outcome variable showed that the effect from type disappeared when it was entered into the model, whereas time remained significant. A close examination of 45 words (productive test) revealed significantly higher scores for gamers (compared with non-gamers) at all vocabulary frequency levels, and for particularly difficult words. Overall, findings from Sample B regarding gaming habits and vocabulary (i.e., use of advanced or infrequent words in essays) reflected the results from Sample A, making it possible to conclude that playing COTS games matters for L2 learner vocabulary.

Keywords: *Gaming, L2 Vocabulary, Digital Wild, Extramural English*

Language(s) Learned in This Study: *English*

APA Citation: Sundqvist, P. (2019). Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary. *Language Learning & Technology*, *23*(1), 87–113. https://doi.org/10125/44674

Introduction

The use of digital games in teaching and learning foreign and second languages (L2s) is a topic that has gained traction in second language acquisition (SLA) and computer-assisted language learning (CALL) literature (Reinders, 2012; Thorne, Black, & Sykes, 2009). Digital games provide interesting benefits for learners, such as repeated opportunities for practice, lowered affective filters that encourage risk-taking, explicit information "both on demand and just in time" (Gee, 2007, p. 226), the possibility to take on and play around with different identities and, not least, target language input, interaction, and output (Gee, 2007; Reinders, 2012). However, despite an increased number of studies on the relation between digital gameplay and language learning, empirical large-scale studies are scarce, as are studies that focus on gaming outside institutional settings, in the digital wilds. This study hopes to partially fill that void. Moreover, instead of involving adult learners, which has usually been the case (see Reinders, 2012), this study targets teenagers. Considering the importance of gaming to many young people and its potential for L2 learning, it is surprising that so little scholarly attention has been paid to this specific learner group (cf. Pinter, 2014).

After the social turn in SLA (Block, 2003), there was general acknowledgement of the relevance and value of sociologically- and ecologically-oriented approaches in L2 research (Thorne, Sauro, & Smith, 2015). This study falls into the latter category. More specifically, and in relation to CALL and gaming research, the study takes a player-learner-oriented perspective (Reinhardt, 2017). The main objectives are to examine the relation between playing commercial-off-the-shelf (COTS) games (Van Eck, 2009) in the digital wild

and L2 English vocabulary learning and to offer comparisons with non-gamers' vocabulary. The greatest attention is paid to productive vocabulary. Output, production, and interaction is the type of language practice that many young people get while playing COTS games, making it particularly relevant to study productive vocabulary more closely. Further, Levy (2015) says that "large and small-scale studies are necessary to provide breadth and depth, sometimes through mixed-methods approaches, to reach a deeper understanding of the processes involved" (p. 565). In line with his suggestion, this study draws on both quantitative and qualitative data collected from two samples of learners and adopts a mixed-methods approach. Next, before the research questions are stated, there is a literature review concerning the digital wilds as well as gaming and L2 learning.

Literature Review

Digital Wilds

As pointed out by Wagner (2004), "the real potential for a social approach to language learning lies outside the classroom" (p. 615). It is central for L2 development that learners get the chance to participate in relevant and meaningful interactions, and such interactions take place outside institutions, in the wild. This term, *the wild*, was coined by Hutchins (1995), who used it in reference to real life, situated cognition. Language learning in the wild, then, has to do with using the resources available outside institutional contexts for learning an L2 (e.g., Theodórsdóttir, 2011). As suggested by Kasper (2004), "ordinary conversation can be a particularly productive environment for L2 learning" (p. 553).

Language learning in the *digital* wild (or wilds) differs in that it occurs in "noninstitutionally located *online* cultures" (Thorne et al., 2015, p. 216, emphasis added). In the call for papers for this special issue, *CALL in the digital wild* is said to encompass any type of informal language learning that takes place in digital spaces, communities, and networks, and it should be independent of formal instructional contexts (Sauro & Zourou, 2017). That is the definition used in this article. The focus is on possible learning from COTS games (i.e., pleasure-oriented in contrast to education-oriented games) played voluntarily by learners in their free time, often in their homes using a computer, tablet, or smartphone on broadband access, but also on the bus or in cafes—that is, in contexts that represent the digital wild. Further, research has pointed to the complex semiotic ecologies of multiplayer games, including game-generated texts, collaboration between players, and websites supporting gameplay (Thorne, Fischer, & Lu, 2012). It may be added that "language is not learned only from or through games, but as constitutive of the ecology of the broader discourses surrounding games" (Reinhardt & Thorne, 2016, p. 423). Ryu (2013) demonstrates this L2 learning in beyond-gaming culture, such as when players enrich their gaming experiences by discussing game-related issues in online fora. The term *digital games* refers to all types of games played on various digital devices.

The definition of CALL (of L2 English) in the digital wild is very similar to the definition of *extramural English* (EE; Sundqvist & Sylvén, 2016), or simply *extramural* L_n (for any L2), in that the learning is initiated by the learner—not a teacher or institution. However, extramural L_n may also encompass learner-initiated language learning taking place offline, which makes that term broader. In this paper, English in the digital wild and EE are used interchangeably.

Gaming and L2 Learning

Reinhardt and Thorne (2016) write the following:

> The language use in, around, and about games has increased in quantity, quality, and diversity, as game playing has become a truly global, interactive, multiplayer, and often multilingual practice. As increasing numbers of L2 learners play digital games outside the classroom, and games are produced in an increasing variety of game genres and languages, it has become easier to imagine digital games as authentic, consequential, and widely applicable L2 learning resources. (p. 416)

What they say is supported by findings from a meta-analysis of the relative effectiveness of digital game-

based learning types that revealed that meaningful and engaging games yielded a large effect size, whereas drill and practice games yielded a small effect (Chiu, Kao, & Reynolds, 2012). Their conclusion aligns with suggestions by others, for example, that massively multiplayer online games (MMOs) may be particularly beneficial for L2 learning (e.g., Peterson, 2012; Rankin, Gold, & Gooch, 2006). MMOs are played on commercial international servers and involve hundreds and even thousands of players in simultaneous gameplay.

So, how can gaming and L2 be researched? Reinhardt (2017) suggests that researchers may approach a problem from one of three perspectives: the perspective of the game, the player-learner, or pedagogy. This study represents player-learner oriented research. As clarified by Reinhardt, in such research, "a learner is in effect both a player and a learner whose experience is impacted by variables like gender, age, L2 proficiency, and game literacy" (p. 211). A strength of such research, Reinhardt says, is that it "can be highly authentic and ecologically valid, especially when using vernacular games and focused on learning" (p. 211). A weakness, however, lies in establishing which behaviors and outcomes are linked with, for example, titles or genres.

L2 Vocabulary

Literature on gameplay and L2 learning first appeared in the 1990s. In his evaluation of computer games for L2 learning, Hubbard (1991) asked some key questions that are still valid: for example, whether the quality of the interaction is linguistically rich enough so that it affords learning. Another early study focused on incidental vocabulary acquisition and revealed that the participants' vocabulary improved (game-specific words) thanks to intensive game-work (Cheung & Harrison, 1992). In incidental vocabulary learning, there is no intent to learn, or learning occurs as a by-product of doing something else, such as communicating (Laufer & Hulstijn, 2001). Studies such as these are relevant to this study and accounted for in what follows.

In two American studies using the simulation game The Sims among adult L2 English learners, positive results were found regarding the participants' L2 English vocabulary after having supplemented material to the game to make vocabulary input more comprehensible (Miller & Hegelheimer, 2006; Ranalli, 2008). In another study from the US, L2 English university students assumed virtual identities and played EverQuest II (a massively multiplayer online role-playing game, MMORPG), with positive findings regarding vocabulary gains, presumably thanks to in-game interactions with non-playing characters (Rankin et al., 2006). Also in the US, by letting a Japanese-as-a-foreign-language student play a baseball action game, the player-learner learned sports vocabulary thanks to repetition and contextual cues (deHaan, 2005a). In an experimental study among undergraduates from a computer science university in Japan (deHaan, Reed, & Kuwada, 2010), learners were either players or watchers of an English-language video music game. While both groups recalled vocabulary from the game, the players recalled significantly less vocabulary than did the watchers, most likely due to "the extraneous cognitive load induced by the interactivity of the game" (deHaan et al., 2010, p. 74). Altogether, these studies show the potential of incidental vocabulary acquisition through gameplay.

Whereas the studies above involved adults, others have targeted younger learners. In Sweden, Sundqvist (2009) examined secondary school L2 English learners' involvement in EE and found a connection between digital gameplay and vocabulary size. A follow-up study by Sundqvist and Wikström (2015) investigated the relation between digital gameplay and L2 English vocabulary in tests and essays. Frequent gamers used the most advanced vocabulary in the essays, followed by non-gamers and moderate gamers. The pattern was different for the tests, where frequent gamers scored the highest, followed by moderate gamers and non-gamers, indicating that gameplay aligned more directly with vocabulary test scores than with indicators of vocabulary drawn from essays. Another study (of primary-school learners) revealed that L2 English vocabulary proficiency was positively correlated with frequency of gameplay (Sylvén & Sundqvist, 2012), and a similar finding was established among very young Danish learners (Hannibal Jensen, 2017). Cobb and Horst (2011, p. 25) examined Francophone L2 English learners in Canada (who were instructed to play a mini-game) and found vocabulary gains and increased speed of lexical access; they concluded that a

longer period of game play was needed to stabilize learning. Thus, time was important. In Norway, Brevik (2016) interviewed upper secondary-school learners who were poor readers in first language (L1) Norwegian but strong in L2 English; they were all frequent gamers and reported having learned vocabulary through gaming. These studies underscore the importance of time spent gaming for learning, but they say less about the possible influence of game preference. Thus, there is a gap in the literature regarding the connection between types of games-played and L2 learning. In addition, productive L2 use is under-researched in gaming studies, and studies on younger player-learners from an ecologically-oriented perspective are few. These gaps are addressed in this study.

Categorization of Digital Games

Generally, it has been considered difficult to categorize digital games according to genre, and lately, broadband access "has afforded even more diversification of game type and genre" (Reinhardt & Thorne, 2016, p. 417). For instance, deHaan (2005b) focused on the content of the game when he proposed these categories: sports, virtual pet, simulation, or role-playing and action/adventure games. In contrast, Sundqvist (2013) focused on the number of players involved in simultaneous gameplay when proposing the scale of social interaction (SSI) model for classification. In this model, the potential for L2 English learning is hypothesized to be greater as the scale of the in-game social interaction grows larger. This is due to an increased likelihood of encountering co-players of different nationalities and, as a consequence, a need for a shared language (English). Basically, the model suggests that MMOs are more beneficial for learning English than multiplayer (MP) games which, in turn, are more beneficial than singleplayer (SP) games (for details, see Sundqvist, 2013). As the SSI model can be used in a quantitative approach where overlap between game genres or types must be avoided, it was considered suitable for this study.

Research Questions

Based on data collected from teenage learners, the aims of this study were to examine the relation between playing COTS games in the wild and L2 English vocabulary, with an emphasis on productive vocabulary, and to offer comparisons with non-gamers. Four research questions guided the study:

1. To what extent is there a relation between the time spent playing COTS games and L2 English vocabulary test measures?
2. To what extent is there a relation between four type-of-game-preference groups (i.e., non-gamers, SP, MP, and MMO) and L2 English vocabulary test measures?
3. What does an examination of solution rates (percentage of correct answers) of individual vocabulary items in a productive levels test reveal about gamers' productive vocabulary? Is it different from non-gamers' productive vocabulary? If so, how?
4. What does an examination of infrequent vocabulary in essays reveal about productive vocabulary use among gamers? Is it different from non-gamers' productive use of infrequent vocabulary? If so, how?

Method

The Schools and Teachers

Over the years, the researcher has regularly been invited to give talks to in-service teachers. On such occasions, the researcher used to put out a call for teachers who would be interested in participating in a large-scale study on gaming and vocabulary learning. When teachers at six schools agreed to participate with 16 classes, it was possible to start the project. It ran for three years, with data collected from three cohorts of ninth graders, constituting Sample A (see Appendix A).

Altogether, nine schools took part. They represented a great variety of schools in terms of sizes, students' language backgrounds, grades, and parental educational backgrounds. Moreover, they were spread out

across Sweden in rural areas, small towns, and cities of various sizes. In terms of students' language background, 17% had a L1 other than Swedish, which was close to the national percentage (20% in 2014, Swedish National Agency for Education, 2014). Although the schools and students in Sample A were not identified via a random sampling procedure, it seemed reasonable to consider the sample as a near representation of typical Year 9 students in Sweden.

The researcher worked closely with one contact teacher per school. She trusted her contacts with collecting the data in a reliable way and the contacts trusted the researcher to send corrected tests in return. This worked out well; the researcher could collect data from a large sample and the teachers received valuable information about their students' vocabulary. Nevertheless, this type of research design had limitations. For example, national regulations stipulated that the students were guaranteed to receive 480 hours of English instruction in compulsory school. Thus, the time for collecting data was restricted, which made it necessary to decide on an appropriate balance between practical feasibility and research reliability.

The Participants

Sample A

In total, 1,324 students in 61 classes were invited to participate (see Appendix A). Of these, 1,069 (81%) agreed.[1] All were in Year 9 (ages 15–16), the final year of compulsory school. 528 were male (49%) and 541 were female (51%). It was inevitable that some participants would be absent on some days of data collection. Although the teachers were very helpful in collecting data afterward, it was practically impossible to collect everything from all students. Unfortunately, the productive levels tests from one class were lost in the outgoing mail at one school, and the vocabulary levels tests from two classes were lost when a teacher changed offices. For these reasons, the totals in the results section vary.

Sample B

Sample B was a convenience sample consisting of 16 students at one school (12 girls, 4 boys; Year 9). All were familiar with the researcher, as they had participated in another study while in middle school. They were invited to an interview study of EE, focusing on gaming. Fourteen were L1 Swedish speakers; two were simultaneous bilinguals.

Materials

Several datasets were used. While some were collected from both samples, others could only be collected from one. The different materials listed in Table 1 are discussed more in-depth in the following sections.

Table 1. *Data Collected From Samples A and B*

Data Collected (Used for RQ)	Sample A	Sample B
Time of Collection	2011–2012, 2012–2013, 2013–2014	2015
Questionnaire (RQ1–RQ4)	September	May
Productive Levels Test (RQ1–RQ3)	Fall	n/a
Vocabulary Levels Test (RQ1–RQ2)	Spring	n/a
National Test Essay (RQ4)	n/a	April
National Test Scoring Profile (RQ4)	n/a	April
School Leaving Certificate (RQ4)	June	June
Student Interview (RQ4)	n/a	May

Questionnaire

The questionnaire was used to collect data about students' L2 English learning in the wild through various activities, with a focus on digital gameplay. It had to be kept short to ensure that the teachers would be able

to answer any queries from the students, but mainly because short questionnaires are highly recommended for children (Pinter, 2014). Almost all questions included had previously been used in evaluations of English, reports on youth and media, and research (Oscarson & Apelgren, 2005; Sundqvist, 2009; Swedish Media Council, 2008; Swedish National Agency for Education, 2004). For the purpose of this study, three questionnaire items (1–3) were used (translated from Swedish):

1. I play computer/video games in English (4-point Likert scale: *daily*, *once or a few times per week*, *once or a few times per month*, and *never or almost never*).
2. Approximately how much time per week do you spend on playing computer/video games in English? (4-point Likert scale: *none, because I don't usually play computer/video games*, *less than 3 hours per week*, *3–9 hours per week*, and *more than 9 hours per week*).[2] Those who responded that they played games were also asked to list game titles.
3. Where do you believe you learned most of what you know in English? (4-point Likert scale: *all/nearly all through work in school*, *most through work in school*, *most outside of work in school*, and *all/nearly all outside of work in school*).

The three selected questions were deemed the most suitable ones to validly connect to the broad language habits examined in this study. For Sample A, the questionnaire was collected by the contacts. They followed written instructions; no problems were reported. For Sample B, it was collected by the researcher in conjunction with the interviews.

Vocabulary Tests

Two vocabulary tests were used, the productive levels test (PLT; Laufer & Nation, 1999; Nation, 2001) and the vocabulary levels test (VLT; Nation, 2001). Both were adapted, shortened versions of the original tests in order to suit the participants (for details, see Sundqvist, 2009). Both tests had previously been used successfully with ninth graders (e.g., Sundqvist, 2009; Sundqvist & Wikström, 2015).

The PLT measures productive vocabulary knowledge by use of single sentences with target items, such as *He was riding a bi.....* (bicycle). Altogether, there were 45 items, distributed across four word frequency levels: 17 items from the K2-level (i.e., the second most frequent 1,000 word families), 15 items from the K3-level, 8 items from the K5-level, and 5 items from the university word list (Xue & Nation, 1984), here referred to as *academic words* (see Nation, 2001).

The VLT measures receptive vocabulary knowledge. Six words are listed, where three should be paired up with synonyms or explanations (see Figure 1). Here, the VLT included three parts corresponding to frequency levels K2, K3, and K5 (30 words per level, 90 in total).

1.	apply	
2.	elect	___choose by voting
3.	jump	___become like water
4.	manufacture	___make
5.	melt	
6.	threaten	

Figure 1. An example of the VLT format

The tests were mailed to the contact teachers for administration. To ascertain test administration reliability, the teachers were instructed to carefully follow the written instructions provided. All tests were returned to the researcher for correction. Photo copies were made and corrected originals were mailed back along with a key, student scores, and suggestions for follow-up activities. Vocabulary scores were used by the teachers as part of ongoing formative assessment, which is why all students in a class took the tests (i.e., also non-project students).

Essay

The essay was written as part of the mandatory high-stakes national test of English in 2015.[3] No aids were allowed. Due to test secrecy regulations, the exact test topic formulation and instruction cannot be revealed, but it had to do with writing a text to organizers of a conference.

The essays were transcribed as written. The transcripts were entered into Cobb's (2017) Compleat Lexical Tutor tool (Web VP Classic v.4) to generate counts of tokens and types. Spelling errors and occasional Swedish words were kept, which meant that the counts did not only represent English words. This, however, was considered to be a minor concern regarding validity. In addition, students' use of advanced or infrequent vocabulary was assessed with the help of frequency counts of long words (i.e., polysyllabic words, consisting of three or more syllables; cf. Zipf's law, which states that word length is "inversely proportional to frequency of usage," Malvern, Richards, Chipere, & Durán, 2004, p. 202; see also Sundqvist & Wikström, 2015). The *Longman dictionary of contemporary English online* was used as a reference. Polysyllabic words were counted manually and controlled systematically. The essay instructions included 17 polysyllabic types that were often repeated in students' writing. To control for this, the polysyllabic words from the instructions and the essays were compared, yielding data for each student's use of own polysyllabic types (that is, not copied from the instructions; cf. Sundqvist & Wikström, 2015).

National Test Scoring Profiles and School Leaving Certificates

The national English test scoring profiles contain grades assigned to speaking, listening/reading, and writing. Only the writing grade (A–F) was used here. School leaving certificates were collected from both samples; the final English grade (A–F) was included in the analyses. Grades are aligned with the Common European Framework of Reference for Languages (Council of Europe, 2001) and a passing grade (E) from Year 9 corresponds to level B1.

Interviews

The interviews were semi- structured (Dörnyei, 2007) and focused on eliciting talk about students' informal English language learning practices occurring in the wild, with specific attention to gaming. The researcher used an interview guide and offered the students a poster (describing typical English-mediated activities in the wild) to look at for inspiration, with the intention of foregrounding the students' views, language use, and "individual stories" (Pinter, 2014, p. 172). To the best of her ability, the researcher—who has extensive experience of interviewing both children and adults—conducted the interviews in an age-appropriate way, for example, by adapting her style of speaking when probing for more information (see Mackey & Gass, 2016). The students were divided into pairs or groups of three (see Table 2).

Table 2. *Student Interviews*

Interview	Students in Interview	Minutes
1	Cathrine, Ingela	29
2	Mira, Peter	25
3	Klara, Lovisa	40
4	Bengt, Fredrik	55
5	Daniela, Henrik, Ottilia	41
6	Agnes, Naomi	71
7	Emmy, Greta, Jessica	54
Total		315
Mean		45

Research Questions and Materials

To answer Research Question 1, responses to questionnaire Item 2 were used and viewed as ordinal data. Non-gamers reported not playing anything. Low-frequent gamers reported playing games less than three hours per week, moderate gamers reported three to nine hours per week, and frequent gamers reported more than nine hours per week. To answer Research Question 2, the game titles given in the follow-up question were used to categorize students according to the principles of the SSI model (Sundqvist, 2013; see Table 3).[4] The SSI model was chosen since it allowed for quantitative analysis. In cases where students had listed multiple game titles, perhaps two SP games and one MP game, they were classified according to highest order (in this case, MP).

For Research Question 3, Sample A was split into two groups based on responses to questionnaire Item 1 (Non-gamers = *never or almost never*; Gamers = remaining options).[5] Data from both samples were used to answer Research Question 4.

Table 3. *Distribution of Sample A Participants According to the SSI Model*

Group	*N*	Percentage
Non-gamers	416	38.90
SP	77	7.20
MP	338	31.60
MMO	86	8.00
Unclassified	110	10.30
Missing	42	3.90
Total	1,069	100.00

Analytical Procedures Employed

This study is an example of quantitative-dominant mixed-methods research, or *QUAN+qual*, which is a subtype of mixed-methods research (Johnson, Onwuegbuzie, & Turner, 2007). In QUAN+qual studies, qualitative data and approaches are included into an otherwise quantitative research design. Quantitative data were analyzed using inferential statistics, whereas interview data were analyzed inductively with the goal for findings to emerge from the focus on games and learning (cf. Mackey & Gass, 2016). These procedures allowed for triangulation (Mackey & Gass, 2016).

All statistical tests were run in IBM SPSS Statistics 25. Pearson's chi-squared (χ^2) and Cramér's V (φ_c) were used for tests of association between nominal variables. To calculate significance and effect sizes for tests with numeric variables, independent samples *t*-tests and one-way analyses of variance (ANOVAs) were used, with classical eta squared (η^2) for calculating the effect size. As for ANOVA, Gabriel's post-hoc test was used in order to provide additional indications of which groups differed from which. The non-parametric independent samples Kruskal-Wallis test was employed when the assumptions for ANOVA were not met; effect size was calculated following Jawinski (2017). Spearman's rank order correlation coefficient (r_s) was used in correlation analyses (linear regression) involving ordinal data. Multiple regression analysis was used in examining two predictor variables (Time Played and Types of Games) in relation to an outcome variable (L2 Vocabulary).

For interpretation of effect sizes, Cohen's conventions were used, so that $\varphi_c = .2$ is a small effect size, $\varphi_c = .5$ is medium, and $\varphi_c = .8$ is large (Aron, Aron, & Coups, 2005, p. 192). Regarding eta squared, $\eta^2 = .01$ was a small effect size, $\eta^2 = .06$ was medium, and $\eta^2 = .14$ was large (Dörnyei, 2007, p. 221).

Ethical Considerations

The project adhered to the ethical guidelines used by the Swedish Research Council. Written forms of consent were collected from all participants, who were informed of their rights to withdraw at any time. In this article, pseudonyms are used to ensure anonymity.

Results

Research Question 1

The results for non-gamers, low-frequent, moderate, and frequent gamers on the vocabulary tests are shown in Table 4. The frequent gamers scored the highest on the PLT, followed by the moderate gamers, which in turn were followed by the non- and low-frequent gamers. This pattern was repeated for the VLT. There were significant correlations ($p < .000$) between time spent playing COTS games and scores on both tests; the correlation coefficient was slightly greater for the VLT ($r_s = .31$) compared with the PLT ($r_s = .28$).

Table 4. *PLT and VLT Scores Across Groups Based on Gameplay per Week*

	PLT			VLT		
	M	***SD***	***N***	***M***	***SD***	***N***
Non-Gamers	16.88	9.24	404	57.25	16.85	380
Low-Frequent Gamers	16.86	9.24	155	56.46	17.72	145
Moderate Gamers	20.45	10.36	168	62.02	18.92	162
Frequent Gamers	24.42	10.72	250	70.34	17.54	238
Total	19.42	10.33	977	61.33	18.39	925

Research Question 2

Responses to questionnaire Item 2 yielded 136 game titles that were classified as SP, MP, or MMO games, in accordance with the SSI model (see Appendix B). Based on their answers about gaming habits and game titles, students were coded as Non-gamers or as SP, MP, or MMO following the principle of highest order. Sometimes there were insufficient data; these students were coded as Unclassified.

Table 5. *Cross-Tabulation of Groups (Type-of-Game-Preference Versus Gameplay per Week)*

Gameplay (Hours per Week)	**Non-Gamers (0)**	**Low-Frequent (< 3)**	**Moderate (3–9)**	**Frequent (> 9)**	**Total**
Non-Gamers (*n*)	416	0	0	0	416
SP (*n*)	0	42	26	9	77
MP (*n*)	4	61	102	171	338
MMO (*n*)	0	3	15	68	86
Unclassified (*n*)	3	55	33	4	110
Total (*N*)	423	161	176	267	1,027

The variable with the type-of-game-preference groups was cross-tabulated with the time variable gameplay per week (see Table 5). A chi-squared test revealed that there was a significant relation between these variables, with a medium effect size ($\chi^2 = 1078.742$, $df = 9$, $p < .001$, $\varphi_c = .626$).[6] In other words, it was common that students who had a preference for playing SP also reported not spending a great deal of time on gaming, whereas students who had a preference for playing MP or MMO reported the opposite. For

obvious reasons, the non-gamers did not report any time.

The results on the relation between type-of-game-preference groups (i.e., non-gamers, SP, MP, and MMO) and L2 English vocabulary test measures are presented in Figure 2 (for the PLT) and Figure 3 (for the VLT).

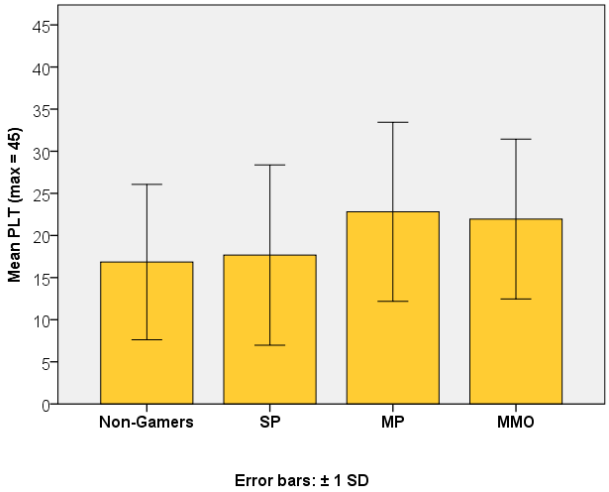

Figure 2. Mean scores for the PLT are shown above for the four type-of-game-preference groups: Non-Gamers, SP, MP, and MMO.

The MP group had the highest mean score, followed by the scores for the MMO, SP, and Non-gamer groups (Figure 2). When comparing these means, the Levene's test was significant ($p < .001$), indicatinig that the variances were significantly different. Therefore, the assumption of homogeneity of variances had been violated, and so the Kruskal-Wallis test was employed instead ($\chi^2 = 222.58$; $dfn = 3$; $dfd = 867$; resulting in $F = 74.19$; $p < .001$; $\eta^2 = .20$). In short, for the PLT, the differences between the groups were significant and the effect size was large.

This pattern of mean scores (i.e., MP > MMO > SP > Non-Gamers) was repeated for the VLT (see Figure 3). ANOVA showed that the differences between the groups were significant ($F_{(3, 818)} = 25.43$; $p < .001$). Gabriel's post-hoc test revealed that the MP group had a significantly higher mean than the non-gamer ($p < .001$) and SP gamer ($p < .001$) group, and so did the MMO group (non-gamer: $p < .001$; SP: $p = .007$). There was a medium to large effect size ($\eta^2 = .085$).

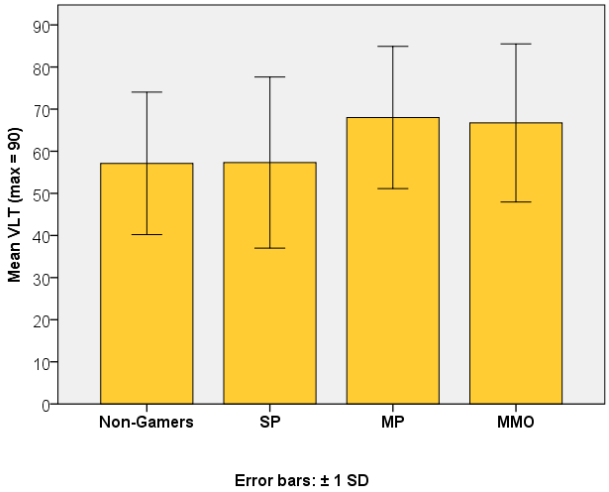

Figure 3. Mean scores for the VLT are shown above for the four type-of-game-preference groups: Non-Gamers, SP, MP, and MMO.

These results showed that Type of Game was related to L2 Vocabulary, but it was also related to Frequency of Playing (time), which was equally related to L2 Vocabulary. Since the two predictor variables (Time and Type) were found to correlate with one another, the question arose whether there might be a mediating effect between them. In order to assess this, both predictor variables were entered into a multiple regression analysis with L2 Vocabulary as the outcome variable (for PLT and VLT, respectively). For the PLT, the effect from Type of Game disappeared when it was entered into the multiple regression model (standardized beta coefficient = -.09; p = .205), whereas Time remained significant (beta = .39; p < .001). Similarly, for the VLT, the effect from Type of Game disappeared (beta = -.07; p = .328), while Time remained significant (beta = .38; p < .001). This suggested that frequency of Playing was the best predictor of L2 vocabulary, and that Type of Game only appeared to predict L2 vocabulary due to mediation. For a sense of the practical significance of the effect identified for Time, students scored on average 3.1 points higher on the PLT per level of the Time variable (unstandardized beta = 3.14).

Research Question 3

The PLT results on the solution rate per word at each frequency level are presented in Figures 4–7. The solution rate represents the proportion or percentage of correct answers for a specific vocabulary item or target word. For each target word, there are two bars, one representing the solution rate for gamers (N = 664), the other non-gamers (N = 362; Missing: N = 43).

There was a significant difference in mean K2-scores between gamers (10.01, SD = 4.46) and non-gamers (8.43, SD = 4.34; p < .001). An examination of the 17 words tested revealed that the solution rates were significantly different between the groups for 13 items (Figure 4). In all cases, the gamers had a higher solution rate. No significant difference (p > .05) between gamers and non-gamers was found for *cream*, *pupils*, *introduced*, and *popular*.

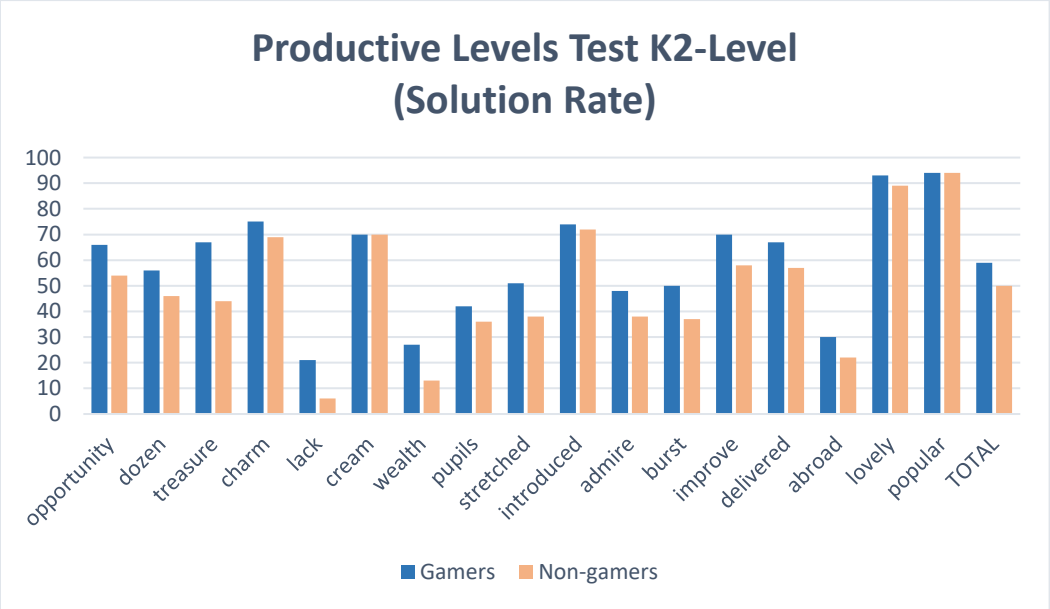

Figure 4. Solution rates per K2-level word and in total are shown above for both gamers and non-gamers.

There was also a significant difference in mean K3-scores between gamers (6.09, SD = 3.49) and non-gamers (4.53, SD = 3.14; p < .001). Of the 15 words at this level, the solution rates were significantly different between the groups for all but one (*gown*), again with the gamers ahead of the non-gamers when the individual items were analyzed (see Figure 5). As expected, the words at the K3-level were more difficult than at the K2-level, reflected in the lower solution rates for the totals, for both gamers and non-gamers. Very few students knew the words *gown*, *proclaimed*, *perceived*, and *slender*. In contrast, more than 70% of the students knew *naked*, most likely because of the Swedish cognate *naken*. On the other

hand, there are Swedish cognates also for *import* and *nerves* but these items, nevertheless, appeared difficult for the students.

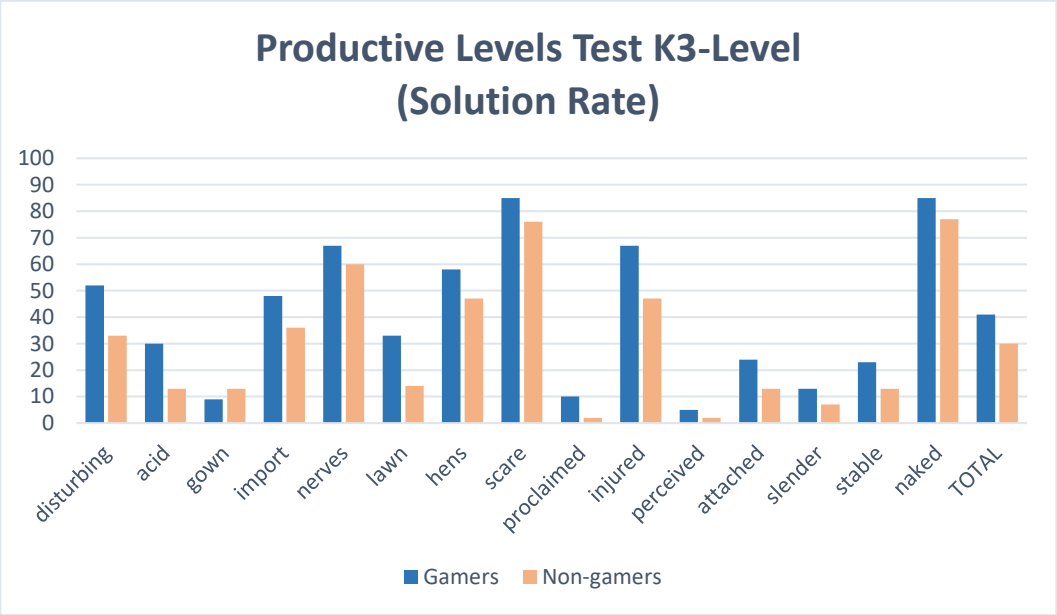

Figure 5. Solution rates per K3-level word and in total are shown above for both gamers and non-gamers.

A similar result was found for the words at the K5-level. Again, gamers had a significantly higher mean score (3.05, *SD* = 1.96) than non-gamers (2.09, *SD* = 1.39; $p < .001$). Of the eight words, the solution rates were significantly different for *oath*, *vault*, *ledge*, *cavalry*, and *mature* (see Figure 6). Gamers outperformed non-gamers. The high solution rate overall for *ignore* can be explained by the Swedish cognate *ignorera*.

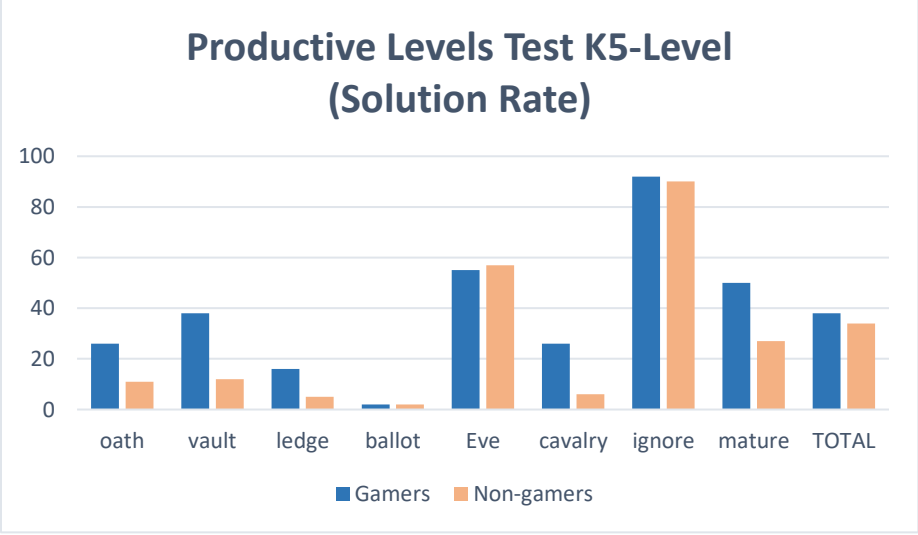

Figure 6. Solution rates per K5-level word and in total are shown above for both gamers and non-gamers.

At the academic level, five words were tested. Gamers had a higher mean (1.78, *SD* = 1.45) than non-gamers (1.49, *SD* = 1.29; $p = .002$). No difference was found for *section* and *motive* (both have Swedish cognates), whereas the gamers had significantly higher solution rates for *inspect*, *saturated*, and *rely* (see Figure 7). There is a Swedish cognate for *inspect*, which probably explains why it was the most well-known word at this level. Interestingly, *saturated* was known by Gamers only.

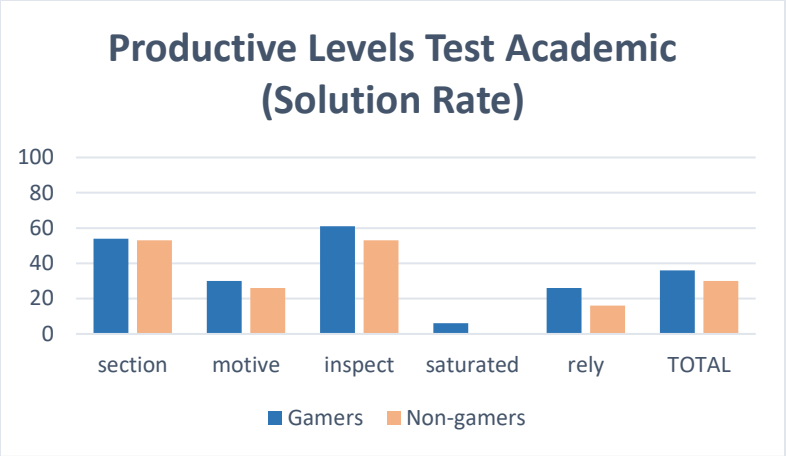

Figure 7. Solution rate per word and in total, Gamers and Non-gamers, Academic words.

It is noteworthy that for particularly difficult words (i.e., words that had low solution rates compared to other words at the same level), the solution rates for the gamers were double (or more than double) those of the non-gamers (K2: *lack* and *wealth*; K3: *acid*, *lawn*, *proclaimed*, *perceived*, and *slender*; K5: *oath*, *vault*, *ledge*, *cavalry*, and *mature*; Academic: *saturated*). In fact, there was a tendency that the proportion of such words increased at higher levels (K2: 2/17; K3: 5/15; K5+Academic: 6/13). Thus, not only were gamers' average solution rates significantly higher at all tested vocabulary levels, they also excelled at particularly difficult words. In sum, the results revealed that gamers had more advanced productive vocabulary than non-gamers.

Research Question 4

Counts of tokens and types in student essays are presented in Table 6. The type–token ratio is a measure of complexity and, thus, indicative of advanced vocabulary use. Table 7 gives the counts of students' use of polysyllabic tokens, polysyllabic types, and own polysyllabic types and also lists the students' own polysyllabic types. Table 8 presents the demographic information of the students, their engagement in EE, and so forth. The level of engagement in EE was established based on interview and questionnaire data, where *high* meant that the students devoted several hours daily to English in the digital wilds (n = 9). For *medium*, the dedication was not as pronounced (n = 4), even though they too were present in the digital wilds, with the remaining three *low* students considerably less so.

Table 6. *Results of VP Classic Analysis of Tokens and Types in Essays*

Student	Tokens (*N*)	Types (*N*)	Type–Token Ratio
Agnes	422	185	.44
Bengt	254	103	.41
Cathrine	373	147	.39
Daniela	320	139	.43
Emmy	344	176	.51
Fredrik	257	128	.50
Greta	453	200	.44
Henrik	424	183	.43
Ingela	255	128	.50
Jessica	459	212	.46

Klara	399	168	.42
Lovisa	454	166	.37
Mira	474	177	.37
Naomi	464	189	.41
Ottilia	529	233	.44
Peter	436	155	.36
M	395	168	.43

Table 7. *Use of Polysyllabic Words in the Essays*

Student	**Polysyllabic Tokens (*N*)**	**Polysyllabic Types (*N*)**	**Own Polysyllabic Types (*N*)**	**Own Polysyllabic Types**
Agnes	32	20	14	*character-everyone-favorite-happening-importance-improving-membership-memory-outgoing-participate-participating-possible-technology-understand*
Bengt	15	8	4	*easier-hospital-interested-telephones*
Cathrine	21	10	5	*another-bullying-communication-example-understand*
Daniela	17	13	11	*addicted-another-available-capable-easier-electronic-everything-experience-family-realized-wherever*
Emmy	19	8	5	*amazing-anxiety-favourite-probably-themselves*
Fredrik	22	15	11	*beginning-commercials-community-computers-concentrating-confidence-everyone-situation-suffering-technology-wonderful*
Greta	29	20	16	*anonymous-another-anything-apartment-escalate-everyone-family-impression-influence-insecure-opinion-ordinary-realizing-realize-responsibility-studying*
Henrik	28	20	17	*activities-activity-Africa-amazing-another-desire-educate-everyone-excited-imagine-medical-opinion-opportunity-perspective-physical-video*
Ingela	14	12	6	*communicate-electronics-everything-example-expensive-family*

Jessica	35	31	28	*activity-anorexia-automatically-bananas-beautiful-customers-encourage-especially-everywhere-extrovert-genetics-happiness-idiotic-industry-insecure-interest-obviously-outgoing-outsider-positive-probably-rejection-society-surprisingly-technology-themselves-trustworthy-understand*
Klara	6	6	5	*already-everyone-interest-listening-Saturdays*
Lovisa	22	11	7	*carbon dioxide-community-creating-factories-happening-hopefully-understand*
Mira	24	17	11	*another-celebrities-changing rooms-connection-easier-electronics-example-expensive-included-interesting-mobile phone*
Naomi	17	11	10	*alcohol-another-community-difference-especially-everybody-example-headmaster-popular-serious*
Ottilia	42	28	20	*anywhere-bullying-challenges-community-computers-considered-ethnicity-everybody-everyone-everything-example-hopefully-including-inspired-interesting-personally-possible-probably-together-unfortunately*
Peter	24	12	10	*alcoholic-apartment-dangerous-expensive-family-negative-positive-themselves-unhealthy-video*

The interviews added valuable information about students' gaming habits. While the majority had ticked questionnaire options that indicated a certain amount of gameplay, the analytical work clarified that six seemed to consider gaming as a significant part of their lives.

Fredrik used to play extensively for some years (mainly MP and MMO games) until his gaming, according to himself, went too far: "I became pissed and just threw all games consoles around and now they're all broken except for like maybe half of one, so it was an effective end to my gaming." Fredrik produced 22 polysyllabic tokens and 11 own polysyllabic types in his essay (Grade B), which was on par with or better than the mean for the highest scoring same-age students on a comparable essay task in a previous study (Sundqvist & Wikström, 2015). His type–token ratio was also high (see Table 6). This ratio, however, is affected by length, and Fredrik's essay was among the shortest.

Henrik was a serious gamer with a preference for multiplayer sports games. His essay was excellent in many ways. It was long, and the polysyllabic counts were indeed indicative of advanced vocabulary use (Table 7). Jessica was also serious about gaming and, similar to Henrik, she played every day. She described how she could be a bit "on-and-off" when it came to gaming, but at the time of the interview, she was definitely "on" (adding "please, don't tell my dad"). In terms of learning, Jessica said that she was "writing a lot in the game" and that "it is possible to speak, but I don't dare," thereby describing a threshold many L2 players may experience when gaming with strangers (cf. Sundqvist, 2015), as Jessica was doing in Heroes of Order & Chaos, an MMORPG.

Table 8. *Demographics, Grades, Gameplay, and Beliefs*

Name	Gender	EG*	FEG*	LOE*	Class	L1	Frequency of Gameplay	Time Spent on Gameplay (Hours per Week)	Examples of Games Played	Beliefs About Where English is Learned
Agnes	F	A	B	High	4	Swedish	Never/almost never	0	n/a	Mostly in school
Bengt	M	D-	E	Low	1	Swedish	A few times per week	< 3	FIFA, NHL	Mostly in school
Cathrine	F	E+	E	Low	1	Swedish	Never/almost never	0	n/a	Mostly in school
Daniela	F	C-	C	Medium	2	Swedish	Never/almost never	0	n/a	Mostly in school
Emmy	F	B	A	High	2	Swedish	Daily	< 3	Juice Jam	Mostly outside school
Fredrik	M	B	B	High	1	Swedish	A few times/month	< 3	n/a	Mostly in school
Greta	F	B	B	High	2	Serbian and Swedish	A few times/week	< 3	Nemo's Reef, 8 Ball Pool	Mostly outside school
Henrik	M	A	A	High	2	Swedish	A few times/week	> 9	NBA 2K15	Mostly outside school
Ingela	F	C-	C	High	1	Swedish	A few times/month	0	n/a	Mostly in school
Jessica	F	B	B	High	2	Swedish	Daily	3–9	Heroes of Order & Chaos	Mostly in school
Klara	F	C-	C	Medium	3	Swedish	Never/almost never	0	n/a	Mostly in school
Lovisa	F	C-	C	Medium	4	Swedish	Never/almost never	0	n/a	Mostly in school
Mira	F	D+	D	Low	1	Swedish	A few times/month	0	n/a	Mostly in school
Naomi	F	C+	D	High	4	Swedish	Daily	< 3	Dota2, Assassin's Creed	Mostly outside school
Ottilia	F	A	B	High	2	Swedish and English	A few times/month	0	n/a	Mostly outside school
Peter	M	C+	C	Medium	1	Swedish	A few times/week	3–9	GTA	Mostly in school

Notes. EG = Essay Grade; FEG = Final English Grade; LOE = Level of Engagement in EE Activities

The remaining three students had very different approaches to gaming in the wild. Naomi described a habit of occasional intense gameplay ("I have to play"), mainly during holidays. She would prepare with salads, finger food, and candy, and then go all-in, often pulling an all-nighter playing Assassin's Creed (SP). In short, Naomi was serious when she was gaming, but she was not a serious gamer in the same sense as Henrik and Jessica. Naomi's essay and productive vocabulary were of average quality, except for the length: she produced many tokens. Although speculative, it is possible that without her gaming experience, she would not have been as productive.

Greta and Emmy described how difficult it could be to stay away from gaming. Compared with Naomi, they played more regularly. In comparison with Henrik and Jessica, however, they played much less. Emmy preferred playing League of Legends (MP), but only with Swedish friends, while Greta enjoyed casual

games, such as Kim Kardashian (SP). Greta's essay was of high quality, as indicated not least by her polysyllabic counts. In terms of advanced vocabulary, Emmy's essay had a good type–token ratio but was otherwise poorer than Greta's; it was still awarded a higher grade, but there are, obviously, other aspects than vocabulary that may affect assessment.

Among the remaining essays written by non-gaming students, those by Agnes and Ottilia stuck out in terms of advanced vocabulary use (especially own polysyllabic types), whereas the rest were of average quality or below.

In their personal stories (Pinter, 2016), many of the students attested to developing their English skills in the digital wilds (that is, not only the gamers). Regarding vocabulary production, ways of developing writing skills came up when Agnes told Naomi that she used a WhatsApp chat group to stay in touch with Italians she had made friends with during a study abroad experience. The girls' conversation continued, then Naomi suddenly started talking about texting friends in English (Excerpt 1).

Excerpt 1. *Agnes and Naomi*

Naomi: *well I don't normally write to people but well like this I write to Swedish people in English*

I: *yeah*

Naomi: *I'm not sure*

I: yeah

Naomi: *we just start writing in Swedish and then we start bri- well like this we just bring in English words*

Agnes: *yes*

Naomi: *and then we just start talking in English*

Agnes: *sometimes I also do so with [name]*

Naomi: *and it's like this it's not on purpose y'know we just go on like this haha so cute*

Agnes: *like why do you write in English*

Naomi: *I dunno we can just have like a totally normal conversation and then just like what are you doing yes but sort of like that*

Agnes: *but with whom with [name]*

Naomi: *[name] or if it's like I sort of want to flirt with someone the you can start writing in English and I'm like okay*

I: *but you*

Agnes: ***are you a***

Naomi: ***you're so cute*** *sort of just stuff sort of like that*

Alicia: ***you're so cute are you and angel did you fall from the sky***

Naomi: *sort of*

Agnes: *yes*

Naomi: *yes things like that*

Notes. I = Interviewer; translated from Swedish; bold = English in the original.

Excerpt 1 shows how easily English was used by these girls in conversations on their smartphones, and during the interview. In fact, both girls described switching frequently between Swedish and English in

everyday conversations—both in text messages and oral interactions—especially with their best friends. Agnes gave an apt description of how fluent her English had become: "Using English is like washing my hands after having been to the toilet," indicating that English is internalized (Lantolf & Thorne, 2006, p. 179) and seemingly part of her identity. Other students gave similar accounts, not least the gamers. Henrik revealed that he found it totally normal to read English articles online every day, and his peers Daniela and Ottilia agreed. Reading everything online in English did not bother them "at all" (quoting Ottilia).

By and large, the results of the qualitative analysis to a great extent support the conclusions drawn in RQ1–RQ3, but in relation to productive vocabulary in essays instead of in vocabulary tests.

Discussion

The results about the relation between time spent on playing COTS games in the digital wild and L2 English vocabulary largely corroborate previous research (e.g., Hannibal Jensen, 2017; Sundqvist & Wikström, 2015). It needs to be mentioned though that in comparison, this study stands out in terms of its scale. While some have suggested that MMOs may be more beneficial for L2 learning than other types of games (e.g., Peterson, 2012; Rankin et al., 2006), to my knowledge, this study is the first (of size) to specifically zoom in on the relation between the types of games played and L2 vocabulary. Using the type-of-game-preference groups, there was an identical pattern for both productive and receptive vocabulary. The MP group scored the highest, followed by the MMO, SP, and non-gamer groups—but with no significant difference between the MP and MMO group nor between non-gamers and SP game players. It was concluded that the types of games played could be tied to the scores. Thus, both time played and types of games appeared connected with L2 vocabulary. However, the multiple regression analysis revealed that it was time that mattered the most, as type became non-significant in the model, underscoring the great importance of time on task for vocabulary learning. Expressed differently, at least the categorization of games used here could not predict L2 vocabulary per se. These findings, combined with the fact that game preference seemed associated with essay results and final grades in English for the interviewed students, call for more research on the topic of game preference.

The examination of productive vocabulary revealed that gamers' average solution rates were significantly higher at all tested vocabulary levels, and they also excelled at words that were notably difficult. This is a key finding. It was beyond the scope of this study to examine whether any of these words were salient in game terminology, but that could be one possible explanation of why gamers outscored non-gamers. As for the word *saturated*, known solely by gamers, the test sentence read "The victim's shirt was satu..... with blood", and it seems likely that those who knew this word knew it either directly from game terminology (e.g., clothing drenched in blood), or from image settings on PCs (i.e., *saturation*).

From a teaching and learning perspective, an implication of these findings is that it seems important for learners, or players (cf. Reinhardt, 2017), to move from playing only SP to also playing MP or MMO games. But above all, they should invest time in gaming for incidental vocabulary learning to happen—and learners who have never tried gaming may well be encouraged to do so. The SSI model (Sundqvist, 2013) was revised in light of these new findings to illustrate the division between SP and MP, and the non-difference between MP and MMO (Figure 8).

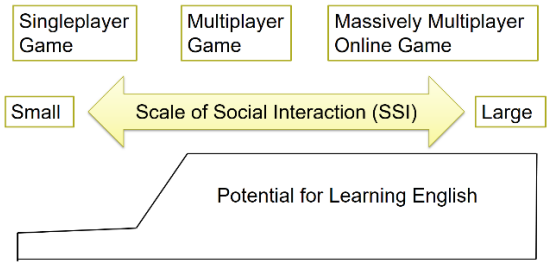

Figure 8. The SSI model revised

Regardless of gaming habits, this study has shown evidence of teenagers who had internalized English and used the language effortlessly in ordinary conversations. Perhaps more importantly, they did not identify themselves as L2 learners, but as L2 users (cf. Carroll, 2000; Kasper, 2004). This was evidenced in many of the interviews and exemplified here by the conversation between Agnes and Naomi (Excerpt 1).

Limitations and Implications

There are some limitations to this study. While it may be authentic and ecologically valid, as discussed in Reinhardt (2017), it is difficult to specify the association between L2 outcomes and game titles or genres. For ethical reasons, such as following regulations about hours of English instruction and respecting teachers' work load, data collection could not take up too much lesson time. Consequently, relatively few items were included in the questionnaire. Another limitation concerns the principle of highest order when classifying students. It is possible that some were coded for a group that was not necessarily the best representation for them. However, to a certain degree, the size of Sample A should compensate for any such misclassifications. Further, other English-mediated activities than gaming may have influenced the findings.

Pedagogical implications of this study include encouraging collaborative classroom work (cf. multiplayer games) and approaching students appropriately when suggesting L2 tasks, depending on how they identify themselves (learners or users). The latter is important not least for learner motivation, as learners benefit from engaging with digital media in the wild and from the affinity spaces for those texts and practices (Gee, 2007).

Conclusions and Outlook

Some important conclusions can be drawn. First, playing COTS games in the wild is clearly related to L2 English vocabulary proficiency. Second, whereas time spent gaming is found to predict L2 vocabulary, types of games appear only to have a mediating effect. However, more research is needed to detangle this specific interplay of variables because it is possible that another type of game categorization than the one employed here (where learners were classified as non-gamers or as gamers preferring to play SP, MP, or MMO games) may yield other findings. Third, compared with non-gamers, gamers show more advanced productive vocabulary and they excel at particularly difficult words in the tests used. Since these findings are based on data collected from a very large sample of teenagers, it makes the study unique. Fourth, the examination of advanced vocabulary use in essays reveal a similar picture in that gamers performed very well. However, so did some non-gamers, and the interviews reveal that other English-mediated activities in the digital wild also seem to contribute to learning (cf. EE in Sundqvist & Sylvén, 2016). Therefore, future research can focus on learning more about the role of different EE variables in player-learner-oriented research (Reinhardt, 2017) as well as on exploring learners-as-users in more depth (Eskildsen & Cadierno, 2015). A corpus of specialized vocabulary in digital games would also be a welcome contribution to the field.

Acknowledgements

This research was supported by the Center for Language and Literature in Education (CSL) and by Research On Subject-specific Education (ROSE), Karlstad University, Sweden. I would like to thank Professor Piet Desmet for his feedback on the design of this study at the CALICO conference in 2016; Peter Wikström, PhD, for assisting me with the multiple regression analysis; and Liliann Byman-Frisén, doctoral candidate, for assisting me with the vocabulary tests and data input. I would also like to thank the anonymous reviewers for their careful reading of the manuscript and many insightful and constructive suggestions.

Notes

1. In reality, the ratio of participation was even higher. Some students were listed as belonging to a specific class, but for various reasons, such as being home schooled or having recently arrived in Sweden, they did not participate in regular English lessons and could, therefore, not participate in the study.
2. A test of internal consistency for Item 1 and Item 2 yielded a Cronbach's α of .92.
3. Samples of this test can be found here.
4. A cross-tabulation of gender and the SSI model groups showed that the groups were unevenly distributed in terms of gender ($\chi^2 = 445.225$, $df = 4$, $p < .001$, $\varphi_c = .658$). Gender falls outside the scope of this study.
5. Questionnaire Item 1 was preferred over Item 2 for this split, since it had been used successfully before (e.g., Sylvén & Sundqvist, 2012).
6. The Unclassified participants were excluded from the chi-squared test and the multiple regression analyses.

References

Aron, A., Aron, E. N., & Coups, E. J. (2005). *Statistics for the behavioral and social sciences: A brief course* (3rd ed.). London, UK: Prentice Hall International.

Block, D. (2003). *The social turn in second language acquisition*. Washington, DC: Georgetown University Press.

Brevik, L. M. (2016). The gaming outliers: Does out-of-school gaming improve boys' reading skills in English as a second language? In E. Elstad (Ed.), *Educational technology and polycontextual bridging* (pp. 39–61). Rotterdam, Netherlands: Sense.

Carroll, D. (2000). Precision timing in novice-to-novice L2 conversations. *Issues in Applied Linguistics, 11*(1), 67–110.

Cheung, A., & Harrison, C. (1992). Microcomputer adventure games and second language acquisition: A study of Hong Kong tertiary students. In M. C. Pennington & V. Stevens (Eds.), *Computers in Applied Linguistics* (pp. 155–178). Clevedon, UK: Multilingual Matters.

Chiu, Y.-h., Kao, C.-w., & Reynolds, B. L. (2012). The relative effectiveness of digital game-based learning types in English as a foreign language setting: A meta-analysis. *British Journal of Educational Technology*, *43*(4), E104–E107. https://doi.org/10.1111/j.1467-8535.2012.01295.x

Cobb, T. (2017). Web VP Classic (Version 4) [Computer software]. Retrieved from https://www.lextutor.ca/vp/eng/

Cobb, T., & Horst, M. (2011). Does Word Coach coach words? *CALICO Journal*, *28*(3), 639–661.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.

deHaan, J. (2005a). Acquisition of Japanese as a foreign language through a baseball video game. *Foreign Language Annals*, *38*(2), 278–282.

deHaan, J. (2005b). Learning language through video games: A theoretical framework, an evaluation of game genres, and questions for future research. In S. P. Schaffer & M. L. Price (Eds.), *Interactive convergence: Critical issues in multimedia* (pp. 229–239). Oxford, UK: Interdisciplinary Press.

deHaan, J., Reed, W. M., & Kuwada, K. (2010). The effects of interactivity with a music video game on second language vocabulary recall. *Language Learning & Technology*, *14*(2), 74–94.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford, UK: Oxford University Press.

Eskildsen, S. W., & Cadierno, T. (2015). Advancing usage-based approaches to L2 studies. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 1–15). Berlin, Germany: De Gruyter Mouton.

Gee, J. P. (2007). *What video games have to teach us about learning and literacy. Revised and updated edition*. New York, NY: Palgrave Macmillan.

Hannibal Jensen, S. (2017). Gaming as an English language learning resource among young children in Denmark. *CALICO Journal*, *34*(1), 1–19. https://doi.org/10.1558/cj.29519

Hubbard, P. (1991). Evaluating computer games for language learning. *Simulation & Gaming*, *22*(2), 220–223.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Jawinski, P. (2017). *Answer to post 'Anyone know how to calculate eta squared for a Kruskal Wallis-analysis'*. Retrieved from https://www.researchgate.net/post/Anyone_know_how_to_calculate_eta_squared_for_a_Kruskal-Wallis_analysis

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, *1*(2), 112–133. https://doi.org/10.1177/1558689806298224

Kasper, G. (2004). Participant orientations in German conversation-for-learning. *The Modern Language Journal*, *88*(4), 551–567. https://doi.org/10.1111/j.0026-7902.2004.t01-18-.x

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. New York, NY: Oxford University Press.

Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26. https://doi.org/10.1093/applin/22.1.1

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33–51. https://doi.org/10.1177/026553229901600103

Levy, M. (2015). The role of qualitative approaches to research in CALL contexts: Closing in on the learner's experience. *CALICO Journal*, *32*(2), 554–568. https://doi.org/10.1558/cj.v32i3.26620

Mackey, A., & Gass, S. (2016). *Second language research: Methodology and design*. Oxford, UK: Routledge.

Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke, UK: Palgrave Macmillan.

Miller, M., & Hegelheimer, V. (2006). The SIMs meet ESL. Incorporating authentic computer simulation games into the language classroom. *Interactive Technology & Smart Education*, *3*(4), 311–328. https://doi.org/10.1108/17415650680000070

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

Oscarson, M., & Apelgren, B. M. (2005). *Nationella utvärderingen av grundskolan 2003 (NU-03). Engelska*. Stockholm, Sweden: Swedish National Agency for Education.

Peterson, M. (2012). Learner interaction in a massively multiplayer online role playing game (MMORPG): A sociocultural discourse analysis. *ReCALL*, *24*(3), 361–380. https://doi.org/10.1017/S0958344012000195

Pinter, A. (2014). Child participant roles in applied linguistics research. *Applied Linguistics*, *35*(2), 168–183. https://doi.org/10.1093/applin/amt008

Pinter, A. (2016). International perspectives on teaching English to young learners. *ELT Journal*, *70*(1), 108–110. https://doi.org/10.1093/elt/ccv067

Ranalli, J. (2008). Learning English with *The Sims*: Exploiting authentic computer simulation games for L2 learning. *Computer Assisted Language Learning*, *21*(5), 441–455. http://dx.doi.org/10.1080/09588220802447859

Rankin, Y. A., Gold, R., & Gooch, B. (2006). *3D role-playing games as language learning tools*. Paper presented at EuroGraphics 2006, Vienna, Austria. Retrieved from http://www.thegooch.org/Publications/PDFs/Rankin_Gold_Gooch.pdf

Reinders, H. (Ed.). (2012). *Digital games in language learning and teaching*. Basingstoke, UK: Palgrave Macmillan.

Reinhardt, J. (2017). Digital gaming in L2 teaching and learning. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology in second language teaching and learning* (pp. 202–216). Hoboken, NJ: John Wiley & Sons.

Reinhardt, J., & Thorne, S. (2016). Metaphors for digital games and language learning. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning and technology* (pp. 415–430). Oxford, UK: Routledge.

Ryu, D. (2013). Play to learn, learn to play: Language learning through gaming culture. *ReCALL*, *25*(2), 286–301. http://doi.org/10.1017/S0958344013000050

Sauro, S., & Zourou, K. (2017). Call for papers for CALL in the digital wilds special issue. *Language Learning & Technology*, 21(1), 186.

Sundqvist, P. (2009). *Extramural English matters: Out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary* (Doctoral dissertation). Karlstad University, Karlstad, Sweden.

Sundqvist, P. (2013). The SSI model: Categorization of digital games in EFL studies. *European Journal of Applied Linguistics and TEFL*, *2*(1), 89–104.

Sundqvist, P. (2015). About a boy: A gamer and L2 English speaker coming into being by use of self-access. *Studies in Self-Access Learning Journal*, *6*(4), 352–364.

Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in teaching and learning: From theory and research to practice*. London, UK: Palgrave Macmillan.

Sundqvist, P., & Wikström, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System*, *51*, 65–76. https://doi.org/10.1016/j.system.2015.04.001

Swedish Media Council. (2008). *Unga och medier 2008*. Stockholm, Sweden: Regeringskansliet.

Swedish National Agency for Education. (2004). *Engelska i åtta europeiska länder*. Stockholm, Sweden: Swedish National Agency for Education.

Swedish National Agency for Education. (2014). Statistik [Grundskolan: Uppdelat per svensk och utländsk bakgrund]. Retrieved from https://www.skolverket.se/skolutveckling/statistik

Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, *24*(3), 302–321. http://doi.org/10.1017/S095834401200016X

Theodórsdóttir, G. (2011). Second language interaction for business and learning. In J. K. Hall, J. Hellerman, & S. Pekarek Doehler (Eds.), *L2 interactional competence and development* (pp. 93–116). Bristol, UK: Multilingual Matters.

Thorne, S. L., Black, R. W., & Sykes, J. M. (2009). Second language use, socialization, and learning in Internet interest communities and online gaming. *The Modern Language Journal*, *93*(S1), 802–821. https://doi.org/10.1111/j.1540-4781.2009.00974.x

Thorne, S. L., Fischer, I., & Lu, X. (2012). The semiotic ecology ad linguistic complexity of an online game world. *ReCALL*, *24*(3), 279–301. https://doi.org/10.1017/S0958344012000158

Thorne, S. L., Sauro, S., & Smith, B. (2015). Technologies, identities, and expressive activity. *Annual Review of Applied Linguistics*, *35*, 215–233. https://doi.org/10.1017/S0267190514000257

Wagner, J. (2004). The classroom and beyond. *The Modern Language Journal*, *88*(4), 612–616. https://doi.org/10.1111/j.0026-7902.2004.t01-21-.x

Van Eck, R. (2009). A guide to integrating COTS games into your classroom. In R. E. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education* (pp. 179–199). Hershey, PA: Information Science.

Xue, G.-y., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, *3*(2), 215–229.

Appendix A. Participating Schools, Classes, and Students

School	Year 1		Year 2		Year 3		3-Year Project	
	Classes	Students	Classes	Students	Classes	Students	Classes	Students
1	4	75	4	51	5	86	13	212
2	1	23	1	23	2	40	4	86
3	1	21	0	0	0	0	1	21
4	4	73	3	66	3	52	10	191
5	4	50	4	57	4	52	12	159
6	2	38	2	35	1	19	5	92
7	0	0	3	56	4	57	7	113
8	0	0	3	69	1	24	4	93
9	0	0	2	51	3	51	5	102
Total	**16**	**280**	**22**	**408**	**23**	**381**	**61**	**1,069**

Appendix B. Game Categorization According to the SSI Model

Game	Singleplayer	Multiplayer	MMO
50 Cent		1	
Age of Empires Online		1	
AION			1
Alice: Madness Returns	1		
Amnesia	1		

Game	Singleplayer	Multiplayer	MMO
Anno 1404	1		
ARMA 2		1	
AruaROSE			1
Assassin's Creed	1		
Baldur's Gate	1		
Batman: Arkham series	1		
Battlefield series (e.g., BF2BC and BF Heroes)		1	
Battlestar Galactica		1	
BioShock	1		
Bloons Tower Defense	1		
Borderlands		1	
Bulletstorm		1	
Burnout		1	
Call of Duty (e.g., Black Ops and Modern Warfare series)		1	
Campus Life	1		
Car Crash	1		
Chivalry: Medieval Warfare		1	
Civilization Revolutions		1	
Counter-Strike (e.g., Global offensive)		1	
Covert Front 2	1		
Crackdown 2		1	
Crysis		1	
Dance Central 4/Just Dance		1	
DarkOrbit			1
Darksiders	1		
DayZ		1	
DC Universe Online			1
Dead Block		1	
Dead Island		1	
Dead Space	1		
Deer Hunter 2005	1		
Dekaron			1
Diablo		1	
Dino Crisis 2	1		
Dirt 3		1	
Dishonored	1		
Dogfight	1		

Game	Singleplayer	Multiplayer	MMO
Don’t Starve		1	
Doom 3		1	
Dota 2		1	
Dragon Nest			1
Europa Universalis		1	
Fallout	1		
Far Cry		1	
Farming Simulator		1	
Fiesta		1	
FIFA		1	
Football Manager	1		
Forza Motor Sport		1	
Gears of War 3		1	
Ghost	1		
GMoD		1	
Gran Turismo 5		1	
Grand Theft Auto	1		
GRID		1	
Guild Wars 2			1
Guitar Hero		1	
Guns of Icarus Online		1	
Half-Life 2 (The Hidden)	1		
Halo: Reach		1	
Happy Wars		1	
Happy Wheels	1		
Harvest Moon: Magical Melody		1	
Heavenly Sword	1		
Heroes of Newerth		1	
Hitman	1		
Horse Saga	1		
Impressive World		1	
Katawa Shoujo	1		
Killing Floor		1	
Kingdom Hearts		1	
L.A. Noire	1		
League of Legends		1	
Left 4 Dead		1	

Game	Singleplayer	Multiplayer	MMO
LittleBigPlanet		1	
MapleStory			1
Mass Effect 2	1		
Medal of Honor	1		
Minecraft		1	
Mirror's Edge	1		
Mount&Blade		1	
Need for Speed		1	
NHL		1	
Nox		1	
Persona 3 Portable	1		
Path of Exile		1	
Payday 2		1	
Pilsner Strip	1		
PlanetSide		1	
Plants vs. Zombies	1		
Portal	1		
Railroad Tycoon 3	1		
Reflex		1	
Roblox			1
Robot Unicorn	1		
RuneScape			1
S4 League		1	
Saints Row 2		1	
Sega Rally	1		
Shayia			1
Skate		1	
Sniper series	1		
Sonic Rush		1	
Splinter Cell: Blacklist		1	
Spore	1		
Spyro the Dragon	1		
Star Wars: The Old Republic (e.g., Cold War)			1
Starcraft II		1	
Street Fighter		1	
Suburbia	1		
Subway Surfers	1		

Game	Singleplayer	Multiplayer	MMO
Super Mario	1		
Supremacy 1914		1	
Tales of Pirates			1
Tamagotchi	1		
Team Fortress		1	
Tetris	1		
The Darkness	1		
The Elder Scrolls series (e.g., Skyrim and Oblivion)	1		
The Legend of Zelda	1		
The Longest Journey	1		
The Simpsons		1	
The Sims	1		
Top Spies	1		
Total War (e.g., Medieval II)		1	
Uncharted	1		
Warcraft III		1	
Warhammer Online			1
Wartune			1
World of Warcraft			1
WRC 3		1	
TOTAL	**53**	**67**	**16**
Percentage	**38.97**	**49.26**	**11.76**

Note. Classification finalized in 2014; links from 2018

About the Author

Pia Sundqvist is currently an Associate Professor of English Education at the University of Oslo and Associate Professor of English Linguistics at Karlstad University. She has a wide range of research interests across applied linguistics, such as informal language learning, CALL (especially gaming), L2 vocabulary acquisition, and the assessment of L2 oral proficiency.

E-mail: pia.sundqvist@ils.uio.no