Article:
Suprapas, I. and González-Fernández, B. (2026) Incidental vocabulary learning through gaming: an exploratory study in an extramural context. ITL - International Journal of Applied Linguistics. ISSN: 0019-0829

https://doi.org/10.1075/itl.25017.sup

Incidental vocabulary learning through gaming

An exploratory study in an extramural context

Iwarin Suprapas and Beatriz González-Fernández
University of Sheffield

Second language (L2) acquisition research emphasises the critical role of out-of-classroom (extramural) input in L2 development. Despite its popularity among L2 learners, the effectiveness of gaming on incidental L2 vocabulary development remains understudied. Thus, this exploratory study investigates whether vocabulary can be acquired from playing computer games and how learner-specific factors (L2-English vocabulary size, gaming habits, engagement) influence learning gains. Twenty-three intermediate English learners played a modified game featuring embedded nonwords in the written game text. Vocabulary knowledge was assessed at form-recall and meaning-recognition levels. An increase in vocabulary knowledge was found for both levels immediately after treatment, with considerable retention after one week. Gaming habits and engagement significantly predicted gains. Findings suggest gaming is an effective extramural activity for incidental L2 vocabulary development, offering important insights for L2 teaching and research.

Keywords: vocabulary, gaming, incidental learning, L2 English, extramural activities

Introduction

Vocabulary is one of the most important elements for successful second language (L2) development. A large amount of vocabulary is required to comprehend different types of English input learners encounter in daily life. For example, 7,000 word families are considered necessary for reading different types of books (Schmitt et al., 2017) and 3,000 word families for TV viewing across various genres (Webb & Rodgers, 2009a, 2009b). Acquiring the necessary vocabulary size for these activities requires significant time and input, with some studies suggesting

https://doi.org/10.1075/itl.25017.sup | Published online: 2 March 2026
ITL - International Journal of Applied Linguistics ISSN 0019-0829 | E-ISSN 1783-1490

that it takes between 500 and 4,000 hours to acquire up to 4,000 word families (McLean et al., 2014; Schmitt, 2008). However, classroom time is usually limited, especially in English-as-a-foreign-language (EFL) contexts. Hence, language input outside of the classroom (a.k.a. extramural input) is considered essential to help learners achieve these vocabulary targets (Webb, 2020). While the potential of some extramural activities for incidental L2 vocabulary development has been thoroughly studied, the most popular leisure activity for L2 learners, gaming, is still largely unexplored (Rodgers & Heidt, 2021). Thus, the present study aims to systematically investigate incidental vocabulary acquisition during independent engagement with commercial off-the-shelf (COTS) games in an extramural context, and to explore how learner-specific factors, such as prior vocabulary knowledge and gaming habits, influence this process.

Background

Incidental vocabulary learning

Incidental vocabulary learning refers to the unintentional acquisition of words while engaging in tasks not explicitly aimed at learning (Schmitt, 2010). To compensate for the limited classroom time and the challenge of providing learners with sufficient L2 input, vocabulary acquisition research has directed its attention towards examining incidental learning that occurs in extramural contexts (Peters, 2018; Richards, 2015).

A substantial body of research on incidental vocabulary acquisition from extramural activities has focused on exploring learning through reading and shows that reading can lead to significant vocabulary gains, especially at the meaning-recognition level, through repeated exposure to words in context (e.g., Chang, 2019; Horst et al., 1998; Waring & Takaki, 2003). More recently, other more popular activities among L2-English learners have been examined, particularly digital audiovisual and multimodal input types, such as watching TV series or movies, and playing games. Studies investigating vocabulary through TV viewing have demonstrated that exposure to TV media, including short clips (Puimège & Peters, 2020), full episodes (Pujadas & Muñoz, 2019), and entire series (Rodgers & Webb, 2020), can support L2 incidental vocabulary development, particularly at the form-meaning mapping level. However, gaming remains a relatively underresearched extramural activity for incidental vocabulary learning. Despite the fact that gaming has increasingly gained popularity as an out-of-class activity among L2-English learners of various age groups and nationalities, including children, adolescents, and university students (see e.g., De Wilde et al. 2020; Peters, 2018;

Rodgers & Heidt, 2021), its role in facilitating L2 vocabulary acquisition outside the classroom is still unclear. Consequently, recent research has begun to explore how games may contribute to L2, particularly vocabulary development.

Gaming and vocabulary learning

Games are typically described as enjoyable, rule-based activities that involve setting and pursuing goals, often played for entertainment or leisure (Juul, 2005). The type of games learners tend to engage with outside formal learning environments are called commercial off-the-shelf (COTS) games (henceforth used interchangeably with 'games'), which are designed primarily for recreational rather than educational purposes. Research on digital game-based language learning (DGBLL) suggests that games possess certain features that can support incidental language acquisition. These features include games' contextual richness (Apperley, 2006), which could aid in deducing meaning, and repeated lexicon across different playthroughs (Reinhardt, 2019). Moreover, the interactive and engaging nature of games, which involves completing purposeful tasks or missions, require the effective use of language to advance to the next stage, much like task-based learning where learners engage in purposeful tasks and utilise language as a tool to achieve defined non-linguistic goals (Reinhardt, 2019; York, 2020). These characteristics collectively highlight the potential of gaming for L2 development.

While still in its infancy, some research has started to examine the relationship between gaming and vocabulary knowledge (e.g., Peters, 2018; Reinhardt, 2017; Reinhardt & Thorne, 2016; Sundqvist, 2013, 2019; Sundqvist & Wikström, 2015; Sylvén & Sundqvist, 2012). Evidence from this body of research, with frequency of gaming gathered from self-reported categorical questionnaires, suggests that playing games outside the classroom positively correlates with learners' vocabulary knowledge (Sundqvist, 2019; Sundqvist & Sylvén, 2014; Sundqvist & Wikström, 2015). For example, Sundqvist & Wikström (2015) investigated how digital gameplay related to Swedish teenagers' school-based English learning and found that students who played more frequently (>5 hours per week [h/week]) tended to outperform both less frequent gamers (<5 h/week) and non-gamers (0 h/week) on standardised vocabulary tests and produced higher-rated essays containing more lower-frequency words. To investigate these patterns on a broader scale, Sundqvist (2019) compared vocabulary knowledge between various categories of gamers (low frequency [<3 h/week]; moderate [3–9 h/week], and frequent [> 9h/week] gamers) and non-gamers among 1069 Swedish learners. The findings aligned with prior research, showing that those who spent the most time playing games in their free time (frequent gamers) tended to score higher on

vocabulary tests (both receptive and productive) than moderate gamers, low-frequency gamers and non-gamers. Additionally, a significant moderate positive correlation ($r = .31$) was observed between time spent gaming and overall vocabulary size, with frequent gamers also demonstrating greater use of advanced vocabulary in their essays. While these earlier studies relied primarily on correlational data, subsequent research has turned to experimental designs to examine whether leisure-based gaming can lead to vocabulary acquisition.

Advancing this strand of experimental research, Calvo-Ferrer and Belda-Medina (2021) explored the effects of gameplay under two different conditions: one involving incidental exposure and another combining gameplay with instructional support. Fifty-four Spanish secondary-school students were divided into two groups: one that played an online multiplayer game with no instruction (incidental), and another that received brief vocabulary pre-teaching along with instructions to produce the target words during the gameplay. As expected, learners in the instructional support condition achieved significantly higher vocabulary outcomes. Yet, both groups demonstrated significant vocabulary gains, confirming that incidental gameplay alone can promote acquisition.

While their study included the examination of vocabulary learning through an incidental condition, their design remained partially classroom-anchored. This limits the study's ability to inform about incidental vocabulary acquisition in settings that better reflect learners' out-of-classroom gaming habits and the leisure-driven purpose for which these games are typically played. Thus, it is precisely this knowledge gap in understanding the vocabulary development and influence of learner variables (e.g., interest, language proficiency) within extramural gaming contexts that the present study aims to address.

Factors affecting incidental vocabulary learning through gaming

Incidental vocabulary acquisition is a gradual and multifaceted process, influenced by both word-related and learner-related factors (Hulstijn, 2012; Reinhardt, 2019; Webb, 2020). A growing body of research has shown how these factors interact to shape and explain lexical learning in out-of-class contexts.

Among word-related factors, word frequency has traditionally been considered crucial for incidental vocabulary acquisition. Lexical items encountered more frequently are believed to build stronger form-meaning connections and are more likely to be retained due to repeated and varied contextual encounters (Teng, 2019; Webb & Chang, 2015). Consequently, experimental studies on incidental learning often meticulously control word frequency to ensure sufficient and consistent exposure across participants (e.g., Pellicer-Sánchez, 2016; Teng, 2019; Webb & Chang, 2015). However, quantifying and controlling word fre-

quency in authentic, dynamic gaming environments presents unique challenges. Unlike linear media such as reading texts or passive viewing, gameplay is interactive and non-linear, driven by player agency (Juul, 2005; Reinhardt, 2019). The amount of exposure to target words in a game is not static but highly dynamic, depending on individual player choices and interactions (e.g., hovering the mouse over an item, clicking on objects, or approaching specific in-game elements). This makes it inherently difficult to precisely quantify or standardise the frequency of exposure of specific words across different players or even within a single player's session. Therefore, while acknowledging frequency's general importance in informing incidental vocabulary learning, the current study focuses on other factors that influence incidental acquisition in such contexts, particularly word saliency and specific learner variables.

Saliency is another critical word-related factor for incidental vocabulary acquisition It refers to the relative importance of a word within a narrative or task structure (Brown, 1993; Lee, 2023). When vocabulary is embedded in a meaningful context, saliency plays a key role in shaping attention and supporting learning (Brown, 1993; Chen & Truscott, 2010). Recent evidence in the context of gaming from Lee (2023) reinforces this point: in her study salient words closely tied to the in-game story (e.g., *miscarriage, midwife*) were significantly more likely to be acquired and retained, even when they occurred infrequently. These findings suggest that the functional importance of a word, that is, its necessity for task comprehension or completion, may outweigh the effect of its frequency of occurrence within the game or learning context. While Lee's work, conducted in a COTS game, explored saliency embedded in narrative context, many COTS games feature words appearing without explicit linguistic context (e.g., as item labels, in menus), where task-relevant nonwords or item labels may become salient through gameplay interaction and in-game progression (Dixon, 2022). Building on Lee's (2023) framework, the present study proposes that quest relevance serves as the functional counterpart to narrative importance as a proxy for salience where target items appear in isolation, without context. Just as keywords' degree of saliency were identified by their relevance to narrative clues for unlocking a plot in Lee's study, the items that appear without context in this study become salient as essential tools for quest completion. Adopting salience as an overarching construct thus allows for an investigation into how the importance of the target words for game progression, regardless of the specific mechanic, influences the incidental acquisition process. This raises the question of whether saliency, operationalised not solely through linguistic context but through quest relevance and gameplay utility, can similarly facilitate vocabulary learning in decontextualised formats.

Beyond word-level considerations, learner-related factors that are believed to influence incidental learning are prior vocabulary size and gaming habits.

Research on audiovisual input has shown that learners with larger vocabulary sizes tend to acquire more words incidentally (Peters & Webb, 2018; Puimège & Peters, 2020). In the gaming domain, frequent gamers have been found to demonstrate higher vocabulary knowledge than their less frequent gamers, possibly due to increased exposure to game-related language (Sundqvist, 2019; Sundqvist & Sylvén, 2014; Sundqvist & Wikström, 2015). This observed outcome may also reflect that the frequency of gameplay fosters greater cognitive efficiency, a concept central to Cognitive Load Theory, which dictates that the human working memory has a limited capacity (Sweller, 1994). Prior research suggested that the simultaneous processing of demanding game mechanics and linguistic input can lead to cognitive overload (DeHaan, 2005; DeHaan et al., 2010), particularly because these complex multimedia tasks involve verbal, visual, and spatial information processing (e.g., Plass et al., 2003). For example, DeHaan (2005) observed that participants' successful word acquisition correlated with their growing familiarity with the game's mechanics, arguing that reduced cognitive load from game management consequently frees up attentional resources for incidental language processing (see also DeHaan et al., 2010). However, these findings remain correlational, and the causal role of gaming in vocabulary development under extramural, incidental conditions has yet to be fully established.

Furthermore, the interactive nature of gameplay introduces engagement as another critical variable in understanding vocabulary outcomes through gaming. Unlike passive forms of media input, digital games require direct interaction with game elements, which can facilitate meaningful engagement (Reinhardt, 2019). Engagement is a multi-dimensional construct, generally referring to learners' level of interest and participation in a given task or activity (Philp & Duchesne, 2016). One of these dimensions is behavioural engagement. In classroom settings, behavioural engagement has often been assessed through indicators such as frequency of verbal contribution or task completion (Gettinger & Walter, 2012). In digital environments, like e-books, engagement has been quantified through interactions with the interface (Roskos et al., 2012; Xu et al., 2021; Zhou & Yadav, 2017). For example, Xu et al. (2021) found that children's interactions with "hotspots" (clickable features embedded in the story) enhance both engagement and learning outcomes, illustrating how digital input can facilitate incidental learning through interactive engagement. Although specific tasks may vary by genre, most games involve meaningful contexts, player interaction, and goal-oriented progression (Juul, 2005). Notably, digital games represent a more autonomous, interactive, and non-linear form of input compared to structured classroom tasks or linear digital materials like e-books (e.g., Xu et al., 2021). Player agency allows learners to choose paths, manage resources, and complete quests in flexible sequences, leading to individualised experiences (Apperley, 2006). This inherent dynamism and

variability introduce complexity into how engagement is understood and measured. Unlike classroom environments where behavioural engagement might be tracked through verbal output or turn-taking (Philp & Duchesne, 2016), gaming might necessitate other observable behaviours, such as time spent exploring or progress metrics like items collected. These alternative engagement indicators, particularly in informal and self-directed gameplay settings, merit closer investigation. Understanding how behavioural engagement affects learning gains in such contexts, and whether it supports vocabulary acquisition, is essential for assessing how COTS games contribute to L2 vocabulary development.

The present study

This exploratory study aims to address the aforementioned knowledge gaps by investigating the extent to which vocabulary can be incidentally acquired through playing a COTS game outside the classroom, specifically focusing on the acquisition of form-meaning knowledge at both receptive and productive levels. It also examines how saliency as a word-related factor and other individual learner factors, including L2-English vocabulary size, gaming habits, and in-game engagement, may impact learning outcomes following these research questions:

1. How much vocabulary can be incidentally learnt from playing an action-adventure game outside of the classroom at the level of form-meaning link in receptive and productive mastery?
2. How do word-related (i.e., saliency) and learner-related (i.e., prior vocabulary knowledge, gaming habits, and in-game engagement) factors affect this learning?

Methodology

Participants

Twenty-three postgraduate gamers from EFL contexts participated in the study and underwent the treatment (no control group due to small participants number).[1] They were recruited online as volunteers. Among the participants, 16 were male and 7 were female, with ages ranging from 20 to 36 years ($M = 25.36$, $SD = 4.41$). Participants represented a range of first language (L1) backgrounds:

1. The reduced number of participants meant that a control group could not be included in this study. We acknowledge this as a key limitation, but believe the research is still highly informative as an initial, exploratory study on this under-researched topic.

while the majority (60%) were Thai, the remainder comprised speakers of Chinese, Spanish, Polish, Gujarati, Greek, Malay, and Hindi. All participants had prior experience with English, with an average vocabulary size (as measured by the VST: Nation & Beglar, 2007) of 10,318 word families (*SD* = 2,400). Thus, participants demonstrated sufficient vocabulary knowledge to play the target game, *The Elder Scrolls V: Skyrim Special Edition*, achieving 98% lexical coverage on the game (achieved with 9,000 word families, as per Rodgers & Heidt, 2021). To investigate gaming habits, participants self-reported how many hours per week (h/week) they spent gaming through a multiple-choice question comprising four categories: 0 h/week, <3 h/week, 3–5 h/week, and >5 h/week. All participants identified as gamers, and their responses fell within only two of those categories, with 56.52% reporting playing games for more than 5 hours per week (i.e., frequent gamers) and the remainder reporting playing 3–5 hours per week (i.e., less frequent gamers). Participants reported an interest in a range of game genres, including adventure, auto-chess, role-playing games (RPGs), sandbox titles, and simulation games, with five participants reporting having played the selected game (Skyrim) before (but no significant effect of game familiarity was found, see results). Importantly, all participants expressed the belief that games can support language and vocabulary learning, and each reported having acquired new vocabulary through prior gameplay experiences.

Materials

The Elder Scrolls V: Skyrim Special Edition (Bethesda Softworks, 2016; henceforth, *Skyrim*) was chosen as the target game in this study. It is a single-player action-adventure game, rich in lore and storyline. Action-adventure games, particularly those with simulation or narrative elements, are well-suited for exploring incidental vocabulary acquisition for several reasons. Unlike repetitive or combat-focused game genres, action-adventure games tend to involve rich storylines, context-driven tasks, and frequent interactions with non-player characters (NPCs) (Apperley, 2006). These features naturally embed vocabulary within meaningful and functionally relevant contexts, allowing learners to encounter words multiple times across varying situations. The communicative and task-based nature of these games mirrors many of the conditions believed to support language learning (see section Gaming and vocabulary learning), including contextualised exposure, goal-oriented interaction, and learner autonomy (Baltra, 1990; Reinhardt, 2019; Apperley, 2006). This aligns with the narrative structure of *Skyrim*, in which the player assumes the role of the Dragonborn and undertakes both main and side quests to battle dragons and prevent the world's destruction.

In addition to completing the in-game quest, participants were instructed to find as many riches (coins) as possible as a side objective, serving as a proxy for their behavioural in-game engagement. Since locating coins required participants to actively explore their surroundings and interact with various objects, the total number of coins collected provided a quantifiable measure of their engagement with the game environment.

Target items

Target items selection

Twenty target items, representing concrete nouns (objects, creatures, consumables, etc.), were first identified from the introductory quest, *Unbound*, in *Skyrim* (see Appendix A). These items were selected based on specific criteria within the *Creation Kit Skyrim* (an official authoring environment allowing researchers to edit game assets and script custom content; Bethesda Softworks, 2011) scripting environment. Within this environment, interactable nouns were extracted only if they appeared at least twice in the quest (see Appendix A for frequency of occurrence[2]) and were presented in a consistent lexical form, allowing only for inflections (e.g., *sword* or *boots*). Where an item could occur with variable adjectival modifiers (e.g., *fur boots*), the adjective was removed so that the lexical target comprised the bare noun (e.g., *boots*).

The original target words were then categorised into three groups based on their quest relevance, serving as a proxy for saliency (see Lee, 2023; Dixon, 2022). These included: (1) quest-related items, which were the most salient as participants were explicitly instructed to interact with them through quest instructions; (2) equipment, which participants needed to aid them in completing the quest but were not explicitly instructed to interact with; and (3) miscellaneous items, which were located around the map unrelated to the main quest, representing the least salient category.

Nonword replacement

To ensure incidental vocabulary acquisition was solely driven by the experimental design, each original target word was replaced by a unique nonword. These nonwords were randomly selected from the ARC Nonword Database (Rastle et al., 2002) based on the following criteria: neighbourhood size ($min = 1$, $max = 5$), number of body neighbours ($min = 1$, $max = 5$), and number of phonological

2. Given the interactive/dynamic nature of gaming, it is difficult to control exactly how many times each participant was exposed to each word. Thus, frequency of occurrence in this study was only employed as a criterion for target word selection and not as a factor for analysis.

neighbours (*min* = 1, *max* = 5). Pseudohomophones were omitted from the selection pool, and only orthographically existing onsets and bodies, and legal bigrams, were included in the generation criteria. Appendix A shows the original target items and their corresponding nonword replacements (See Figure 1 for a sample).

These target nonwords were integrated in the game by overwriting the original noun and replacing it by its assigned nonword (e.g., from *boots* to *claits*) using *Creation Kit Skyrim*. Mirroring the game's features, these items were presented in the written mode (without audio), in isolation (without context) and co-occurring with the corresponding original in-game imagery (see Figure 2).

Saliency category	Nonword replacement	Original target item
Quest	Quares	Bear
	Sprunt	Chest
Equipment	Snarm	Arrow
	Bepth	Helmet
Miscellaneous	Smock	Basket
	Pranced	Broom

Figure 1. Sample target non words and their corresponding in-game items

Figure 2. Sample co-occurring imagery

Measurement instruments

Vocabulary learning was assessed via a form-recall and a meaning-recognition test (see Appendix B). Both tests consisted of the 20 target nonword items and 10 real-word filler items from the game, resulting in 30 test items per session. The filler items were selected from Bawa's (2018) list of gaming vocabulary and cross-compared with words that appeared during the quest, and served as distractors from the focus on non-words.

Form-recall

To assess participants' ability to recall the correct form of the target nonwords, a fill-in-the-blanks format vocabulary test was used. Participants were required to recall the word form, with each prompt situating the target word within a concise, contextualised English sentence adapted from the Oxford English Dictionary (OED). This approach was chosen (rather than providing L1 meanings) to account for the variability of participants' first languages (Laufer & Goldstein, 2004). In addition, the first letter of the target words was provided as a clue (see Example 1). Internal consistency reliability was excellent for both the immediate (Chronbach's alpha (α) = .96, 95% CI [.93, .98]) and delayed (α=.97, 95% CI [.94, .99]) post-tests.

Write the appropriate word in the space to complete the sentences. The word begins with the letter given.

1. A p _____ is a large, heavy hammer with a long handle and usually spikes. It is used as a weapon.
   Answer: _____

Figure 3. Form-recall sample item

Meaning-recognition

A multiple-choice format test was used to assess the participants' learning gains for meaning-recognition. Target words were presented in short, non-informative sentences. Participants were asked to choose the correct answer from four options, which included three authentic-word distractors. The distractors were other words that appeared in the game quest, either in the same word category or physically related to the target nonword, e.g., *sword* and *mace* as distractors for *warhammer*. Additionally, an "I don't know" option was included to prevent guessing. Internal consistency reliability was good-to-excellent (α=.89, 95% CI [.80, .94] for the immediate post-test and α=.92, 95% CI [.86, .96]) for the delayed post-test).

1. *Choose the correct meaning of the word in [square brackets]. If you do not know the answer, please choose "I don't know".*

1. He is carrying a [phloob].
   a. Warhammer
   b. Sword
   c. Mace
   d. Broom
   e. I don't know

Figure 4. Meaning-recognition sample item

Questionnaire

Prior to gameplay, participants completed the Vocabulary Size Test (VST; Nation & Beglar, 2007) to assess their L2 vocabulary size. They also filled out a structured online questionnaire which gathered biographical data (age, gender, nationality, L1, educational level, self-rated English proficiency), detailed gaming habits (weekly play hours, preferred genres, years of experience), and prior familiarity with *Skyrim* and other role-playing games (see Appendix C). These learner-level variables, including gaming habits and vocabulary size, were later entered as covariates alongside prior vocabulary size in the statistical analyses. Familiarity with the game was excluded as a covariate from the models due to its lack of statistical significance across all analyses.

Procedure

Data were collected across three online sessions, each one week apart. During the first session (approximately 50 minutes), participants completed a background questionnaire (see Appendix C) to gather information on their biographical details and gaming habits, as well as the VST. They were then provided with step-by-step installation instructions for the game, including how to apply the necessary modifications to embed the target nonwords into the game environment.[3] The following week, during the treatment session, participants played *Skyrim* for approximately 30 minutes with additional time allowed for character creation (up

3. The game installation instructions focused purely on functional steps (e.g., download, move file, start game) and did not mention the study's vocabulary aims. The technical nature of the required .esp file format to implement the nonword modifications to the game served as a safeguard, ensuring participants could install the modified content without having the means to inspect the specific nonword modifications.

to 15 minutes) to reflect a natural gaming experience and enhance participants' motivation. Immediately following the gameplay, they completed the form-recall and meaning-recognition tests (see Appendix B), in this order, to assess their incidental vocabulary learning. The testing session lasted approximately 30 minutes. One week later, the delayed post-test session took place, where participants were administered the same tests to examine vocabulary retention.

Scoring and analysis

Test responses were scored dichotomously, assigning 1 point for each correct response (learned) and 0 for incorrect responses (not learned). For the form-recall test, a strict scoring criterion was applied, requiring accurate recall of the word's spelling to confirm retrieval of the correct lexical form. This was favoured over a lenient scoring due to the small number of target words (20) and the advanced proficiency of participants (average VST score >10,000 word families).

All statistical analyses were conducted in R (version 4.4.2; R Core Team, 2024). Descriptive statistics were generated, and normality assumptions were checked using Shapiro-Wilk tests. For one-sample and paired comparisons, either *t*-tests or Wilcoxon signed-rank tests were employed, depending on distributional assumptions. Effect sizes (Cohen's *d* and rank-biserial *r*) were also calculated.

To model item-level data, generalised linear mixed-effects models (GLMMs) were fitted using the *lme4* package (Bates et al., 2015), including random intercepts for participants and items. Model fit was summarised with marginal and conditional R^2. Post-hoc comparisons and marginal-means plots were obtained for detailed analysis. Full details of the statistical procedures and findings are reported in the Results section.

Results

Descriptive analysis

Table 1 presents the descriptive statistics for the immediate and delayed posttests, categorised into two levels of word knowledge: form-recall and meaning-recognition. In the immediate posttest, participants scored an average of 32.4 per cent ($SD = 35.7$) on the form-recall test and 63.5 ($SD = 26.6$) on the meaning-recognition test. This shows considerable learning occurred immediately from playing the game. The mean score dropped slightly in the delayed posttest with an average of 29 ($SD = 37.35$) in the form-recall and 58 ($SD = 30.3$) in meaning-

recognition, suggesting that most of the vocabulary learnt was retained after a week, particularly at the meaning-recognition level.

Table 1. Descriptive statistics: Raw and percentage scores

Test time	Aspect	*Raw (SD)*	*M% (SD)*	Range	CI
Immediate	Form-recall	6.48 (7.14)	32.4 (35.7)	0–100	[17.0, 47.8]
	Meaning-recognition	12.70 (5.33)	63.5 (26.6)	5–100	[52.0, 75.0]
Delayed	Form-recall	5.80 (7.47)	29.0 (37.4)	0–100	[11.5, 46.5]
	Meaning-recognition	11.60 (6.05)	58.0 (30.3)	0–95	[44.1, 72.4]

Vocabulary gains over time

A Shapiro-Wilk test indicated that form-recall scores were non-normally distributed ($W = .91$, $p = .001$), while meaning-recognition scores showed normality ($W = .97$, $p = .321$). Thus, to determine whether the scores in the immediate posttest were significant, a one-sample Wilcoxon signed-rank test and *t*-test were conducted with the form-recall scores and meaning-recognition scores, respectively. As the study employed nonwords as target items, it was assumed that participants had no prior knowledge of either the word forms or their associated meanings (any correct answers on meaning recognition at pretest level would have simply indicated uninformed guessing). Therefore, participant scores were compared against a baseline value of 0.

Results from the Wilcoxon signed-rank test indicated that form-recall scores were significantly above zero ($V = 136$, $p < .001$), with a large Pratt-corrected rank-biserial effect size ($r = .70$, 95% *CI* [.46, .87]). Similarly, the one-sample *t*-test for meaning-recognition scores revealed a significant effect ($t(22) = 11.42$, $p < .001$), with a large effect size (Cohen's $d = 2.38$, 95% *CI* [1.57, 3.18]). These findings provide strong evidence that participants incidentally acquired a significant amount of vocabulary through playing the computer game.

To assess whether the vocabulary learned was retained one week after gameplay, paired-sample comparisons were conducted between immediate and delayed post-test scores for each type of word knowledge. For form-recall, a Wilcoxon signed-rank test revealed a non-significant decrease ($V = 56.5$, $p = .18$). The rank-biserial correlation ($r = .45$, 95% *CI* [–.09, .78]) suggested a moderate decline, though the confidence interval included zero, indicating statistical uncertainty. In contrast, for meaning-recognition, a paired-samples *t*-test showed a statistically significant decline from immediate to delayed post-test ($t(20) = 2.41$, $p = .025$), with a medium effect size (Cohen's $d = .53$, 95% *CI* [0.06, 0.98]). These

findings suggest that while most form-recall knowledge was retained after a week, recognition-based knowledge exhibited a more significant decline.

Effect of word- and learner-related factors

To explore the effect of the target factors (prior vocabulary size, gamer-type, in-game engagement, and item saliency) on the incidental learning of target non-words, two generalised linear mixed-effects models (GLMMs) with a logit link were fitted using the *lme4* package (Bates et al., 2015), one for each testing period and type of word knowledge. Model fitting began with a core model that included gamer type and item saliency (item category), with prior vocabulary size and in-game engagement (indexed by the number of coins collected) entered as covariates. Due to its foundational role in L2 vocabulary acquisition (Nation, 2013; Webb, 2007), prior vocabulary knowledge was kept in the model regardless of its individual statistical significance.

A step-by-step backward model selection procedure was applied (see Appendix D). Non-significant predictors were excluded one by one, starting with those with smaller *z*-values, and each new model was explicitly compared with the previous one. Nested models were compared using the chi-squared test for significant differences, as well as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) scores, until all remaining factors were significant, ensuring the selection of the best-fitting model. Random intercepts were specified for participants and items. The best fitting models for each word knowledge type are reported in Tables 2 and 3.

Form-recall

The best-fitting model for form-recall (see Table 2) included the fixed effects of Test Time, Gamer type, Prior Vocabulary Knowledge, In-game Engagement (coins collected), Item Category (for saliency), and the Test Time x In-game Engagement interaction, alongside random intercepts for participants and items. Results showed that the type of gamer had a significant direct effect on form-recall learning outcomes ($p = .0068$), indicating that frequent gamers were significantly more likely to correctly recall target word forms than less-frequent gamers. The Test Time x Gamer Type interaction was non-significant (and thus excluded from the final model), suggesting that the effect of the type of gamer remains for both immediate acquisition and retention of word form knowledge.

In-game engagement, measured by the scaled number of coins collected, was a significant positive predictor of form-recall knowledge. Simple slopes analysis of the significant Test Time x In-game Engagement interaction revealed that the positive effect of engagement was strong and significant at the immediate

posttest ($p < .001$), suggesting that increased coin collection was associated with a greater likelihood of recalling word forms immediately. However, this effect was significantly attenuated at the delayed posttest ($p = .0071$ for the interaction term), resulting in a non-significant simple slope of in-game engagement for the delayed posttest ($p > .05$), indicating that the benefit of engagement on word form retention faded over one week.

Item saliency showed a significant effect on form-recall: items in the miscellaneous category (i.e., unrelated to the quest or gameplay) had significantly lower odds of being recalled compared to Quest items ($p = .0177$). Pairwise comparisons confirmed the word form of the quest-related items had significantly higher odds of being recalled compared to miscellaneous items ($OR = 3.32$, $p = .0465$, Tukey-adjusted). This effect was present across the time points in the model.

Finally, prior vocabulary knowledge did not significantly affect the acquisition or retention of form-recall ($p = .798$). This finding will be covered in the Discussion section.

Table 2. Best-fitting model for form-recall

Fixed effects	***β (Estimate)***	***SE***	***z***	***p***
Intercept	0.144	0.833	0.173	0.862
Test Time (Delayed)	−0.384	0.245	−1.564	0.118
Prior Vocab Knowledge	−0.161	0.632	−0.256	0.798
Gamer (Less-frequent)	−3.495	1.292	−2.707	.007**
In-game Engagement	1.769	0.69	2.562	.010*
Item Category (Equipment)	−0.167	0.425	−0.393	0.694
Item Category (Misc)	−1.201	0.506	−2.372	.018*
Delayed x In-game Engagement	−0.648	0.241	−2.694	.007**
Random effects	**Variance**	**SD**		
Participant ID	6.018	2.453		
Item	0.335	0.579		

Meaning-recognition

The optimal model for meaning-recognition (Table 3) retained the covariates Prior Vocabulary Knowledge and In-game Engagement, along with the factor Gamer type. The model also included the significant Test Time x Gamer type interaction. The results showed that gamer type was a significant predictor both for immediate ($p = .0378$) and delayed ($p = .0004$) posttests, suggesting that frequent gamers have greater odds of learning and retaining meaning recognition

knowledge than less-frequent gamers. Furthermore, comparing the time points, for less-frequent gamers, the odds of correct meaning recognition at the immediate posttest were 2.585 times higher than the odds in the delayed posttest one week later ($OR = 2.585$, $p = .0004$), suggesting a significant decline in retention for this group. For frequent gamers, the odds were not significantly different between the two posttests ($p = 0.8609$), indicating retention over the one-week period.

In-game engagement (coins) was a significant positive predictor of meaning-recognition knowledge across both time points ($p = .0093$). This suggests that higher levels of engagement were associated with greater odds of correctly recognising word meanings.

Item saliency was not included in the optimal model for meaning-recognition due to lack of significance. Similarly, prior vocabulary size showed no significant effect on the acquisition or retention of meaning-recognition knowledge ($p = .979$).

Table 3. Best-fitting model for Meaning-Recognition

Fixed Effects	*β (Estimate)*	*SE*	*z*	*p*
Intercept	1.459	0.469	3.112	.002**
TestTime (Delayed)	0.047	0.266	0.175	0.861
Prior Vocabulary Knowledge	0.008	0.31	0.027	0.979
Gamer (Less-frequent)	−1.311	0.631	−2.077	.038*
In-game Engagement	0.814	0.313	2.601	.009**
Delayed x LessfrequentGamer	−0.996	0.378	−2.635	.008**
Random Effects	**Variance**	**SD**		
Participant ID	1.383	1.176		
Item	1.362	1.167		

Discussion

Incidental vocabulary learning through COTS games

The results from the immediate posttest revealed that learners gained knowledge of a significant number of target items. As expected, the average scores were significantly higher at the meaning-recognition level (63.5%) than in form-recall (32.4%). This is in line with vocabulary acquisition models for EFL learners which show that receptive mastery is acquired before productive mastery (González-Fernández, 2025; González-Fernández & Schmitt, 2020). The study

also demonstrates that COTS games can facilitate the acquisition of both recognition and recall knowledge of new lexical items, and that a large proportion of the knowledge were also retained over the course of one week (29% for form-recall, 58.42% for meaning-recognition). While meaning-recognition scores showed a statistically significant 5.08% decline, form-recall scores did not decrease significantly, suggesting that this type of knowledge, while being harder to achieve, was also harder to forget. This resonates with Barclay & Pellicer-Sanchez's (2021) findings that vocabulary with a higher learning burden (i.e., that took longer to acquire or required more cognitive effort) demonstrated better retention (lower decay) over time. In this study, the more "burdensome" form-recall knowledge aspect was retained to a greater extent than the easier, meaning recognition aspect. While the lack of significant decay for form-recall is encouraging, it must be interpreted cautiously, as the small initial acquisition gains may have constrained the potential for knowledge decline over time. This possibility should be explicitly accounted for and tested in future experimental designs.

The results also support longstanding claims in DGBLL research that games can serve as rich environments for language development, particularly vocabulary acquisition (Reinders & Wattana, 2014; Reinhardt, 2017; Sundqvist, 2013). Importantly, these outcomes are directly comparable to, and in some respects surpass, gains reported in studies using other input modalities. For instance, the immediate meaning-recognition gain of 59.2% in the present study notably exceeds those reported from both shorter-term TV viewing studies (17–31.8% meaning recognition; Peters et al., 2016) and longitudinal TV viewing studies (5.97–8.34% incidental meaning recall; Pujadas & Muñoz, 2019). Furthermore, the study's form-recall gain of 35.4% significantly surpasses incidental form-recall gains (13.02–14.30%) from longitudinal TV viewing (Pujadas & Muñoz, 2019). A possible explanation is that, unlike TV series, where learners passively receive input, COTS games offer multisensory enrichment through visual, auditory, and interactive elements, and require action and decision-making, potentially supporting stronger form-meaning mapping. Players not only read target items but also manipulate them within the game world, equipping weapons, collecting items, and receiving feedback through sound and animation. This rich, multimodal input is believed to enhance memory encoding through elaboration and deeper processing (Carpenter & Olson, 2012; Mayer et al., 2015), and the greater increase in vocabulary knowledge shown in this study compared to those of other types of incidental learning seem to point to this conclusion. In sum, the results highlight the great potential of gaming as a useful form of interactive language input outside of the classroom. However, the absence of a control group and the potential information interference between vocabulary tests across testing points means that the

findings of this exploratory study need to be interpreted with caution and further experimental designs are needed to confirm these results.

Learner-related factors: Gamer experience, engagement, and cognitive load in vocabulary learning

GLMMs were used to explore how gamer type, in-game engagement, and prior vocabulary size influenced vocabulary acquisition through gaming. Surprisingly, prior vocabulary size did not significantly influence learning outcomes in any of the models. This contrasts with previous findings from audiovisual input studies (e.g., Peters & Webb, 2018; Puimège & Peters, 2020), where learners with larger prior vocabularies tend to acquire more lexis. Potential reasons for this discrepancy may lie in the use of nonwords and the presentation of these words. In this study, target words were presented in isolation (without linguistic context) with co-occurring imagery. This simplified presentation likely enabled even learners with smaller vocabularies to easily establish the form-meaning link, thus potentially neutralising the advantage typically held by more proficient learners who rely on linguistic contextual clues (e.g., Reynolds, 2016; Reynolds, 2020). Additionally, the limited variability in the participants' vocabulary size scores (*Mean* = 10,318, *SD* = 2,400 word families) may have contributed to the lack of a significant effect of prior vocabulary knowledge, as the overall high proficiency of the group made it difficult to detect any predictive relationship. Future research should consider this and further explore the effect of prior vocabulary during gaming.

The influence of the type of gamer was evident across both vocabulary aspects. While there was a strong, overall advantage for frequent gamers in the odds of correctly recalling word forms, the results for meaning-recognition revealed a more complex pattern through a significant interaction between type of gamer and test time. Specifically, the meaning-recognition knowledge gained by frequent gamers was remarkably stable, showing no significant decline from the immediate to the delayed test. In contrast, less-frequent gamers experienced a significant decay in meaning recognition knowledge over the one-week retention interval. This suggests that frequent gaming may not just lead to better initial acquisition but, more importantly, to more durable knowledge consolidation. This advantage also extended to the form-recall level, which is generally considered harder to acquire (González-Fernández, 2025; González-Fernández & Schmitt, 2020). These findings echo earlier work linking frequent gaming habits to richer receptive and productive vocabularies, particularly Sundqvist & Wikström (2015) and Sundqvist (2019), who found that more frequent gamers possess larger vocabulary sizes. By demonstrating benefits at the challenging

form-recall level, the current data add crucial nuance to this line of research, indicating that frequent gaming may allow learners to incidentally acquire more vocabulary at a recall level, potentially supporting its productive use in other contexts. The results also align with prior findings that more experienced gamers might be able to manage cognitive load more efficiently, potentially freeing up attentional resources for language processing (DeHaan et al., 2010). Their familiarity with game mechanics may also support deeper engagement with the linguistic input embedded in gameplay, facilitating vocabulary learning (DeHaan, 2005).

In-game engagement (coins collected) was a significant predictor of vocabulary learning, but its influence on knowledge durability differed notably between the two outcomes. For form-recall, the positive effect of in-game engagement was strong and significant only at the immediate posttest. The significant interaction with the time of testing demonstrated that this initial benefit significantly faded by the delayed posttest. This suggests that active exploration and interaction successfully enhance the immediate noticing and encoding of the word form; however, this short-term benefit alone may not translate into durable, long-term memory for the form. In contrast, for meaning recognition, the overall effect of in-game engagement remained a significant positive predictor across the time points in the model. This finding supports Laufer and Rozovski-Roitblat's (2015) argument that the quality of attention, rather than sheer exposure, is crucial for incidental acquisition. Active exploration and interaction appear to enhance noticing and elaboration, both key conditions for vocabulary growth (Webb & Nation, 2017). This result further resonates with the concept of behavioural engagement (Philp & Duchesne, 2016), which refers to the visible and measurable signs of a learner's involvement in a task. While traditional classroom settings or linear digital materials (e.g., e-books) often track behavioural engagement through verbal contributions or interface interactions (Gettinger & Walter, 2012; Xu et al., 2021), this study adopted a measure specifically adapted to the affordances of gameplay. Unlike some prior studies, such as Lee (2023), which primarily rely on self-reported interest (emotional engagement), the current approach captures more direct, observable task behaviour within the game. In this study, coin collection served as a practical and quantifiable proxy for behavioural engagement. It required players to navigate the game world, return to previously visited areas, and interact with objects containing embedded lexical triggers, thereby reflecting their sustained attention, persistence, and willingness to interact with game elements. The act of collecting coins did not simply reflect player activity but represented goal-oriented task engagement with potential learning affordances. This novel, quantifiable measure of in-game engagement highlights the distinct learning dynamics afforded by games and underscores the need for further research using similar

direct behavioural indices to fully ascertain their incidental L2 vocabulary benefits.

Word-related factors: Saliency

Item saliency played a key role in vocabulary learning through gaming, particularly at the form-recall level. Saliency was embedded in the gameplay structure in this study through the functional relevance of each item for quest progression and success. Items varied in how essential they were to complete the task, which likely shaped attentional focus and influenced vocabulary outcomes. In this sense, saliency in gameplay may be more closely tied to task utility than to textual prominence, distinguishing it from saliency in reading-based learning environments (Chen & Truscott, 2010). Notably, similar patterns were found by Lee (2023), who demonstrated that saliency, defined in her study as the narrative importance of a word, was the strongest predictor of both vocabulary gain and retention in the narrative-driven game she used. She found that even low-frequency words were well-retained if they were central to the storyline (e.g., *miscarriage, midwife*), underscoring that functional and narrative salience can override raw frequency in driving lexical uptake, at least while playing games. Consistent with this, results for the form-recall GLMM showed participants in the current study demonstrated higher recall and recognition for items like *thwist* (potion) and *kavs* (coins), both of which were integral to completing the main or side quests. These findings suggest that saliency, whether narrative or task-based, increases learner attention and facilitates vocabulary learning, and is a key factor in explaining incidental vocabulary learning through gaming

Conclusion and future research

This study examined incidental vocabulary learning through gameplay in a COTS action-adventure game. Findings suggest that learners can acquire and retain a considerable amount of vocabulary at both the form-recall and meaning-recognition levels through gameplay, particularly when tasks are engaging and lexical items are functionally salient. Frequent gamers not only outperformed less experienced players, but their meaning-recognition knowledge proved significantly more durable, showing no retention decay over one week, and behavioural engagement, indexed by in-game parameters (coins collected), was linked to vocabulary gains.

These results contribute to growing evidence that commercial games can function as rich input sources for language learning outside the classroom. However, the study was exploratory in nature and thus presents some limitations. First, the small sample size limits the generalisability of the findings, and the absence of a control group prevents us from knowing whether some learning or guessing of nonwords could have occurred without the intervention. Further research should collect data from a larger sample and include a control group as baseline. In relation to the first limitation, the assessment of two word-knowledge aspects across testing times may have led to test interference. The contextualised presentation used in the form-recall test, in particular, may have inadvertently provided clues about the words' meanings that may have subsequently influenced performance on the meaning-recognition test. Future studies should consider alternative formats to assess form recall knowledge that further minimise meaning interference. Third, the use of a categorical operationalisation for gaming frequency limits the ability to precisely model the relationship between gaming frequency and vocabulary outcomes, and thus future research should consider this variable as a continuous measure. Finally, while the use of nonwords allowed a level of experimental control needed for initial exploration of vocabulary learning through gaming, future work could consider examining the learning of low-frequency real words to enhance ecological validity. Despite these limitations, this study offers promising initial evidence that commercial games can significantly support incidental L2 vocabulary learning at the form-recall and meaning-recognition level. Importantly, it offers novel theoretical and methodological insights that can shape future research in this area.

Funding

Open Access publication of this article was funded through a Transformative Agreement with University of Sheffield.

References

Apperley, T. H. (2006). Genre and game studies: Toward a critical approach to video-game genres. *Simulation & Gaming*, *37*(1), 6–23.

Baltra, A. (1990). Language learning through computer adventure games. *Simulation & Gaming*, *21*(4), 445–452.

Barclay, S., & Pellicer-Sánchez, A. (2021). Exploring the learning burden and decay of foreign language vocabulary knowledge. *ITL — International Journal of Applied Linguistics*, *172*(2), 259–289.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bawa, P. (2018). Massively multiplayer online gamers' language: Argument for an M-gamer corpus. *The Qualitative Report*, *23*(11), 2714–2753.

Bethesda Softworks. (2011). *The Elder Scrolls V: Skyrim Special Edition Creation Kit* [Computer software]. Bethesda Softworks.

Bethesda Softworks. (2016). *The Elder Scrolls V: Skyrim Special Edition* [Video game]. Bethesda Softworks.

Brown, C. (1993). Factors affecting the acquisition of vocabulary: Frequency and saliency of words. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary* (pp. 263–286). Ablex.

Calvo-Ferrer, J. R., & Belda-Medina, J. (2021). The effect of multiplayer video games on incidental and intentional L2 vocabulary learning: The case of Among Us. *Multimodal Technologies and Interaction*, *5*(12), Article 80.

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 92–101.

Chang, A. C.-S. (2019). Effects of narrow reading and listening on L2 vocabulary learning. *Studies in Second Language Acquisition*, *41*(4), 769–794.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalisation on incidental vocabulary acquisition. *Applied Linguistics*, *31*(5), 693–713.

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure: Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, *23*(1), 171–185.

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

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

Dixon, D. H. (2022). Linguistic environments of digital games. *CALICO Journal*, *39*(2), 192–214.

Gettinger, M., & Walter, M. J. (2012). Classroom strategies to enhance academic engaged time. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 653–673). Springer.

González-Fernández, B., & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics*, *41*(4), 481–505.

González-Fernández, B. (2025). How is vocabulary learnt? An acquisitional sequence of L2 word knowledge. *TESOL Quarterly*, *59*(2), 755–784.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring second-language vocabulary through reading. *Reading in a Foreign Language*, *11*(2), 207–223.

Hulstijn, J. H. (2012). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Wiley.

Juul, J. (2005). *Half-real: Video games between real rules and fictional worlds*. MIT Press.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399–436.

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research*, *19*(6), 687–706.

Lee, S.-M. (2023). Factors affecting incidental L2 vocabulary acquisition and retention in a game-enhanced learning environment. *ReCALL*, *35*(3), 274–289.

Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign-language words. *Current Biology*, *25*(4), 530–535.

McLean, S., Hogg, N., & Kramer, B. (2014). Estimations of Japanese university learners' English vocabulary sizes using the Vocabulary Size Test. *Vocabulary Learning and Instruction*, *3*(2), 47–55.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye tracking study. *Studies in Second Language Acquisition*, *38*(1), 97–130.

Peters, E. (2018). The effect of out-of-class exposure to English-language media on learners' vocabulary knowledge. *ITL: International Journal of Applied Linguistics*, *169*(1), 142–168.

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, *63*, 134–148.

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551–577.

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, *36*, 50–72.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior*, *19*(2), 227–241.

Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *42*(3), 525–549.

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, *47*(4), 479–496.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology: Section A*, *55*(4), 1339–1362.

Reinders, H., & Wattana, S. (2014). Can I say something? The effects of digital gameplay on willingness to communicate. *Language Learning & Technology*, *18*(2), 101–123.

Reinhardt, J. (2017). Digital gaming in L2 teaching and learning. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 202–217). Wiley.

Reinhardt, J. (2019). *Gameful second and foreign language teaching and learning: Theory, research, and practice* (1st ed.). Springer.

Reinhardt, J., & Thorne, S. L. (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). Routledge.

Reynolds, B. L. (2016). The effect of L2 learners' vocabulary size on the incidental acquisition of vocabulary through the reading of a novel. In Chiayi University Department of Foreign Languages (Ed.), *NCYU inquiry in applied linguistics: The 2015 issue — Theme: Linguistics applied across borders* (pp. 1–22). Crane Publishing Co., Ltd.

Reynolds, B. L. (2020). The effects of nonce words, frequency, contextual richness, and L2 vocabulary knowledge on the incidental acquisition of vocabulary through reading: More than a replication of Zahar et al. (2001) & Tekmen and Daloğlu (2006). *International Review of Applied Linguistics in Language Teaching*, *58*(1), 75–102.

Richards, J. C. (2015). The changing face of language learning: Learning beyond the classroom. *RELC Journal*, *46*(1), 5–22.

Rodgers, M. P. H., & Heidt, J. (2021). Levelling up comprehensible input and vocabulary learning. In J. Chik et al. (Eds.), *Pop culture in language education: Theory, research, practice* (pp. 126–138). Routledge.

Rodgers, M. P. H., & Webb, S. (2020). Incidental vocabulary learning through viewing television. *ITL — International Journal of Applied Linguistics*, *171*(2), 191–220.

Roskos, K., Burstein, K., & You, B.-K. (2012). A typology for observing children's engagement with e-books at preschool. *Journal of Interactive Online Learning*, *11*(2), 47–66.

Schmitt, N. (2008). Instructed second-language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching*, *50*(2), 212–226.

Sundqvist, P. (2013). The SSI model: Categorisation of digital games in EFL studies. *European Journal of Applied Linguistics and TEFL*, *2*(1), 89–96.

Sundqvist, P. (2019). Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary. *Language Learning & Technology*, *23*(1), 87–113.

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

Sundqvist, P., & Wikström, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System*, *51*, 65–76.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312.

Sylvén, L., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, *24*(3), 302–321.

Teng, F. (2019). The effects of context and word exposure frequency on incidental vocabulary acquisition and retention through reading. *The Language Learning Journal*, *47*(2), 145–158.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*(2), 130–163.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46–65.

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge.

Webb, S., & Chang, A. C.-S. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*(6), 667–686.

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Webb, S., & Rodgers, M. P. H. (2009a). The lexical coverage of movies. *Applied Linguistics*, *30*(3), 407–427.

Webb, S., & Rodgers, M. P. H. (2009b). Vocabulary demands of television programs. *Language Learning*, *59*(2), 335–366.

Xu, Y., Yau, J. C., & Reich, S. M. (2021). Press, swipe, and read: Do interactive features facilitate engagement and learning with e-books? *Journal of Computer Assisted Learning*, *37*(1), 212–225.

York, J. (2020). How to teach languages with "Among Us". *Ludic Language Pedagogy*, *2*, 269–283.

Zhou, N., & Yadav, A. (2017). Effects of multimedia story reading and questioning on preschoolers' vocabulary learning, story comprehension, and reading engagement. *Educational Technology Research and Development*, *65*(6), 1523–1545.

Appendix A. Original target items, nonword replacements, frequency, and saliency

The frequency of occurrence (number of times a target word appeared in an interactive object or quest objective) reported below was utilised exclusively as a criterion for initial target word selection, ensuring a baseline level of exposure (minimum two occurrences per item). We chose not to include word frequency as a predictor in the Generalised Linear Mixed Models (GLMMs) for two primary methodological reasons.

First, due to the interactive and open-world nature of the game, the actual exposure frequency varied idiosyncratically for each participant based on their navigation and unique in-game decisions. This, therefore, presents a methodological constraint to determine the exact frequency for each individual.

Alternatively, we opted to use the number of coins collected metric as a quantifiable proxy for behavioural engagement and active interaction, which captures the quality and attention given to the input, rather than relying on a static measure of potential lexical exposure.

Table A1. Original target items, nonword replacements, frequency, and saliency

Saliency category	*Nonword replacement*	*Original target item*	*Frequency (in quest)*
Quest	*Quares*	*Bear*	*3*
	Sprunt	*Chest*	*3*
	Kavs	*Coin*	*11*
	Thwist	*Potion*	*3*
	Blawked	*Spider*	*9*
Equipment	*Snarm*	*Arrow*	*4*
	Claits	*Boots*	*13*
	Swilge	*Bow*	*2*
	Slodge	*Dagger*	*3*
	Thests	*Gauntlet*	*12*
	Dwined	*Greatsword*	*3*
	Bepth	*Helmet*	*2*
	Relp	*Shield*	*2*
	Cheigh	*Sword*	*3*
	Phloob	*Warhammer*	*2*
Miscellaneous	*Smoch*	*Basket*	*12*
	Prauced	*Broom*	*2*
	Rhogged	*Bucket*	*7*
	Scact	*Door*	*10*
	Sweered	*Lockpick*	*3*

Appendix B. Vocabulary tests (form-recall & meaning-recognition)

This appendix provides some examples of vocabulary tests used to assess participants' incidental vocabulary learning. Both the form-recall and meaning-recognition tests comprised 20 target nonword items and 10 real-word filler items, totalling 30 items per test. A full list of target nonwords and their corresponding original items is available in Appendix A All tests were administered online via Google Forms.

Form-recall and meaning-recognition test formats

All tests' formats, including form-recall and meaning-recognition of target nonwords and distracters are provided in this section. As stated in the methodology, the tests were distributed online through Google Form. However, the form formats were disorganised when converted into PDF. Therefore, the questions were re-typed into word document formats for a more organised composition.

Form-recall questions

Write the appropriate word in the space to complete the sentences. The word begins with the letter given. It is okay if you do not know the answer to the questions.

Example

0. *I had to cover my eyes when coming into the room, as it was too b_____ for my eyes.*
Answer: bright

1. A p_____ is a large heavy hammer with a long handle and usually spikes. It is used as a weapon.
   Answer: _____
2. I use a c_____ which is a long sharp blade to fight my enemies.
   Answer: _____
3. They use a s_____ to shoot arrows.
   Answer: _____

Meaning-recognition

Choose the correct meaning of the word in [square brackets]. If you do not know the answer, please choose "I don't know."

1. He is carrying a [phloob].
   ☐ Warhammer ☐ Sword ☐ Broom ☐ Mace ☐ I don't know
2. He attacks the enemy with a [cheigh].
   ☐ Warhammer ☐ Battleaxe ☐ Sword ☐ Broom ☐ I don't know
3. He drops a [swilge].
   ☐ Bow ☐ Sword ☐ Arrow ☐ Broom ☐ I don't know

Note: Square brackets were used in the meaning recognition part since bold or italics were not available on Google Form.

Appendix C. Background questionnaire (summary)

This appendix provides a summary of the questionnaire used to collect participant background and gaming habit data. The questionnaire was administered online via Google Forms and comprised the following sections:

I. *Biographical data*

This section collected basic demographic information and L1 details.

- **Example question:** "Age" (short answer)
- **Example question:** "Gender" (multiple choice: Male, Female, Prefer not to say)

II. *Gaming habits*

This section gathered data on participants' general gaming frequency and preferred genres.

- **Example question:** "How often do you play games on average per week in the past 3 months?" (multiple choice: 0 hours, less than 3 hours, 3–5 hours, more than 5 hours)
- **Example question:** "Which game genres do you play?" (checkbox list: Adventure, Autochess, Battle Royale, etc.)

III. *Gaming and language learning beliefs*

This section assessed participants' perceptions regarding the utility of gaming for language and vocabulary acquisition.

- **Example question:** "Gaming is a useful vocabulary learning resource." (Likert scale: Strongly Agree to Strongly Disagree)

IV. *Vocabulary size*

This section prompted participants to complete an external vocabulary test and report their scores.

- **Example instruction/question:** "Please complete the Vocabulary Size Test. Click on this link: [https://www.lextutor.ca/tests/vst/]. Please paste the scores shown in the white box on the left-hand side of the screen here." (short answer)

V. *About the game*

This section determined participants' prior familiarity with the target game, *The Elder Scrolls V: Skyrim Special Edition*.

- **Example question:** "Have you played 'The Elder Scrolls V: Skyrim Special Edition' before?" (multiple choice: Yes, No)

Appendix D. GLMM model selection outputs

This appendix provides the detailed R output for the Generalised Linear Mixed-Effects Models (GLMMs) fitted in this study. For each dependent variable and time point, a model selection table is presented, showing the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood, deviance, number of parameters (npar), and results from Chi-squared ($\chi 2$) tests for nested model comparisons. Marginal R^2 (R^2m) and Conditional R^2 (R^2c) values are also included for each model.

D.1 *Form-recall model comparison*

For the form-recall, the full model was selected as the optimal model, as no significant improvement in fit was achieved by removing predictors.

Table D1. Form-recall model comparison

Model	Fixed effects formula (key interactions)	npar	AIC	logLik	$\chi 2$	df	*p*
Full model	TestTime × (VS+Gamer+Coins+ItemCat)	14	560.74	−266.37	—	—	—
Step 1	Full Model — (TestTime×Gamer)	13	559.29	−266.65	0.5497	1	0.4584
Step 2	Step 1 — (TestTime×VocabSize_scaled)	12	558.09	−267.05	0.7954	1	0.3725
Step 3 (Optimal)	Step 2 — (TestTime×ItemCat)	10	558.02	−269.01	3.933	2	0.1399

D.2 *Meaning-recognition model comparison*

Table D2. Meaning-recognition model comparison

Model	Fixed effects formula (key interactions)	npar	AIC	logLik	$\chi 2$	df	*p*
Full model	TestTime × (VS+Gamer+Coins+ItemCat)	14	824.69	−398.34	—	—	—
Step 1	Full Model — (TestTime×ItemCat)	12	821.06	−398.53	0.3765	2	0.8284
Step 2	Step 1 — (TestTime×VocabSize_scaled)	11	819.49	−398.74	0.4264	1	0.5138
Step 3	Step 2 — (ItemCat)	9	815.56	−398.78	0.0694	2	0.9659
Step 4 (Optimal)	Step 3 — (TestTime×Coins_scaled)	8	817.04	−400.52	3.4843	1	0.06195

Address for correspondence

Iwarin Suprapas
School of English, Jessop West
University of Sheffield
1 Upper Hanover Street
Sheffield, S3 7RA
United Kingdom
isuprapas1@sheffield.ac.uk
https://orcid.org/0009-0001-8359-3985

Co-author information

Beatriz González-Fernández
School of English, Jessop West
University of Sheffield
b.gonzalez-fernandez@sheffield.ac.uk

Publication history

Date received: 29 July 2025
Date accepted: 11 February 2026
Published online: 2 March 2026