Journal of Research in Applied Linguistics
ISSN: 2345-3303 – E-ISSN: 2588-3887 – https://rals.scu.ac.ir
Published by Shahid Chamran University of Ahvaz

Please cite this paper as follows:

Rasti-Behbahani, A. (2025). Involvement load hypothesis vs. technique feature analysis in digital game-based vocabulary learning activities. *Journal of Research in Applied Linguistics*, *16*(1), 55-70. https://doi.org/10.22055/RALS.2024.46889.3300

Shahid Chamran University of Ahvaz

Research Paper

Involvement Load Hypothesis vs. Technique Feature Analysis in Digital Game-Based Vocabulary Learning Activities

Amin Rasti-Behbahani

Center for Preparatory Studies, Dhofar University, Salalah, Oman; *aminrasti@du.edu.om*

Received: 11/05/2024 Accepted: 24/07/2024

Abstract

Frameworks like the involvement load hypothesis (ILH) and technique feature analysis (TFA) have proved acceptable success rates in predicting precisely the effectiveness of paper-based incidental vocabulary learning (IVL) activities. However, the prediction precision of these frameworks has not been yet studied and compared in digital game-based vocabulary learning (DGBVL) activities. To do so, 37 male and 29 female Omani-Arabic speakers (18-20 years old) were randomly assigned to a DGBVL activity that induced low, moderate, or high levels of involvement load. The results of the pretest showed that the 20 selected target words, or names of concrete inanimate objects, were new to the participants. After completing the DGBVL activities, the results showed that (1) receptive and productive knowledge gain was enhanced by DGBVL activities differentially depending on the activity type, (2) TFA was more accurate than ILH in prediction, and (3) the prediction of both frameworks was not error-free. Pedagogically, prospective teachers and researchers are recommended to use TFA for designing effective DGBVL activities; however, they must be aware that relying only on the collective score of TFA may not always lead to desired outcomes.

Keywords: Involvement Load Hypothesis; Technique Feature Analysis; Digital Game-Based Learning; Incidental Vocabulary Learning; Video Games.

1. Introduction

There have been many attempts to design effective incidental vocabulary learning (IVL) activities from glossing to digital game-based vocabulary learning (DGBVL; Rasti-Behbahani & Shahbazi, 2022) because IVL gains can improve the quality of the lexicon and can add information that may not be learned intentionally (Webb, 2020). However, the success rates of the proposed interventions are relative and change in different contexts (Yanagisawa & Webb, 2022). Then, the necessity for a comprehensive framework that can predict the success rate of teacher-designed activities becomes evident. To achieve this aim, Laufer and Hulstijn (2001) borrowed the depth of processing hypothesis from psychology.

The depth of processing hypothesis assumes that vocabulary items that are processed qualitatively and quantitatively richer will leave more profound traces on memory and will be retained longer; hence, vocabulary learning interventions that promote richer processing can enhance IVL gains effectively (Luafer & Hulstijn, 2001). Based on this, Laufer and Hulstijn (2001) introduced the involvement load hypothesis (ILH), a cognitive-motivational and trifold framework that assisted with measuring the impact and success rate of IVL activities. Although ILH was supported by many studies, its indexing method and predictive power were criticized (Jahangiri & Abilipour, 2014; Folse, 2006; Keating, 2008; Kim, 2011; Martínez-Fernández, 2008), which led to its expansion, for example, technique feature analysis (TFA; Nation & Webb, 2011), and other modifications (Yanagisawa & Webb, 2022; Zou, 2017).

Among all the proposed frameworks, ILH and TFA are the most popular. However, there is a debate on their accuracy in indexing methods for predicting the success rate of IVL activities. Despite recent attempts (Ehsani & Karimi, 2022; Gohar et al., 2018; Hu & Nassaji, 2016; Khoshsima & Eskandari, 2017), the scope of the findings is limited to only

traditional paper-and-pencil IVL activities, whereas DGBVL activities have been left unexplored while integration of new technologies into learning contexts is inevitable (Díez-Arcón & Martin-Monje, 2023; Traxler et al., 2023). Therefore, this study constitutes the first attempt to compare the precisions of the ILH and TFA frameworks in indexing the success rate of DGBVL activities.

2. Literature Review

2.1. Involvement Load Hypothesis (ILH)

ILH was proposed by Laufer and Hulstijn (2001) to provide a framework that can measure the success rate of IVL activities. They assumed that an IVL activity that can provide more opportunities for word processing both qualitatively and quantitatively, would lead to greater and more durable IVL knowledge gains. Moreover, they added that the effectiveness of the activities can be moderated through three components: need, search, and evaluation.

The need is the motivational component of the hypothesis and refers to the extent of a learner's willingness to learn a word. It is moderate (score 1) when a learner is asked to learn an L2 word by an external agent like a teacher; however, it is strong (score 2) when a learner feels the necessity to learn the word. Search is moderate (score 1) when a learner looks for or tries to infer the meaning of an L2 word. It is strong (score 2) when a learner looks for the word form (Nation & Webb, 2011). Evaluation is moderate (score 1) when a learner compares the meanings of L2 words to choose the most suitable one for a context, e.g., fill-in-the-blank activities, and is strong (score 2) if a learner must decide on the L2 form, e.g., sentence or composition writing. "The combination of these factors with their degrees of prominence constitutes the *task-induced involvement load*. The higher the involvement load, the more effective the vocabulary task is" (Khoshsima & Eskandari, 2017, p. 51).

Not all studies in the literature enhanced or found full support for the ILH main hypothesis (Folse, 2006; Jahangiri & Abilipour, 2014; Keating, 2008; Kim, 2011; Martínez-Fernández, 2008). They mainly argued against the collective effect of components and added that components can contribute to IVL gains separately; that the weight of evaluation component contribution is more than that of the other two components (Keating, 2008; Kim, 2011); and that the important factors such as frequency and retrieval (Folse, 2006) that can enhance IVL were not considered. Hence, due to the importance of ILH, three main modifications have been proposed to improve the indexing and ultimately predictive power of ILH.

Reflecting on the critical contribution of evaluation in IVL, Zou (2017) proposed a modified version of ILH in which the evaluation component has four levels: absent, moderate, strong, and super strong. Evaluation is moderate in activities such as fill-in-the-blank because such activities do not involve information-processing techniques such as chunking, hierarchical organization, or pretask planning. However, in sentence writing and composition writing, evaluation is strong and super strong respectively because they involve information-processing techniques either partially or fully. Karami and Esrafili's (2021) study supported Zou's augmented ILH framework and enhanced emphasis on the contribution of evaluation.

Yanagisawa and Webb (2022) proposed the ILH Plus framework, which is a byproduct of a comprehensive meta-analysis study exploring the contribution of ILH to the prediction of IVL gains. Their framework emphasizes the individual effect of components and shows that depending on conditions defined by variables, such as test type, test intervals, types of tasks, frequency, and mode (spoken or written), the contribution of different ILH components varies. Therefore, they introduced two new formulas for indexing task-induced levels of involvement load (to read more, see Yanagisawa & Webb, 2022).

Finally, there is the TFA (see Table 5) checklist by Nation and Webb (2011). This checklist includes many factors that have been found effective on IVL. They justified the necessity for a new quantifying method due to the incomprehensiveness of ILH and a lack of recognition of previous findings and important factors that may affect IVL (Nation & Webb, 2011). The checklist is comprised of five main categories: motivation, noticing, retrieval, generation, and retention. In total, there are 18 components, and each component receives a score of either 0 or 1. If a component is triggered by the vocabulary learning activity, it is assigned a score of 1; hence, the highest score of TFA is 18.

The predictive power of ILH and TFA, but not ILH Plus and the augmented ILH framework, was compared recently. Hu and Nassaji (2016) compared the predictive power of ILH and TFA by comparing the performance of 96 Taiwanese college-level participants who tried to acquire the receptive knowledge of 14 academic words. Four different vocabulary learning tasks (text + multiple-choice questions, text + definition selection, reading + fill-in-the-blanks, and text + rewording sentences) were designed so that each induced a different level of involvement based on TFA and ILH. After comparing the during and immediate after-the-task performances of the participants, the results showed that TFA predicted the effectiveness of vocabulary learning tasks more precisely than ILH. Khoshsima and Eskandari (2017), who compared different tasks with different indexes in both TFA and ILH frameworks, provided support for the superiority of TFA explanatory power over ILH. Immediate receptive and productive knowledge of ten infrequent academic words were measured by the vocabulary knowledge scale in this study. In another attempt with different task designs (sentence writing, composition writing, and text + comprehension questions) and TFA and ILH scores, Gohar et al. (2018) studied the predictive powers of the two frameworks. After administering an immediate posttest that measured the receptive knowledge of ten low-frequency academic words, they found that TFA had a better explanatory capability than ILH, but only for variances between pretest and posttest and not for during the task performance. Finally, Ehsani and Karimi (2022) assigned 110 advanced EFL learners to five vocabulary learning tasks (text + true/false, text + definition selection, text + fill-in-the-blanks, text + rewording sentences, and text + comprehension questions) for learning 10 pseudowords including verbs, adjectives, and nouns. The tasks differed in their structure and ILH and TFA scores. After an immediate and short-delayed posttest (1 week) for measuring the receptive knowledge of the target words, they found that TFA had predicted the outcome of both immediate and delayed posttests more precisely than ILH had done.

However, San Mateo-Valdehíta and Criado de Diego (2021), who studied the effect of three tasks inducing different ILH and TFA loads on the immediate receptive and productive acquisition of 12 infrequent and 12 pseudo-Spanish words, found that ILH had better explanatory power than TFA in foreseeing the impacts of tasks. The superiority of TFA to ILH has been related to the number of factors that it analyzes (Hu & Nassaji, 2016; Jahangiri & Abilipour, 2014). It is notable that as there are not enough studies on the predictive power of ILH Plus and the augmented ILH framework, they are not compared in this study.

2.2. Digital Game-Based Vocabulary Learning (DGBVL)

DGBVL, or implementing computer games for teaching vocabulary (Rasti-Behbahani & Shahbazi, 2022), has been widely reported as one of the most important informal language learning activities (Amengual-Pizarro, 2024) that can enhance IVL effectively. For example, Camacho-Vásquez and Ovalle (2019) found that DGBVL activities were capable of enhancing bachelor students' knowledge of infrequent academic words. Khazaie and Mashhadi (2020) found that integrating even non-English mobile games into L2 vocabulary learning pedagogy can also have a significant effect on vocabulary knowledge gains. Rasti-Behbahani and Shahbazi (2022) found that adventure DGBVL activities have the potential to enhance IVL, especially the productive recognition of form-meaning. Calvo-Ferrer and Belda-Medina (2021) concluded that multiplayer games, in which learners have opportunities for word use, can enhance IVL more effectively than games with opportunities limited to encountering new words in the game without any previous input. Sundqvist and Nilsson (2022) found that investing time in multiplayer games can significantly impact IVL enhancement. Yang and Jeaco (2023) found that their concordance-based game supported lexicogrammatical knowledge of English words, phrases, collocations, and many other aspects of word knowledge. The outcomes of this study were supported by Pham (2023) who reported that by adding game elements, such as a scoreboard, reward system, and so on, the participants' uptake of infrequent words was noticeably increased. Overall, the impact of DGBVL activities on improving the quantity and quality of vocabulary learning is both significant and undeniable (Vnucko & Klimova, 2023).

Digital games can enhance IVL gains because firstly they improve the learning condition by maximizing engagement and motivation, reducing anxiety, promoting learners' self-confidence, encouraging risk-free attempts, enhancing authenticity, and making DGBVL activities goal-oriented (Peterson et al., 2022; Vassilieva & Drugov, 2019; Vnucko & Klimova, 2023). Next, the inherent elements of digital games, such as audiovisual effects, story, challenge, competition, sense of winning, and immersion (Alexiou & Schippers, 2018), promote opportunities for learning (Hamari et al., 2014) and IVL gains (Peterson et al., 2022). Owing to these factors, opportunities for vocabulary learning, such as repetition, instantiation, dual encoding, interactivity, and feedback, can be enhanced, which can lead to effective IVL gains in DGBVL activities (Rasti-Behbahani, 2021).

DGBVL activities can be effective "when [they are] implemented in conjunction with carefully designed supplementary materials" (Peterson et al., 2022, p. 81). This idea was supported by Calvo-Ferrer and Belda-Medina (2021) and Vinter et al., (2022) who found that an age-appropriate and modified-input DGBVL activity enhanced vocabulary learning more effectively. Hence, it is valid to infer that modifying DGBVL activities by adjusting their levels of involvement load can enhance IVL gains. However, choosing the best framework, for example, ILH or TFA, for modifying DGBVL activities is important because the frameworks help with the decisions on activity designs. Also, an imbalanced design may lead to unsuccessful attempts for IVL gain. Conversely, the available literature cannot help with choosing the best framework because firstly, most of the relevant studies were conducted in a paper-and-pencil context, whereas DGBVL activities were ignored. Secondly, they have focused only on comparing the explanatory power of the frameworks in the acquisition of receptive knowledge while productive knowledge is widely neglected. Finally, delayed posttests are either absent or administered only a week later instead of a minimum of 3 weeks later (Schmitt, 2010). Including a long-delayed posttest is essential because, in the long run, the effect of many factors that either enhance or hamper ILH or TFA will diminish drastically (Yanagisawa & Webb, 2022). Hence, this study was an attempt to expand the comparative attempts to find the most accurate framework and to fill in those gaps in the literature by answering the following questions:

- To what extent do the DGBVL activities contribute to IVL gains?
- To what extent could TFA or ILH predict the DGBVL activities contributions?

3. Methodology

This empirical study had a quasi-experimental design, and the quantitative data were collected through pretests and posttests before and after treatments respectively.

3.1. Participants

Through convenient sampling—by visiting English classes and asking English learners to participate in this study—66 randomly selected Omani-Arabic speakers (37 males, 29 females; 18-20 years old) participated in this study. They were basic level (level 1 out of 3) students in the foundation program at Dhofar University. Their English proficiency, evaluated by an online standard test called the Cambridge English Proficiency Test (CEPT), was A2 (CEFR). Although proficiency level has a short-term effect on ILH contribution to IVL and its effect diminishes significantly later (Webb et al., 2023), 1,000 (1K) and 2,000 (2K) most frequent English words on the Updated Vocabulary Levels Test (uVLT; Webb, Sasao, & Balance, 2017) were administered to check the participants' vocabulary size and ensure their homogeneity in word knowledge, and understanding of a walkthrough, or game-guide, instructions. Their average in 1K was 83.43% (M = 25.03, SD = 3.12), in 2K was 73.53% (M = 22.06, SD = 3.01), and overall was 78.46% (M = 23.54, SD = 3.41). This suggests that their vocabulary knowledge was high enough (González-Fernández & Schmitt, 2020) to understand the walkthrough text because 97% of the words in the walkthrough were between 1K and 2K. All the participants had played games on PC and knew how to control a PC game by mouse and keyboard. Also, none of them had played the game that was used in this study before.

3.2. Materials

3.2.1. The Digital Game

A commercial adventure digital game called Azada: In Libro was selected for the current study. In this game, a gamer must stop an evil magician from conquering Azada. There were two reasons why I chose a commercial adventure digital game: Firstly, it makes learning meaningful by boosting intrinsic motivation (Pivec et al., 2003); secondly, adventure games have interesting puzzles, deep dialogues, and engaging stories that boost motivation (Chen et. al., 2020) and support vocabulary learning.

In this game, puzzles are object-based; that is, a gamer must solve game puzzles by finding specific items and using or placing them in their proper locations (see Figure 1). The game is controlled by mouse clicks.

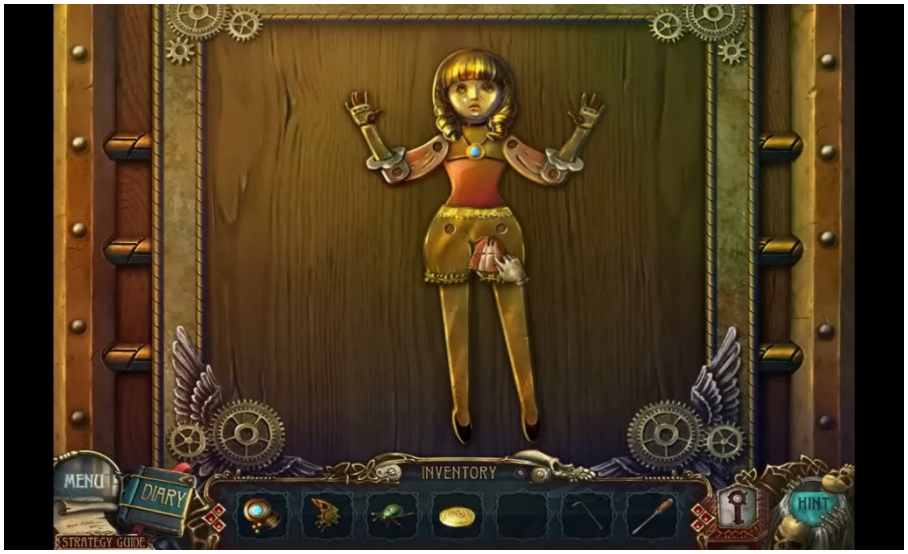

Figure 1. Object-Oriented Puzzle: The Doll's Skirt Is Being Installed to Unlock a Safe

3.2.2. Walkthrough

After downloading a walkthrough, or game guide, from the game's official Website, its first two chapters were modified for this study. Firstly, pictorial clues were removed; secondly, the sentences were numbered; finally, target word fonts were in bold but the participants were not informed of the reason. IVL can be enhanced effectively if target words can attract readers' attention (Boers, 2022) and word-form focus can play a significant role in word learning (Hu & Nassaji, 2016). Time-on-task duration and the negative effects of extensive interaction could have been reduced by providing the walkthrough (Rasti-Behbahani & Shahbazi, 2022).

All the participants needed to read the walkthrough and follow the steps in order and as closely as possible to finish the chapters while they were playing the game.

3.2.3. Target Words

After checking the frequency profile (on https://www.lextutor.ca/vp/comp/) of the walkthrough text and assuming lower chances of familiarity with infrequent words, 20 low to infrequent concrete nouns of inanimate objects were selected (see Table 1). Unfamiliarity with the target words was later confirmed (see Table 9). To select them, I considered four criteria: Firstly, without knowing them, the participants could not finish critical steps in the walkthrough. Secondly, they did not occur more than twice throughout the walkthrough to control the effect of multiple exposures (Folse, 2006; Rasti-Behbahani & Shahbazi, 2022). Thirdly, they had to be nouns (names of inanimate objects) because the acquisition of nouns is quicker than adjectives, verbs, and adverbs, and this could have reduced the learning load and word difficulty (Sökmen, 1997). Fourthly, their frequency band had to be K4 or higher to reduce the chances of A2-level learners' familiarity with the target words (Milton, 2010). Notably, the target words were also pilot-tested by asking ten other A2-level nonparticipant students who were studying at level 1 to translate the target words into Arabic. All the words were new for the pilot test takers:

Table 1. *Target Words*

	Freq.	K	Target Words	Freq.	K
Cord	2	4	Bracelet	2	7
Rod	2	4	Pouch	2	7
Rack	2	4	Spade	2	7
Cylinder	2	5	Clover	2	8
Debris	2	5	Revolver	2	8
Purse	2	5	Spout	2	8
Tub	2	5	Cyclone	2	10
Mattress	2	5	Faucet	2	13
Crate	2	6	Pestle	2	14
Scroll	2	6	Anklet	2	16

3.2.4. Activity Design

After modifying the walkthrough, I made three different versions of it. In the first version, I added a single Arabic definition for each target word on the right margin (meaning group; see Tables 2, 3, and 4 for step 8 in the walkthrough for each group). The definitions were placed on the page margins because learners prefer marginal glosses to other locations (Boers, 2022):

Table 2. *Meaning-G Walkthrough Text*

8. Put the **faucet** on the bottle.	صنبور

*صنبور means faucet.

In the second version, I added three different Arabic definitions for each target word on the right margin (choice group). It is worth mentioning that each definition was also repeated twice to control their effects (Folse, 2006):

Table 3. *Choice-G Walkthrough Text*

	السلك
8. Put the **faucet** on the bottle.	لفيفة طومار
	صنبور

*السلك means cord; لفيفة طومار means scroll; and صنبور means faucet.

There was no definition in the third version (recall group) on the walkthrough page margins. Instead, the 20 target words along with their respective Arabic definitions and a sample sentence for each word were listed and delivered on a separate paper to the participants:

Table 4. *Recall Walkthrough Text*

8. Put the **faucet** on the bottle.

3.3. Task Indexing

All the three activities were indexed based on both ILH and TFA frameworks by the researcher. They were later checked and approved by two applied linguistics professors.

3.3.1. ILH

Based on ILH, the meaning activity was indexed as follows: need = 2, search = 0, evaluation = 0. The need was strong because the participants needed to know the meanings of the words to complete the game. However, because the meanings were provided, they did not need to search for them or do any evaluation.

The choice activity, in addition to a strong need, triggered search and evaluation (need = 2, search = 1, evaluation = 1). However, the search and evaluation were not strong because the participants tried to infer the correct meaning, which triggered a moderate search and evaluation (Laufer & Girsai, 2008).

The recall activity triggered a strong need and a moderate search, but not the evaluation (Need 2, search 1, and evaluation 0). This task triggers the search because it persuades an attempt to find the meaning through inferring and recalling (Zou, 2012). Also, attempts to recall tidbits about the target words and the words around them can be inferred as a search for meaning (Folse, 2006). Moreover, this ILH combination is rarely studied (Yanagisawa & Webb, 2021).

Overall, based on the ILH framework, the order of the effectiveness of the tasks on IVL gains is as follows: choice (4) > recall (3) > meaning (2).

3.3.2. TFA

According to TFA, the meaning activity triggered the components 2, 4, 5, 12, 15, and 18 in TFA (Index 6) and induced a low level of involvement load. The choice activity activated a few more components in TFA: 2, 4, 5, 7, 10, 12, and 18 (Index 7) and increased the level of involvement load induction to a moderate level. The recall activity triggered

even more components than the previous two activities including 2, 4, 5, 7, 9, 10, 11, 12, 15, and 18 (Score 10) which led to a higher level of involvement load.

Adventure digital games are motivating so component 2 was activated by all the activities (Pivec et al., 2003); the target words were written in bold so all the tasks activated component 4. Due to the implementation of vocabulary learning techniques, component 5 was triggered by all the activities; the target words were exposed twice in different sentences, so components 12 was scored for all the activities; and component 18 was activated because words like left and right, were not taught spontaneously (Nation & Webb, 2010). However, component 15 was triggered exclusively by meaning and recall activities because having Arabic definitions could guarantee a successful form-meaning link. Components 7 and 10 were activated by the choice and recall activities because of, at least, two encounters for each target word form (recall) or its meaning (choice). In the recall activity, component 9 was scored because Arabic definitions were not placed on the page margins, and component 11 because the participants had seen the target words a day before their main activity.

Then, based on the TFA frameworks, the order of the effectiveness of the activities on IVL is as follows: recall (10) > choice (7) > meaning (6).

Table 5. *ILH and TFA Scores*

	ILH Criteria	Scores		
		Meaning	*Choice*	*Recall*
1	Need	2	2	2
2	Search	0	1	1
3	Evaluation	0	1	0
	Maximum Scores	2	4	3
	TFA Criteria	Scores		
		Meaning	*Choice*	*Recall*
	Motivation			
1	Is there a clear vocabulary learning goal?	0	0	0
2	Does the activity motivate learning?	1	1	1
3	Do the learners select the word?	0	0	0
	Noticing			
4	Does the activity focus attention on the target words?	1	1	1
5	Does the activity raise awareness of new vocabulary learning?	1	1	1
6	Does the activity involve negotiation?	0	0	0
	Retrieval			
7	Does the activity involve retrieval of the word?	0	1	1
8	Is it productive retrieval?	0	0	0
9	Is it recall?	0	0	1
10	Are there multiple retrievals of each word?	0	1	1
11	Is there spacing between retrievals?	0	0	1
	Generation			
12	Does the activity involve generative use?	1	1	1
13	Is it productive?	0	0	0
14	Is there a marked change in context that involves the use of other words?	0	0	0
	Retention			
15	Does the activity ensure successful linking of form and meaning?	1	0	1
16	Does the activity involve instantiation?	0	0	0
17	Does the activity involve imaging?	0	0	0
18	Does the activity avoid interference?	1	1	1
	Maximum score	6	7	10

3.4. Instruments

3.4.1. Updated Vocabulary Level Test (uVLT)

On uVLT (reliability = 0.93), first and second frequency bands (1K and 2K), comprising 30 receptive recognition questions each, were administered. In this test, the participants needed to match a definition with one of the six alternatives (see Table 6). A paper-based version of uVLT was administered and there was no time limit for completing the test:

Table 6. *uVLT Question 1*

	boy	rent	report	size	station	thing
How big or small something is						
Place buses and trains go						
Young man						

3.4.2. Achievement Tests

Two paper-based achievement tests measured the participants' receptive and productive knowledge gains. In the receptive test, the participants needed to recognize the proper Arabic definitions for the bolded English forms in sentences (see Table 7). In the productive test, the participants had to fill in the blanks by recalling the proper English form of the target words by using the sentences and cues provided (see Table 8). There were 20 questions in each test and each correct answer was scored one. The tests were used in two stages: pretest and delayed posttest with no time limit for completing them. Administering them as a pretest served two purposes: It could reveal whether, firstly, the selected target words were new, and secondly, the DGBVL activities were effective in learning the target words. Notably, in the receptive test, options were chosen among the target words to reduce the chances of lucky guess; the productive test was administered first, and target words were randomly distributed in each test to avoid the effect of learning from the earlier test:

Table 7. *Receptive Test: Question 3*

3. The murderer used an old **revolver**.
a) خرطوشة b) شاکوش c) رصاصة d) مسدس

* خرطوشة means cartridge; شاکوش means hammer; رصاصة means bullet; and مسدس means revolver

Table 8. *Productive Test: Question 20*

20. In the past, **an**_____ was worn on feet by Asian women to show that they were married.

Tests were pilot-tested by administering them to 11 unrecruited A2-level 1 English learners from the same training center. The results of Cronbach's alpha indicated an acceptable level of internal consistency (Taber, 2018): receptive test = 0.79, and productive test = 0.81.

3.5. Procedure

This study had three stages. In the first stage, after contacting the teenage English learners personally to find volunteers, asking them to sign a consent letter, and informing their parents of their participation, volunteers sat for the paper-based uVLT. They also sat for paper-based receptive and productive tests (pretest) to check their familiarity with the target words. After ensuring their acceptable vocabulary knowledge, they were randomly given meaning (N = 22), choice (N = 22), and recall (N = 22) activities.

In the second phase, a day before the main DGBVL activities, the word list including the target words, their Arabic definitions, and sample sentences was distributed to the recall participants. They had 10 min to read it and were not informed of the purpose of the word list. They returned the word list after they finished reading it. Next, in a computer lab, all the participants were instructed on how to play the game and complete the DGBVL activities. All the participants tried to complete the first and second chapters of the game by following the walkthrough steps. I was there during their play and helped them with technical issues. The participants spent between 86 and 98 min on their tasks mostly because of the cut scenes that were video clips showing more details on the story and users were not able to skip them.

In the third stage, I administered the delayed posttests four weeks later because firstly, I wanted to measure the efficiency of the DGBVL activities in long-run and their effects on word retention; also, if the effect of a vocabulary

learning activity can last for at least three weeks, that learning is stable and durable (Schmitt, 2010). Both tests were administered at the same time. Finally, no time limit was set for completing the delayed posttests.

3.6. Data Analysis

To answer the research questions, I analyzed the elicited quantitative data by SPSS (version 26). The quantitative data were analyzed by one-way analysis of variance to find if the DGBVL activities could enhance IVL gains and to understand which group in which test had a better performance, LSD post-hoc test to compare the individual effects of activities and their differences, and a hierarchical multiple regression analysis to compare the precision of both frameworks in predicting the outcome of each DGBVL activity (Calvo-Ferrer & Belda-Medina, 2021).

4. Results

Before analyzing the data, Shapiro-Wilk and Levene homogeneity of variance tests were run, and the results indicated that variances were homogeneous, and the assumptions for regression analysis, that is, linearity, homoskedasticity, independence of errors, normality, and independence of independent variables were also met:

Table 9. *ANOVA (Pretest Scores)*

	Task	Mean	Standard Deviation	*F*	*p* Value
Receptive (Pretest)	Meaning	1.13	0.63	1.70	.19
	Choice	1.45	0.59		
	Recall	1.45	0.73		
Productive (Pretest)	Meaning	1.13	0.71	0.09	.91
	Choice	1.09	0.75		
	Recall	1.04	0.65		

To answer the research questions, firstly, the homogeneity of participants' knowledge of the target words before the main task was checked. The results of ANOVA, which was run on pretest scores (see Table 9), showed that there was no significant difference ($p > 0.05$) in the participants' knowledge of target words. Also, their mean scores suggested that they had little knowledge of the target words:

Table 10. Descriptive Statistics (Gain Scores)

	Task	Mean	*SD*
Receptive	Meaning	10.40	1.94
	Choice	7.68	2.53
	Recall	16.22	2.44
Productive	Meaning	11.77	2.06
	Choice	8.86	2.58
	Recall	16.63	1.89

Table 11. *ANOVA (Gain Scores)*

		df	Mean Square	F	p
Receptive	Task	2	419.15	77.67	0.00
	Within Groups	64	5.39		
	Total	66			
Productive	Task	2	339.28	69.95	0.00
	Within Groups	64	4.85		
	Total	66			

To answer the first question, a one-way ANOVA was conducted on the participants' gain scores to examine the differences between participants in their IVL gains across the three DGBVL activities. Table 10 shows that the participants with the recall activity had the best performances in both receptive ($M = 16.2273$) and productive (16.6364) tests. Next, surprisingly, the participants with the meaning activity performed better in both receptive ($M = 10.4091$) and productive (11.7727) tests than those with the choice activity ($M_{Receptive} = 7.6818$, $M_{Productive} = 8.8636$). According to

ANOVA results (see Table 11), the differences are also statistically significant ($F_{Receptive}$ = 77.677, $F_{Productive}$ = 69.957; p = 0.000):

Table 12. *LSD Post-Hoc Multiple Comparisons Across Tasks*

Tests	Activity	Between-Tasks	Mean Difference	Std. Error	*p*
Receptive (Gain Scores)	Meaning	Choice	2.72	.70	.00
		Recall	-5.81	.70	.00
	Choice	Meaning	-2.72	.70	.00
		Recall	-8.54	.70	.00
	Recall	Meaning	5.81	.70	.00
		Choice	8.54	.70	.00
Productive (Gain Scores)	Meaning	Choice	2.90	.66	.00
		Recall	-4.86	.66	.00
	Choice	Meaning	-2.90	.66	.00
		Recall	-7.77	.66	.00
	Recall	Meaning	4.86	.66	.00
		Choice	7.77	.66	.00

To compare the individual effects of activities and their differences, an LSD post-hoc was then conducted. Table 12 shows that each activity contributed to both receptive and productive IVL gains differentially, and their effects are significantly different ($p < 0.05$):

Table 13. *Assumptions Check Against the LSD Results*

Tests	Comparisons	Mean Differences	ILH Assumption	TFA Assumption
Receptive	Meaning > Choice	2.72	✗	✗
	Recall > meaning	5.81	✓	✓
	Recall > Choice	8.54	✗	✓
Productive	Meaning > Choice	2.90	✗	✗
	Recall > meaning	4.86	✓	✓
	Recall > Choice	7.77	✗	✓

Table 13 shows that one of the comparisons was not consistent with ILH or TFA assumptions across both receptive and productive tests. That is, the participants with the meaning activity performed better than those with the choice activity. Moreover, we can see that other comparisons are consistent with the TFA assumptions, whereas only one of them is consistent with ILH assumptions. Hence, according to Table 13, the predictive power of TFA is more accurate than ILH:

Table 14. *Components Distribution Percentage in ILH*

Activity	Need	Search	Evaluation
Meaning	2 (100%)	0	0
Choice	2 (50%)	1 (25%)	1 (25%)
Recall	2 (66%)	1 (34%)	0

*Meaning = 100/2 = 50 — Meaning: 0-100 = 2
Choice = 100/4 = 25 — Choice: 0-50: 2; 51-75: 3; 76-100: 4
Recall = 100/3 = 33.333 — Recall: 0-66: 2; 67-100: 3

Table 15. *Components Distribution Percentage in TFA*

Activity	Motivation	Noticing	Retrieval	Generation	Retention
Meaning	1 (16.66%)	2 (33.33%)	0	1 (16.66%)	2 (33.33%)
Choice	1 (12.5%)	3 (37.50%)	2 (25%)	1 (12.50%)	1 (12.50%)
Recall	1 (9.09%)	3 (27.27%)	4 (36.36%)	1 (9.09%)	2 (18.18%)

*Meaning = 100 / 6 = 16.66 — Meaning: 0-16.66 = 1; 16.67-50 = 3; 51-66.66 = 4; 66.67-100 = 6
Choice: 100 / 8 = 12.5 — Choice: 0-12.5 = 1; 12.6-50 = 4; 51-75 = 6; 76-87.5 = 7; 87.6-100 = 8
Recall: 100 / 11 = 9.09 — Recall: 0-9.09 = 1; 9.10-36.36 = 4; 36.37-72.72 = 8; 72.73-81.81 = 9; 81.82-100 = 11

Finally, gain scores were used to examine which of the two frameworks could explain more differences in learners' vocabulary gains. To do so, the participants' gain scores were first used to calculate the weighted scores of each activity for each participant. Next, a hierarchical multiple regression was run to check which of the two frameworks had a higher contribution to IVL gains from the pretests to the posttests.

Following Hu and Nassaji's (2016) conversion method, weighted scores, or the score that each participant receives based on the weight of each component in each framework (Hu & Nassaji, 2016), were calculated as follows: A percentile rank and a score were assigned to the triggered components of each framework based on each task. Next, the learners' gain scores were converted into percentiles. Finally, the learners' percentile scores were converted into weighted scores. The following formula was used to calculate the percentiles for each task: *N*/20 x 100 = *p*% (*N* = score of the task, 20 = number of the target words; Hu & Nassaji, 2016). For example, if a participant's gain score was 9 for the meaning activity, then their percentile score was 45% (9/20 x 100 = 45). Based on the five components of TFA, a score of 45% falls between the 16.67% and 50% percentile rank. As a result, I took this to be roughly equal to index 3 in TFA or 2 in ILH (see Tables 14 and 15):

Table 16. *Multiple Regression Analysis of ILH and TFA Predictive Power*

		R2	Δ*R2**	Δ*F**	*df*	*Sig. F* Change
Receptive	Model 1					
	1. ILH	0.63	0.63	112.95	(1, 64)	0.00
	2. TFA	0.79	0.15	45.48	(1, 63)	0.00
	Model 2					
	1. TFA	0.78	0.78	234.09	(1, 64)	0.00
	2. ILH	0.79	0.00	1.39	(1, 63)	0.24
Productive	Model 1					
	1. ILH	0.56	0.56	81.84	(1, 64)	0.00
	2. TFA	0.77	0.21	58.56	(1, 63)	0.00
	Model 2					
	1. TFA	0.77	0.77	217.37	(1, 64)	0.00
	2. ILH	0.773	0.00	0.01	(1, 63)	0.91

*ΔR2: change in R2; ΔF: change in *F*.

Table 16 shows the results of a hierarchical multiple-regression for evaluating the ILH and TFA frameworks' contributions to both receptive and productive vocabulary gains. The forced entry method was selected for running the hierarchical multiple regression in SPSS. Gain scores in receptive and productive tests were defined as dependent variables and weighted scores of the tasks in both ILH and TFA as predictor variables. In the regression models for receptive and productive knowledge, first ILH was entered into the equation then TFA. In the second regression model for receptive and productive knowledge, first TFA was entered into the equation then ILH.

In model 1 (receptive), after entering ILH into the equation for testing its contribution to receptive knowledge gain, ILH explained 63.8% of the variances, which was statistically significant; next, TFA was entered and it could explain 15.2% additional variances in receptive gains, which was also statistically significant. Conversely, in model 2 (receptive), first TFA was entered then ILH into the equation. TFA could explain 78.5% variances in receptive gains, which was statistically significant, while ILH could only explain 0.5% additional variances, which was not statistically significant.

In model 1 (productive), ILH was entered first into the equation and it could explain 56.1% of gains in productive knowledge of the target words. Next, TFA was entered and it could explain 21% additional variances. Both predictions were statistically significant. Finally, in model 2 (productive), TFA was entered first, and it predicted 77.3% variances, which was also statistically significant. However, after entering ILH, it could not explain any additional variances (0%), which was not statistically significant.

The regression analysis results suggest that TFA was significantly much stronger than ILH in predicting and explaining receptive and productive incidental knowledge and lexical gains in the DGBVL context.

5. Discussion

This study aimed to compare the precision of ILH and TFA frameworks in indexing the success rate of DGBVL activities and their contributions to IVL gains. With respect to the first research question, it was found that the three

DGBVL activities with different levels of involvement load (LIL) enhanced both receptive and productive knowledge of the target words, which supports previous findings (Camacho-Vásquez & Ovalle, 2019; Pham, 2023; Rasti-Behbahani & Shahbazi, 2022; Yang et al., 2023). It seems possible that this result might be due to the effect of inherent elements of digital games that support learners' cognitive-emotional engagement (Alexiou & Schipper, 2018) through three phases (Hamari et al, 2014). In the first phase, digital games provide motivational resources and affordances that, in the second phase, can change learners' motivation; then, in the third phase, these motivational changes can lead to behavioral changes and outcomes that can be guided toward the desired outcomes (Hamari et al., 2014). Therefore, in this study, the digital game might have supported IVL by enhancing motivation (Rasti-Behbahani, 2021), which in turn, not only resulted in emotional but also profound cognitive engagements. Also, because the DGBVL activities were guided and aimed at vocabulary knowledge acquisition, all of them could also result in the desired outcome, that is IVL gains.

For the second research question, results showed that TFA was more precise than ILH in predicting the contribution of DGBVL tasks which supports previous findings (Ehsani & Karimi, 2022; Gohar et al., 2018; Hu & Nassaji, 2016; Khoshsima & Eskandari, 2017). There are explanations for this from two perspectives: framework components and the structure of DGBVL activities. From the framework components perspective, TFA comprises more components than ILH (TFA = 5 vs ILH = 3), which means TFA is sensitive to more factors that can affect IVL gain (Hu & Nassaji, 2016). Therefore, owing to the high level of sensitivity, TFA can statistically explain more variances, which means it has a more precise predictive power. Moreover, the accuracy of TFA escalates even higher due to the examination of each component from different scopes in contrast with ILH, which just examines three components by the absence or two-level presence measure (Ehsani & Karami, 2022). In other words, TFA reflects on the multidimensionality of word knowledge (Nation, 2022), which is fairly neglected in ILH. Hence, despite the existence of unidentified, but effective, factors that can moderate the quality and quantity of IVL gains, TFA accuracy in DGBVL contexts can still remain high.

Conversely, ILH failure in prediction can also be explained by the extent of its component coverage and its approach to treating the components equally (Yanagisawa & Webb, 2021). ILH superficially examines one motivational and two cognitive factors, which might or might not have played any significant role in DGBVL activities to enhance IVL. In other words, DGBVL activities may rely on factors beyond need, search, and evaluation to enhance IVL gains. Moreover, if those cognitive-motivational factors could have played any role, ILH might not have been able to perceive them precisely because it just considered their absence or two-level presence. Therefore, in DGBVL contexts, ILH predictions suffer from significantly low accuracy due to their limited and superficial analyses.

The superiority of TFA over ILH can also be discussed by looking at the structure of DGBVL activities. The superiority of the recall activity, which was also reported by Calvo-Ferrer and Belda-Medina (2021), can be explained by considering the role of spaced repetition and generation or meeting target words in new forms and contexts (Nation, 2022). Webb et al. (2023) found that the effect size of spaced repetition on IVL gains is much larger than frequent exposure in one sitting. In addition, San Mateo-Valdehíta and Criado de Diego (2021) emphasized the importance of the generation in IVL gains. By looking at the frameworks, it is obvious that in contrast to ILH, TFA is sensitive to both repetition types and generation and examines their effects from multiple angles. Therefore, TFA could sense and measure those factors that are neglected in ILH and consequently offer a more accurate indexing method and score.

On the other hand, ILH hypothesized the choice activity as the most effective of all. By looking at the choice task, we can see that the component of search is triggered, which can hinder IVL gains (Yanagisawa & Webb, 2022). Moreover, the evaluation was not strong enough to compensate for the negative effect of the search component (Ehsani & Karami, 2022). Therefore, the ILH framework misinterpreted the choice activity and indexed it wrongly as the most effective of all. Besides, because the ILH framework does not examine the effect of repetition, it misinterpreted the recall activity and indexed it as the second most effective.

Also, this study enhances the assumption of the individual effect of components (Hu & Nassaji, 2016) and supports the discussion by Yanagisawa and Webb (2022) that posits long-term retention of task-induced involvement load should be evaluated with respect to factors different from those involved in short-term retention. From this perspective, TFA was successful in predicting the effectiveness of *all* three tasks because by looking at the retention component individually, both meaning and recall activities received 2 points while the choice activity received only 1 point which means both meaning and recall activities could lead to better retention. Moreover, meaning activity was more

effective than choice because it ensured the successful linking of form and meaning. Also, the recall activity was superior to meaning because it triggered spaced-repetition components (Folse, 2006) in addition to all the components that were triggered by the meaning activity.

Finally, although it was reported that the effect of involvement load might diminish or vanish when the long-term retention of words is examined (Yanagisawa & Webb, 2022), the results of this study showed that TFA can also be sensitive to the long-term effect of DGBVL activity-induced involvement loads and predict their effects on both receptive and productive knowledge gains. Hence, in contrast to ILH, TFA predictions about the long-term effects of DGBVL activities can be reliable.

6. Conclusion

Overall, a few significant conclusions can be drawn from this study. The first outcome delineates that DGBVL activities can support IVL effectively regardless of activity types and their induced levels of involvement load. Next, prospective teachers and researchers can rely on the predictive power of TFA. By doing so, they can design DGBVL activities that can guarantee fairly successful short- and long-term IVL gains. Also, regarding the multidimensionality of word knowledge, this study showed that TFA can help with designing more effective DGBVL activities for enhancing both receptive and productive knowledge. After that, the results of this study imply that the assumption of the collective effect of components in both TFA and ILH seems inefficient, and current frameworks require new methods of interpretation and indexing of IVL activity effects like ILH+. Finally, due to the multidimensionality of word knowledge and the emergence of new learning milieus that introduce unfamiliar but effective factors, ILH and TFA frameworks, which are proposed mainly based on the paper-and-pencil research outcomes, must evolve, and be upgraded by adding new components or aspects.

The findings of this study should be interpreted cautiously due to its limitations. Firstly, in this study, only the acquisition of nouns was examined, which means that the acquisition of other lexical classes has remained unexplored. Moreover, the target words were concrete nouns whose acquisition is both quicker and easier than other types of nouns (Nation, 2022). This point should be considered in future studies. In addition, the lack of a control group leads to uncertainty about the effect of the DGBVL activities, which could be addressed in future studies. Furthermore, the findings of this study are limited to only adventure digital games while the effect of DGBVL activities should be studied in the other genres. Last but not least, the small number of participants can statistically limit the generalizability of the findings, which should be avoided in upcoming studies.

Information on Informed Consent or any Data Privacy Statements

All the participants voluntarily participated in this study and showed their willingness to participate by signing the consent letters. Also, due to their age range and cultural sensitivity, their parents were informed and asked for their permission. The participants and their families were also aware of the right to leave at any stage of the study.

Author Contributions

Material designs and instrument selection, data collection and analysis, and report composition and dissemination of the findings were done by the author.

Conflict of Interest

The author has no conflicts of interest to declare.

Ethics Board Approval Statements

Dhofar University Research Council (DURC) approved the ethical soundness of this study before conducting and data collection.

Funding

The author received no funding for this study.

References

Alexiou, A., & Schippers, M. C. (2018). Digital game elements, user experience and learning: A conceptual framework. *Education and Information Technologies, 23*(6), 2545-2567. https://doi.org/10.1007/s10639-018-9730-6

Amengual-Pizarro, M. (2024). Exploring the relationship between out-of-class engagement with English and academic achievement at university level. *Journal of Research in Applied Linguistics*. https://doi.org/10.22055/rals.2024.46193.3243

Boers, F. (2022). Glossing and vocabulary learning. *Language Teaching, 55*(1), 1-23. https://doi.org/10.1017/S0261444821000252

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), 80-95. https://doi.org/10.3390/mti5120080

Camacho Vásquez, G., & Ovalle, J. C. (2019). The influence of video games on vocabulary acquisition in a group of students from the BA in English teaching. *GIST - Education and Learning Research Journal, 19*, 172-192. https://doi.org/10.26817/16925777.707

Chen, H. J. H., Hsu, H. L., & Chen, Z. H. (2020). A study on the effect of adding L1 glosses in the subtitle of an adventure game for vocabulary learning. *Interactive Learning Environments, 31*(4), 1889-1905. https://doi.org/10.1080/10494820.2020.1863233

Díez-Arcón, P., & Martin-Monje, E. (2023). Language teacher development in computer-mediated collaborative work and digital peer assessment: An innovative proposal. *Journal of Research in Applied Linguistics, 14*(2), 40-54. https://doi.org/10.22055/rals.2023.44054.3080

Ehsani, M., & Karami, H. (2022). Comparing the predictive power of involvement load hypothesis and technique feature analysis. *International Journal of Language Studies, 16*(2), 165-188.

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly, 40*(2), 273-293. https://doi.org/10.2307/40264523

Gohar, M. J., Rahmanian, M., & Soleimani, H. (2018). Technique feature analysis or involvement load hypothesis: Estimating their predictive power in vocabulary learning. *Journal of Psycholinguistic Research, 47*, 859-869. https://doi.org/10.1007/s10936-018-9568-5

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. https://doi.org/10.1093/applin/amy057

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. In *2014 47th Hawaii International Conference on System Sciences* (pp. 3025-3034). IEEE. https://doi.org/10.1109/HICSS.2014.377

Hu, H. C. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement load hypothesis versus technique feature analysis. *System, 56*, 28-39. https://doi.org/10.1016/j.system.2015.11.001

Jahangiri, K., & Abilipour, I. (2014). Effects of collaboration and exercise type on incidental vocabulary learning: Evidence against involvement load hypothesis. *Procedia - Social and Behavioral Sciences, 98*, 704-712. https://doi.org/10.1016/j.sbspro.2014.03.471

Karami, H., & Esrafili, M. (2021). The impact of task type and involvement load index on Iranian EFL learners' incidental vocabulary learning and retention. *Journal of Language Horizons, 5*(1), 251-266. https://doi.org/10.22051/lghor.2020.31501.1311

Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research, 12*(3), 365-386. https://doi.org/10.1177/1362168808089922

Khazaie, S., & Mashhadi, A. (2020). An investigation into the impact of grafting m-game onto the blended module of L2 pedagogy. *Teaching and Learning Research, 14*(2), 39-53. https://doi.org/10.22070/tlr.2020.2511

Khoshsima, H., & Eskandari, Z. (2018). Task effectiveness predictors: Technique feature analysis versus involvement load hypothesis. *Iranian Journal of English for Academic Purposes, 6*(2), 50-69. https://dorl.net/dor/20.1001.1.24763187.2017.6.2.5.3

Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 58*(2), 100-140. https://doi.org/10.1111/j.1467-9922.2011.00644.x

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics, 29*(4), 694-716. https://doi.org/10.1093/applin/amn018

Laufer, B., & Hulstijn, J. (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

Martínez-Fernández, A. (2008). Revisiting the involvement load hypothesis: Awareness, type of task, and type of item. In *Selected proceedings of the 2007 second language research forum* (pp. 210-228). Cascadilla Press. www.lingref.com, document #1746

Nation, I. S. P. (2022). *Learning vocabulary in another language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Cengage Learning Heinle.

Peterson, M., White, J., Mirzaei, M. S., & Wang, Q. (2020). A review of research on the application of digital games in foreign language education. In M. Kruk & M. Peterson (Eds.), *New technological applications for foreign and second language learning and teaching* (pp. 69-92). Information Science Reference/IGI Global. https://doi.org/10.4018/978-1-7998-2591-3.ch004

Pham, Q. (2023). Maximizing vocabulary retention with gamification tools. *Scientific Journal of Tan Trao University, 8*(4). https://doi.org/10.51453/2354-1431/2022/837

Pivec, M., Dziabenko, O., & Schinnerl, I. (2003, July). Aspects of game-based learning. In *3rd International Conference on Knowledge Management, Graz, Austria* (pp. 216-225).

Rasti-Behbahani, A. (2021). Why digital games can be advantageous in vocabulary learning. *Theory and Practice in Language Studies, 11*(2), 111-118. https://doi.org/10.17507/tpls.1102.01

Rasti-Behbahani, A., & Shahbazi, M. (2022). Investigating the effectiveness of a digital game-based task on the acquisition of word knowledge. *Computer Assisted Language Learning, 35*(8), 1920-1945. https://doi.org/10.1080/09588221.2020.1846567

San Mateo-Valdehíta, A., & Criado de Diego, C. (2021). Receptive and productive vocabulary acquisition: Effectiveness of three types of tasks. Results from French students of Spanish as a second language. *Onomázein, 51*, 37-56. https://doi.org/10.7764/onomazein.51.05

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan. http://dx.doi.org/10.1057/9780230293977

Sökmen, A. J. (1997). Current trends in teaching second language vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 237-257). Cambridge University Press.

Sundqvist, P., & Nilsson, R. (2022). Young gamers in the digital wilds: Implications of gaming preferences on L2 English vocabulary learning and teaching. *Babylonia Journal of Language Education, 1*, 14-19. https://doi.org/10.55393/babylonia.v1i.153

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Traxler, J., Barcena, E., & Andujar, A., Jalilifar, A. R., & Mashhadi, A. (2023). Introduction: Teaching languages in times of social and technological change and divide. *Journal of Research in Applied Linguistics, 14*(2), 3-6. https://doi.org/10.22055/rals.2023.18722

Vassilieva, N. V., & Drugov, A. V. (2019). Integrative use of "techno-R" remedial technology and gaming technology in teaching foreign language listening. *Journal of Research in Applied Linguistics, 10* (Proceedings of the 6th International Conference on Applied Linguistics Issues (ALI 2019) July 19-20, 2019, Saint Petersburg, Russia), 463-472. https://doi.org/10.22055/rals.2019.15030

Vinter, A., Bard, P., Lukowski-Duplessy, H., & Poulin-Charronnat, B. (2022). A comparison of the impact of digital games eliciting explicit and implicit learning processes in preschoolers. *International Journal of Child-Computer Interaction, 34*, 100534. https://doi.org/10.1016/j.ijcci.2022.100534

Vnucko, G., & Klimova, B. (2023). Exploring the potential of digital game-based vocabulary learning: A systematic review. *Systems, 11*(2), 57-75. https://doi.org/10.3390/systems11020057

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225-239). Routledge. https://doi.org/10.4324/9780429291586

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated vocabulary levels test: Developing and validating two new forms of the VLT. *ITL - International Journal of Applied Linguistics, 168*(1), 33-69. https://doi.org/10.1075/itl.168.1.02web

Webb, S., Uchihara, T., & Yanagisawa, A. (2023). How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching, 56*(2),161-180. https://doi.org/10.1017/S0261444822000507

Yanagisawa, A., & Webb, S. (2021). To what extent does the involvement load hypothesis predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning, 71*(2), 487-536. https://doi.org/10.1111/lang.12444

Yanagisawa, A., & Webb, S. (2022). Involvement load hypothesis plus: Creating an improved predictive model of incidental vocabulary learning. *Studies in Second Language Acquisition, 44*(5), 1279-1308. https://doi.org/10.1017/S0272263121000577

Yang, T., & Jeaco, S. (2023). Vocabulary building on the move with the Prime Machine: Evaluation of a cross-platform corpus-based game for English vocabulary learning. *International Journal of TESOL Studies, 5*(2), 64-78. https://doi.org/10.58304/ijts.20230206

Zou, D. (2012). *A study of the components of the involvement load hypothesis: How involvement load should be allocated to "Search" and "Evaluation"* [Unpublished doctoral dissertation]. City University of Hong Kong. https://scholars.cityu.edu.hk/en/theses/theses(b31b00f2-8ff8-4fe4-8bad-aae418f59e95).html

Zou, D. (2017). Vocabulary acquisition through cloze exercises, sentence-writing and composition-writing: Extending the evaluation component of the involvement load hypothesis. *Language Teaching Research, 21*(1), 54-75. https://doi.org/10.1177/1362168816652418