OPEN ACCESS

Australian Journal of Applied Linguistics
ISSN 2209-0959
https://www.castledown.com/journals/ajal/

Australian Journal of Applied Linguistics, 9, 103759 (2026)
https://doi.org/10.29140/ajal.2026.103759

Instructional formats and digital game-based vocabulary learning activities: The role of extraneous cognitive load

AMIN RASTI-BEHBAHANI*[a] (Corresponding author)
NAGWA MOHAMMED KHALLAF[a]
YAZDAN CHOUBSAZ[b]

[a]*Dhofar University, Oman*
[b]*Sultan Qaboos University, Oman*

Abstract

This research examined the impact of minimizing extraneous cognitive load on incidental vocabulary learning during digital game-based vocabulary learning (DGBVL) activities. The participants were 70 Arabic-speaking university freshmen (CEFR A2) who were randomly assigned to an Integration Group, where the extraneous load was minimized, and a Control Group, where it was not. The students played an arcade-style online game that taught them ten low-frequency pseudowords. Then they were given Immediate and Delayed post-tests, which measured eight areas of word knowledge, including both receptive and productive skills. A 2 × 2 repeated-measures MANOVA revealed significant main effects of group and time. The integration format, which was less cumbersome for their working memory, allowed the participants to experience more vocabulary gains and better retention than the control group. The advantages were particularly strong for the productive aspect of word knowledge, which demands more from working memory. These results support the hypothesis that the proper minimization of extraneous cognitive load may lead to an increase in both learning efficiency and retention. It therefore emphasizes the significance of teacher intervention in the design of DGBVL with extraneous cognitive load as a factor.

Keywords: Digital game-based vocabulary learning (DGBVL), incidental vocabulary learning, instructional formats, split-attention, integration format, separate format, cognitive load theory (CLT), video games

Introduction

Instructional formats play a critical role in language learning (File & Adams, 2010). However, the success of any instructional format depends on many factors, among which cognitive load theory (CLT) serves as a guiding framework. CLT emphasizes the limitations of working memory and the importance of managing mental effort to facilitate meaningful learning (Brünken et al., 2010). According to CLT, a poorly designed instructional format, like a separate format, can impose unnecessary loads that can hinder learning and retention because, according to CLT, when learners must divide their attentional resources between different sources of relevant information, the split-attention effect occurs, which hinders learning (Brünken et al., 2010). To avoid this, an integration format is recommended (Al-Shehri & Gitaski, 2010).

However, empirical studies on the effect of integration and separate instructional formats on incidental vocabulary learning (IVL)—learning components of word knowledge as a byproduct of a language learning activity designed for a different purpose (Ashcroft et al., 2019)—have reported mixed results. While the integration format is considered a more effective format theoretically and can often enhance reading comprehension (Al-Shehri & Gitaski, 2010; Yeung et al., 1998), several studies have found that it may not necessarily improve IVL outcomes (Awad, 2022; File & Adams, 2010; Ashcroft et al., 2019). In contrast, studies like File and Adam's (2024) discuss that separate formats can even be more effective for IVL. These inconsistencies may come from the absence or limited control over contextual factors like word exposure and repetition (Nation, 2024). This suggests that the effectiveness of IVL depends not only on the instructional format with a managed cognitive load but also on the mental effort needed for processing the vocabulary items.

At the same time, Digital Game-based Vocabulary Learning (DGBVL) has emerged as a promising way to support both receptive and productive IVL through interactive and engaging environments that can support the depth of processing of the vocabulary items (Chen & Hsu, 2019; Rasti-Behbahani, 2023; Rasti-Behbahani & Shahbazi, 2022; Vnucko & Klimova, 2023). However, research shows that utilising digital games without proper adjustments may also result in impaired IVL gains (Rasti-Behbahani, 2023). One adjustment that can enhance a proper adaptation of DGBVL activities is managing their cognitive load demands because research has shown how this demand can affect the success rate of DGBVL activities (deHaan et al., 2010; Mohsen, 2016; Vyvey et al, 2018), particularly concerning extraneous cognitive load (EL) (Chen et al., 2020; Rasti-Behbahani, 2023). These mismanagement issues have caused inconsistencies in studies that use both integration and separate (causing split-attention) formats in DGBVL activity designs (Chen et al., 2020; Rasti-Behbahani, 2023).

Given these findings, it is still unclear how minimizing EL affects IVL gain and retention in digital game-based contexts. To address this gap, the present study examines whether minimizing EL improves the effectiveness of DGBVL activities in promoting better acquisition and retention of components of word knowledge.

Components of Word Knowledge

Acquiring vocabulary is a demanding task for language learners because word knowledge is a multidimensional construct (Webb, 2008). Our latest understanding of word knowledge components is based on what Nation introduced in 2001 (see Table 1). In this framework, learning a word means acquiring *aspects,* including form, meaning, and use, and *dimensions,* comprising receptive and productive knowledge. Nation (2001) defines receptive knowledge as being able to understand and recognize a word in a written or spoken format, and productive knowledge as being able to use words in writing or speaking.

Table 1. *Components of word knowledge (adapted from Nation, 2001, p. 27)*

Form	Spoken	R	What does the word sound like?
		P	How is the word pronounced?
	Written	R	What does the word look like?
		P	How is the word written and spelled?
	Word parts	R	What parts are recognizable in this word?
		P	What word parts are needed to express the meanings?
Meaning	Form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	Concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	Associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	Grammatical functions	R	In what patterns does the word occur?
		P	In what pattern must we use this word?
	Collocations	R	What words or type of words occur with this one?
		P	What words or type of words must we use with this?
	Constraints on use (register, frequency, etc.)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

Note. R = receptive knowledge; P = productive knowledge.

Besides, according to Laufer and Goldstein (2004), the receptive–productive dimension is also conceptualized as comprising two *scopes*—recognition and recall—mainly operationalized within the form–meaning aspect of lexical knowledge (see Table 2).

It has been found that acquiring receptive knowledge precedes productive knowledge of components of word knowledge, and receptive knowledge is retained even longer (Meara, 1997; Mondria & Wiersma, 2004). However, Webb (2005) found that this precedence is conditional and depends on factors like the type of task, experiment design, and ceiling effect on the receptive test. This finding by Webb (2005) is important because it suggests that the rate and route of word knowledge gains in all aspects, dimensions, and scopes are manageable through task and experiment design. In other words, utilizing a proper instructional format may facilitate the gain and retention of components of word knowledge, which is the focus of this study.

Table 2. *Degrees of form-meaning knowledge (adapted from Laufer & Goldstein, 2004, p. 407)*

	Recall	Recognition
Productive (retrieval of form)	Supply the L2 word	Select the L2 word
Receptive (retrieval of meaning)	Supply the L1 word	Select the L1 word

Extraneous Cognitive Load and Instructional Format

The importance of instructional format lies in its role in determining the magnitude of cognitive load, which is influential on word knowledge gain. Cognitive load theory (CLT) posits that working memory—the part of memory responsible for processing and managing inputs—has a limited capacity that must be used for essential processes that enhance meaningful learning (Brünken et al., 2010). However, poorly designed educational interventions, such as information misplacement in a separate format, can often lead to the split-attention effect (Yeung et al., 1998) and cognitive overload that are referred to as 'extraneous load' (Brünken et al., 2010).

Research on the impact of integration and separate formats has produced mixed findings regarding their effectiveness in vocabulary acquisition. Al-Shehri and Gitaski (2010) studied their effects on reading comprehension and vocabulary acquisition. Participants were divided into four groups: half received a reading comprehension text with comprehension questions placed in between the paragraphs (integration), while the other half answered the comprehension questions after the text (separate). Half of each of these groups also had access to an online dictionary. The results showed that although the integration group achieved higher comprehension scores than the separate group, their IVL gains were significantly lower than those of groups with an online dictionary. Furthermore, the results showed no significant difference between the two online-dictionary groups in vocabulary acquisition, suggesting that the instructional format itself had a negligible effect on vocabulary acquisition. In another study, Awad (2022) studied the role of integration and separate formats in academic vocabulary acquisition, while also controlling for proficiency level and first language (L1). The pool of participants consisted of 35 Arabic speakers and 25 Chinese speakers who were categorized into low- and high- English proficiency groups. The integration group read texts partitioned into short sections, and each of the sections had a vocabulary exercise on the same webpage, while the separation group read the whole text first and then did the vocabulary questions on a different webpage. The findings showed that despite insignificant statistical differences, both formats could improve knowledge of word meaning, families, and collocations. However, high proficiency participants in the separate group could answer more questions correctly compared to those with lower proficiency. Moreover, the effect of L1 was statistically insignificant. File and Adams (2010) compared the effect of integration and isolated form-focused instruction on incidental learning of receptive and productive knowledge of 36 low-frequency words. Participants were 20 English learners from various Asian L1 backgrounds. The isolation group received explicit instruction on target words before reading the main text, whereas the integration group encountered and learned the words within the text, with explanations provided by the researchers upon first exposure. They found that both instruction types could enhance vocabulary acquisition significantly compared to a control group. However, the retention test revealed a slight advantage for the isolation group.

Lastly, in a more detailed investigation, Yeung et al (1998) examined the role of cognitive load and learner proficiency in determining reading comprehension and vocabulary learning through the use of explanatory notes in either separate or integration formats. They conducted five experiments wherein both L1 (native) and L2 (ESL) readers were analyzed with different proficiency levels. They found that the integration format enhanced the reading comprehension but did not enhance vocabulary acquisition of low proficiency L1 and L2 learners. Whereas the results were different for higher proficiency L1 and L2 learners, whose vocabulary acquisition was enhanced more effectively than their reading comprehension in the separate format. The researchers concluded that the method of instruction has a strong interaction with the skill of the learner regarding reading comprehension and vocabulary acquisition.

Overall, these studies reveal inconsistency about the influence that instructional formats have on vocabulary acquisition. One reason for these contradictory findings might be that the majority of the studies mainly concentrated on reading comprehension skills and did not take into account the text and context preparation for proper guessing and IVL. Except for File and Adam's, few studies controlled contextual factors such as the number of exposures and the spaced repetition (Nation, 2024), which are known to affect IVL. This may explain the outperformance of the separate format group in File and Adam's study. Hence, for a better understanding of the influence of instructional format on IVL, a more controlled and systematically designed experiment, aligning more closely with CLT and IVL literature, is necessary.

Digital Game-based Vocabulary Learning

Digital game-based vocabulary learning (DGBVL), which is defined in detail by Rasti-Behbahani (2023) as the educational use of video or digital games to extend learners' vocabulary both qualitatively and quantitatively, has been widely reported as an effective method that enhances vocabulary knowledge gain in receptive and productive aspects. A case example is the study of Chen and Hsu (2019) in which a serious game, Slave Trade, was able to support 66 Chinese participants' receptive knowledge acquisition of certain words and historical events as per the comparison of the pre-and posttests results. Rasti-Behbahani and Shahbazi (2022) also found that DGBVL activities could enhance the incidental acquisition of nearly all the dimensions and aspects of word knowledge. They conducted their research involving 124 Persian speakers, divided into experimental and control groups. The group subjected to the experiment learned 10 target words through DGBVL activities, while the control group practiced the same vocabulary through a fill-in-the-blank paper activity. After three weeks, when eight aspects of vocabulary knowledge were tested, the experimental group retained more vocabulary knowledge, especially in the case of productive-recognition knowledge of form-meaning. Soyoof et al. (2026) compared paper, non-gamified digital, and gamified mobile flashcards for 98 young Iranian EFL learners' vocabulary and collocation learning. All formats could support word knowledge gains, but gamified mobile flashcards led to substantial improvements because of their multisensory input and motivational features. The learners found the gamified flashcards more engaging, indicating the effectiveness of DGBVL activities in supporting receptive vocabulary and productive collocation gains. Overall, there is a consensus that DGBVL activities are one of the most effective interventions to increase IVL (Pradheepa et al., 2025; Vnucko & Klimova, 2023).

DGBVL activities were found to support IVL through their inherent elements. Zhang et al (2025) studied the effect of innate elements of digital games, such as challenge, fantasy, multimedia, rewards, and human-computer interactions (HCI), on learning 20 low-frequency (<9K) English vocabulary items. To collect data, they recruited 50 Chinese university students and

used pre- and posttests, eye-tracking devices, and interviews for data collection. They found that engagement with game elements such as HCI and fantasy can enhance vocabulary learning effectively. Moreover, factors such as flow, or the state of deep engagement with a task at hand (Karimi & Nasouri, 2024), motivation, authenticity, repetition, instantiation, dual encoding, interactivity, and feedback (Rasti-Behbahani, 2021) were frequently reported as supportive of IVL in DGBVL activities.

Research has shown that cognitive load management is important in IVL from DGBVL activities. For example, deHaan et al. (2010) examined the role of interactivity in the recognition and recall of English word forms in DGBVL activities. They assigned 80 Japanese teenagers to two groups: one that played a rhythm game, and another that watched the game. After testing both groups, the researchers found that the watchers recognized and recalled more word forms than the players. They attributed this finding to cognitive overload caused by high interactivity, which hindered vocabulary acquisition in the playing group. Mohsen (2016) replicated this study using a simulated surgery game (less interactive) with Arab speakers. He found that DGBVL activities with less interactivity were more effective for vocabulary acquisition, as they did not impose unnecessary cognitive loads on working memory. He emphasized the importance of considering CLT in the design of DGBVL activities for effective vocabulary learning. Lastly, although not directly conducted in the second language learning context, a study by Vyvey et al. (2018) can explain why cognitive load optimization of DGBVL activities is important. They studied the effect of cognitive load and enjoyment manipulation on implicit and explicit pseudo-brand names recall and recognition. They found that high EL—caused by visual clutter, auditory noise, and higher game speed—redistributed the attentional sources and narrowed them down to the core gameplay. That is, due to high cognitive load, participants did not process the brand names deeper (implicit learning), e.g., to relate them to imaginary products the brands advertised in the game, but they narrowed their focus to the necessary elements of gameplay—called "ludic focus"—that helped progression in the game. Hence, in the posttest, participants could recognize and recall the brand names forms more effectively because they were a part of core gameplay (explicit learning). This suggests that when attentional resources are consumed by gameplay demands due to a poor design, learners are less likely to encode lexical items robustly.

Research has also shown that design, or instructional format, is important for learning from digital games because evidence has shown that a well-designed digital game activity and gameplay can enhance vocabulary learning effectively (Peterson & Swier, 2025). Nevertheless, despite the effectiveness of DGBVL activities, research on the DGBVL instructional format—integration vs. separate—has indicated mixed results, mostly due to cognitive load mismanagement (Rasti-Behbahani, 2023). For instance, a study by Chen et al. (2020) examined how adding Chinese translations (L1 glosses) to a digital game text affected vocabulary learning. The researchers modified an existing adventure digital game by placing L1 glosses next to 28 target words in the subtitles for 21 eighteen-year-old Chinese participants. Their results showed that integrating L1 glosses into the game significantly improved IVL. However, in another study, Rasti-Behbahani (2023) studied the effect of instructional format (pre-teaching vs. integration) on IVL through an adventure DGBVL activity and found otherwise. He recruited 33 B1-level Persian teenagers and assigned them to two DGBVL activities—pre-teaching (separate) and integration (glosses)—for learning 20 low-frequency English nouns. The results showed that the pre-teaching group's productive and receptive IVL gains were significantly higher than those of the integration group. Think-aloud data revealed that the integration format led to cognitive overload. The results were consistent with the study by File and Adams (2010). Lastly, Rasti-Behbahani et al. (2024) studied the effect of two integration formats—single vs. multiple

choice glosses—on fifty-four Persian speakers. Participants were assigned to meaning-given, meaning-inferred (Multiple-choice), and no-gloss DGBVL activities, and the results showed that although glosses enhanced IVL, meaning-inferred glosses had a reverse effect. They attributed this to mismanagement of cognitive load in the DGBVL activities.

Overall, poorly designed DGBVL activities with inappropriate instructional formats can increase learners' cognitive load and complicate learning new words incidentally. This problem can be especially common in integration formats that use multiple-choice glosses. As Rasti-Behbahani (2023) explains, DGBVL designs that require learners to engage in intense and unguided searches can negatively affect IVL. Such unguided searching often leads to unnecessary EL and interferes with the learning process (Tuovinen & Sweller, 1999). Therefore, to address this issue, it is important to investigate whether minimizing EL can improve IVL in DGBVL activities, especially in the integration format. This study will address the following question:

1. Does minimizing EL enhance the effectiveness of DGBVL activities in promoting greater word knowledge acquisition and retention?

Methodology

This empirical study employs an experimental design (experimental and control groups/treatments/immediate and delayed posttests). Participants learned target words through DGBVL activities. Vocabulary learning was incidental, as participants were not forewarned about the tests or the primary purpose of the activity (Webb, 2020).

Participants

Participants in this study were 70 male and female Omani Arabic-speaking teenagers (18–19 years old) who were freshmen at the Centre for Preparatory Studies at Dhofar University in Oman with CEFR A2 proficiency level, measured by the online Cambridge English Proficiency Test (CEPT). Their vocabulary knowledge was also measured to guarantee their proper comprehension of the texts used in the DGBVL activities by administering the first and second bands of most frequent English words on the Updated Vocabulary Levels Test (uVLT) (Webb et al., 2017). Their overall score was 90.31% (M = 27.09, SD = 1.24), with 98.22% (M = 29.46, SD = 0.93) in 1K[1], suggesting they had the mastery of 1K (Webb et al., 2017) and high enough vocabulary knowledge (González-Fernández & Schmitt, 2020) to understand the texts in DGBVL activities. However, two participants' scores were found unsatisfactory, and they were removed from the study. Participants were competent in interacting with web-based digital games, had experience in playing, and were aware of the game mechanics of the web-based digital game used in this study.

Materials

Target Words

Ten nouns representing inanimate objects were selected for the study (see Appendix A), as nouns generally impose the least learning burden compared to other lexical categories (Rasti-Behbahani, 2023). To ensure authentic vocabulary acquisition and prevent prior knowledge effects, the words were chosen from the low-frequency English lexicon (verified using Vocab-Profiler on www.lextutor.ca) and replaced with pseudo-words previously used in Webb (2007). Similar to Webb's approach, the acceptability of these words as English vocabulary items was

[1] First 1000 frequent English words band.

tested in brief interviews, during which participants believed the pseudo-words were real English words.

When determining the number of target words, informed by researchers' previous pilot studies and experienced teacher colleagues' recommendations, factors such as potential fatigue and the time required to complete the DGBVL activities and assessments were taken into account.

Digital Game, DGBVL Activity Design, and Cognitive Load Optimization

The game, used in this study, was designed by the researchers on a web platform called Educaplay (www.educaplay.com). Its genre is arcade-style. In the general design of DGBVL activities, each target word corresponded to one jump in the game, with each definition accompanied by an image representing the object on a lotus leaf (see Figures 1 and 2). Participants had to read each short paragraph and decide on the Arabic definition of the target words. After making their choice, they tapped on the corresponding image on a lotus leaf to make the frog jump. If the answer was correct, the frog remained on the lotus, and the participant earned points. If the answer was incorrect, the frog drowned, and no points were awarded. Incorrect jumps could not be repeated. The general instruction provided to participants was:

> "*Help Froggy cross the lake by jumping onto the correct lotus leaves. The texts contain clues to assist you in making the right choices.*"

Since EL effect and IVL were the main factors in this study, it was necessary to first present the target words in texts and then modify the target words and optimize the cognitive load of the surrounding texts. Thus, target words were embedded in short paragraphs, as shorter interventions promote greater interaction (Liu, 2024). Since the students were at the A2 proficiency level, the vocabulary coverage of the composed texts was controlled using VocabProfiler (M_{1K} = 82.8%; M_{2K} = 12.3%; $M_{\leq 3K}$ = 4.9%; 1K and 2K text coverage: 95%) to ensure that 95% of the words fell within the 2,000-word frequency band to assist them with guessing (Hirsh & Nation, 1992). Milton (2010) found that learners with CEFR A2 proficiency are expected to have a vocabulary size between 1500 and 2500.

Figure 1. *Game designed for the control groups*

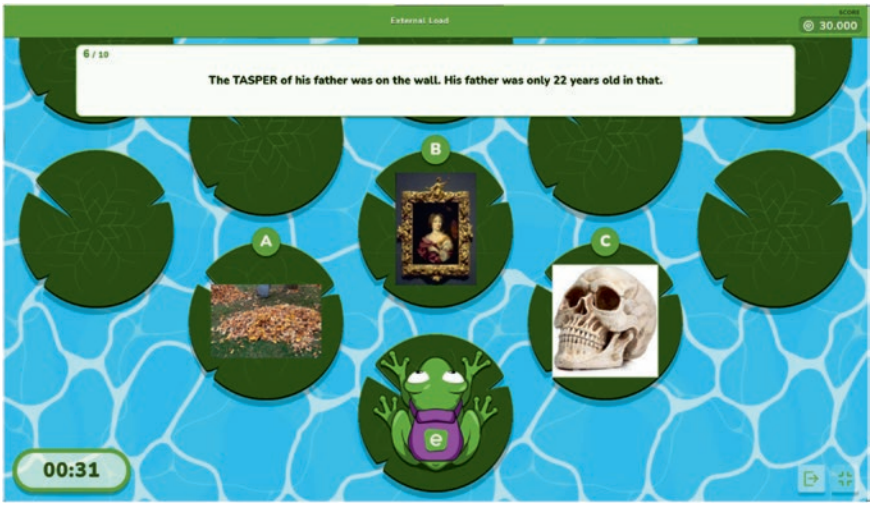

Figure 2. *Game designed for the integration format group*

Additionally, three native English-speaking teachers reviewed and revised the texts according to the scale defined by Webb (2008) (see Appendix B) to ensure all paragraphs were moderately informative because moderately informative contexts are more effective for learning novel words (Lowell & Morris, 2017).

In total, 10 short, moderately informative paragraphs were composed. Each paragraph consisted of three to four sentences and contained one target word. The target words were presented in bold and underlined within each paragraph. Three Arabic definitions were provided alongside each paragraph for each target word because L1 glosses can be more beneficial than other gloss types (Outamgharte et al., 2025). It was ensured that no definition appeared more than twice in the margins (see Appendix C). Paragraphs were later printed on paper and handed over to the participants.

For the cognitive load optimization of the integration format group's (IG) DGBVL activity, the EL was minimized by reducing the split-attention effect. To do so, the sentences containing the target words were incorporated into the game (see Figure 2). When participants had to choose a lotus leaf to jump on, they were presented with images and sentences containing the target words on their screens. This design was based on three key considerations. First, Brünken et al. (2010) argued that retaining information in working memory –such as extracting information from a text and applying it in another context, like a game interface–contributes to EL. Second, Hung (2007) found that integrated designs can effectively minimize EL. Third, since the target words were important for comprehension, integration could minimize EL and enhance comprehension (Yeung et al., 1998).

For the control group (CG), or separate format, no modifications were made to the text. Similarly, the game screen for this version did not display any target words (see Figure 1).

Instruments

Updated Vocabulary Level Test

On uVLT (reliability = 0.89), 1000- and 2000-word levels, which encompass 60 receptive recognition questions, were administered. There are three definitions and six words in each question.

Participants had to match one definition with one word (see Appendix D for a sample question). With no time limit, a paper-based format was administered.

Achievement Tests

Given the multidimensional nature of word knowledge (Nation, 2001), a comprehensive assessment of vocabulary acquisition across multiple dimensions was essential to evaluate the effect of the CLT. Accordingly, eight achievement tests were designed based on the study framework by Webb (2008). To minimize potential interference between tests, they were administered in the sequence: A, F, B, H, E, D, G, and C (explained in detail below), with target words randomly distributed across the tests. Contextual clues were excluded to prevent guessing. This sequence helped what test-takers encounter in the earlier test does not teach them the information necessary for answering the next test. For instance, test A did not provide the orthographical forms of the target words to be helpful for test B, and so on. This order was previously used by Rasti-Behbahani and Shahbazi (2022).

During test administration, participants received and completed each test individually, returning the completed test to the researchers before proceeding to the next. Each test consisted of 10 questions, with each question assigned a weight of 1. The tests assessed different dimensions, aspects, and scopes of productive and receptive word knowledge:

A. *Productive knowledge of orthography:* This was a spelling test in which target words were read aloud twice. Participants were required to write down the words with accurate spelling. Minor spelling errors were penalized.

B. receptive knowledge of orthography: Participants selected the correct spelling of target words, for example:

a. shigom b. shimog c. gomish d. ***gishom*** ☑

C. receptive knowledge of meaning and form (recognition): This multiple-choice test required participants to identify the correct definition of a target pseudoword. For example, option A was the correct definition for the pseudoword *mesut*:

a. سيك ☑ *b.* ةيناطب *c.* ةحول *d.* ةراتسلا

D. receptive knowledge of meaning and form (recall): Participants recalled and wrote the Arabic definitions of the target words.

E. productive knowledge of meaning and form (recognition): Participants matched Arabic definitions with corresponding pseudowords, which included distractor real words to reduce repetition effects. For example, for the word مطاح, the correct choice was C:

a. Cader b. Glove c. ***Ancon*** ☑ *d. Matches*

F. productive knowledge of meaning and form (recall): Participants translated Arabic definitions into pseudowords, recalling the correct pseudoword forms. Minor spelling errors were not penalized.

G. receptive knowledge of association: Participants identified unrelated words among options. Real words were included as distractors. For instance, option B was irrelevant in this example:

a. Ictay b. ***Ancon*** ☑ *c. Oil d. Fire*

H. productive knowledge of association: Participants generated words with syntagmatic or paradigmatic associations to the target words. Loose associations were accepted, and spelling errors were not penalized.

The validity of the tests was confirmed by feedback from English teachers, non-participating volunteers, and an applied linguistics professor, all of whom agreed that the tests were appropriately designed to measure vocabulary knowledge. Additionally, by measuring the correlation coefficient between the control group's immediate and delayed posttests, the reliability of the tests was calculated and demonstrated high consistency (Larson-Hall, 2015), with a reported reliability coefficient of 0.83.

Cognitive Load Measurement

To measure the cognitive load magnitude of the DGBVL activities, a questionnaire adapted from deHaan et al. (2010, p. 81) was used with a high reliability (Cronbach's Alpha = 0.86) (Larsen-Hall, 2015). This questionnaire was selected because deHaan et al. (2010) used it to measure the magnitude of cognitive load and its effect on DGBVL, which aligns with the purpose of our study. The questionnaire had four items that measured mental effort and material difficulty. Its items were designed to differentiate between the cognitive load associated with gameplay and the load stemming specifically from the DGBVL activity linguistic elements. The questionnaire was administered in Arabic.

1. How much mental effort did you invest in playing the digital game? (*nine-point Likert scale from extremely low mental effort to extremely high mental effort*)
2. How easy or difficult was the digital game to play? (*seven-point Likert scale from extremely easy to extremely difficult*)
3. How much mental effort did you invest in studying the language of the DGBVL activity? (*nine-point Likert scale from extremely low mental effort to extremely high mental effort*)
4. How easy or difficult was the language of the DGBVL activity to understand? (*seven-point Likert scale from extremely easy to extremely difficult*)

Before administration, an applied linguistics professor checked the validity of the questionnaire and confirmed that it serves its intended purpose.

Process

The research was set up using an experimental design (see Figure 3). The experimental group experienced an EL-minimized DGBVL activity, while cognitive load was not optimized in the control group's DGBVL activity.

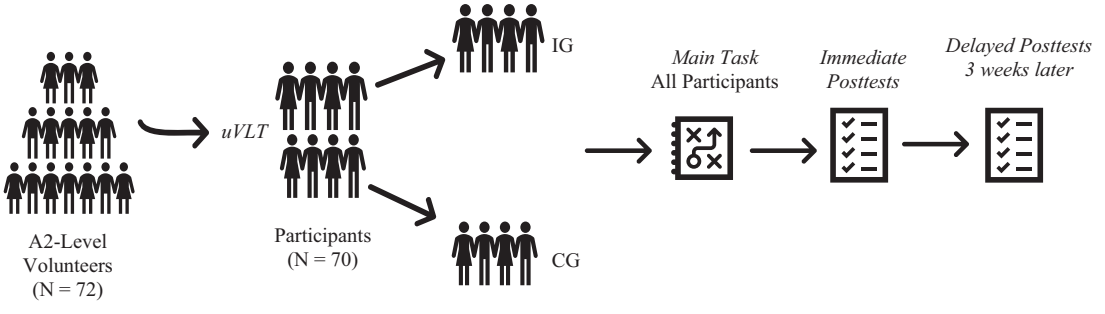

Figure 3. *Graphic illustration of the study process*

After granting the ethical approval (decision number DU-AY-24-25-QUES-002) by the Dhofar University Research Ethics and Biosafety Committee (DUREBC) to conduct the study, a consent form was composed in which the participants were made aware of their rights as research subjects, and it was clearly mentioned that they could drop out willingly at any stage of the study. The consent was written in Arabic, and participants were asked to read and sign it. Seventy-two learners of English at the A2 level were chosen based on their scores on the CEPT. Amongst the 72 volunteers, 2 were left out because their uVLT scores were too low. The rest of the participants were assigned randomly to the IG and CG, with 35 participants in each group. In the designated rooms, the participants of each group came together and did the DGBVL activities individually on their mobile phones or tablets. A leaderboard of the game was shown for the purpose of increasing motivation through competition, and the three best performers in each group were given three packs of chocolate as the prize. On average, the time spent on the DGBVL activities was about 9 minutes.

Immediately after the activities, the posttests were given without any prior indication. The retention of vocabulary was assessed with a three-week delay by administering the same battery of tests. Around one hour was the time taken by the participants to complete all the tests on both occasions. A questionnaire was also distributed to understand the participants' perceived mental effort and task difficulty.

Data Analysis

The quantitative data were analyzed using a 2 (group) × 2 (Time) repeated measures MANOVA in SPSS 26. All necessary assumptions—including normality of the dependent variable, sphericity, and the absence of significant outliers—were tested and met. When the assumption of sphericity was violated, the Greenhouse-Geisser correction was applied. The significance level was set at $p < .05$ (Larsen-Hall, 2015).

Results

The main purpose of this study was to investigate whether minimizing EL enhances the effectiveness of DGBVL activities in promoting greater word knowledge acquisition and retention.

The findings in Table 3 and Figure 4 indicate that minimizing EL may contribute to the effectiveness of DGBVL activities. During both immediate and delayed posttests, the IG outperformed the CG in terms of several parameters of word knowledge. The advantage was more pronounced in terms of productive tests, such as orthography and recall of meaning, where the differences in receptive measures were smaller but still favoured the IG group. A comparable pattern in delayed posttests suggests that minimized EL might have facilitated initial vocabulary learning and retention, even for more cognitively demanding aspects of vocabulary knowledge.

A repeated-measures MANOVA was conducted to examine the effects of group—IG vs. CG—and time—immediate vs. delayed posttest—on eight dimensions of word knowledge (see Table 4). The analysis of Pillai's Trace showed that the group had a significant main effect with $V = .35$, $F(8, 61) = 4.17$, $p < 0.001$, *partial* $\eta^2 = 0.35$, observed power ≈ 1.00. The results show that the IG maintained better performance than the CG throughout all vocabulary assessment tests. The analysis showed a significant main effect of time with $V = 0.70$ and $F(8, 61) = 18.25$, which resulted in $p < 0.001$ and *partial* $\eta^2 = 0.70$, and observed power = 1.00. This indicates that participants showed changes in performance from immediate to delayed posttests. Moreover, there was no significant interaction between time and group with $V = 0.21$, $F(8, 61) = 2.08$, $p = .051$, *partial* $\eta^2 = .21$, and observed power = .79. Therefore, although both groups exhibited score changes throughout the study, the *pattern of*

Table 3. *Descriptive statistics*

	Integration Group				Control Group			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
1st Orthography (Productive) (A)	2.09	1.87	0.00	7.00	0.89	0.63	0.00	2.00
1st Meaning Recall (Productive) (F)	2.23	2.86	0.00	10.00	0.26	0.44	0.00	1.00
1st Orthography (Receptive) (B)	**6.40**	2.33	2.00	9.00	4.60	2.30	2.00	9.00
1st Association (Productive) (H)	1.40	2.48	0.00	10.00	0.32	0.10	0.00	0.00
1st Meaning Recognition (Productive) (E)	**5.77**	3.23	0.00	10.00	4.17	2.38	0.00	8.00
1st Meaning Recall (Receptive) (D)	3.54	3.90	0.00	10.00	1.46	1.62	0.00	4.00
1st Association (Receptive) (G)	3.43	2.00	0.00	7.00	1.91	1.36	0.00	6.00
1st Meaning Recognition (Receptive) (C)	**5.46**	3.00	0.00	10.00	4.66	1.71	3.00	7.00
2nd Orthography (Productive) (A)	2.17	2.20	0.00	7.00	1.14	1.06	0.00	3.00
2nd Meaning Recall (Productive) (F)	2.26	2.98	0.00	10.00	0.11	0.32	0.00	1.00
2nd Orthography (Receptive) (B)	**5.91**	2.97	1.00	10.00	4.06	2.46	1.00	8.00
2nd Association (Productive) (H)	1.60	2.80	0.00	10.00	0.20	0.13	0.00	0.00
2nd Meaning Recognition (Productive) (E)	**4.77**	3.27	0.00	10.00	3.60	2.86	0.00	8.00
2nd Meaning Recall (Receptive) (D)	2.97	3.68	0.00	10.00	0.97	1.20	0.00	3.00
2nd Association (Receptive) (G)	2.77	2.02	0.00	7.00	1.09	1.09	0.00	4.00
2nd Meaning Recognition (Receptive) (C)	**5.34**	3.16	0.00	10.00	3.83	1.48	2.00	6.00
1st Productive (Total)	*11.49*	8.34	2.00	33.00	5.31	3.10	0.00	10.0
1st Receptive (Total)	*18.83*	9.33	5.00	35.00	12.63	5.79	6.00	21.0
2nd Productive (Total)	*10.80*	9.64	1.00	37.00	4.86	3.53	0.00	10.0
2nd Receptive (Total)	*17.00*	10.42	3.00	37.00	9.94	5.38	3.00	17.0

Note. Individual tests (A-G) maximum score is 10; Total tests maximum score is 40.

change did not differ significantly for both groups. Thus, these trends should be interpreted cautiously and not as definitive evidence of differential retention.

The analysis of pairwise comparisons (see Table 5) for the IG revealed pronounced decreases in orthography (receptive) ($M_{difference}$ = 0.51, *SE* = 0.16, p = .002) from the immediate to the delayed posttest, and also in the case of meaning recognition (productive) ($M_{difference}$ = 0.79,

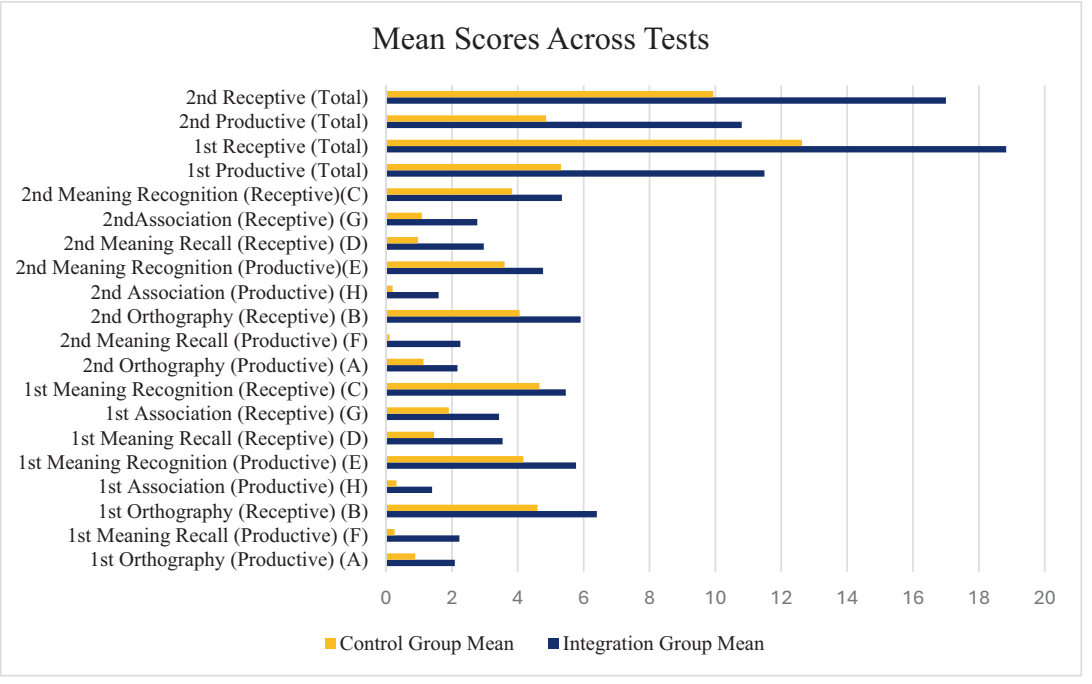

Figure 4. *Visual representation of descriptive statistics*

Table 4. *2 × 2 repeated measure MANOVA results*

Effect	Pillai's Trace	Value	F	df		P	Partial η^2	λ	Observed Power
				Hypothesis	Error				
Between Subjects	Group	0.35	4.17	8.00	61.00	0.00	0.35	33.40	≈1.00
Within Subjects	Time	0.70	18.25	8.00	61.00	0.00	0.70	145.95	1.00
	Time Group	0.21	2.08	8.00	61.00	0.51	0.21	16.7	0.79

Note. Individual tests (A-G) maximum score is 10; Total tests maximum score is 40.

SE = 0.13, $p < .001$), meaning recall (productive) ($M_{difference}$ = 0.53, *SE* = 0.20, $p = .011$), association (receptive) ($M_{difference}$ = 0.74, *SE* = 0.09), $p < .001$), and so on for the final score in productive skills ($M_{difference}$ = 2.26, *SE* = 0.33, $p < .001$). The increase in the receptive total vocabulary also reflected the IG's strong initial gains, and it was just not significant ($p > .05$). The results suggest that the IG, while showing strong initial gains, was still more susceptible to loss of productive knowledge and certain receptive skills in the course of time.

In the CG, significant declines were observed between the immediate and delayed posttests in orthography (productive) ($M_{difference}$ = –0.26, $p = .037$), orthography (receptive) ($M_{difference}$ = 0.54, $p = .004$), meaning recognition (receptive) ($M_{difference}$ = 0.83, $p < .001$), meaning recall

Table 5. *Pairwise comparisons between immediate and delayed posttests for the IG and CG*

Test	Group	Mean Difference (I–D)	SE	95% CI [LL, UL]	*p*
Orthography (Productive)	IG	–0.17	0.21	[–0.58, 0.24]	.410
	CG	–0.26*	0.12	[–0.50, –0.02]	.037
Orthography (Receptive)	IG	0.51**	0.16	[0.20, 0.83]	.002
	CG	0.54**	0.18	[0.19, 0.90]	.004
Meaning Recognition (Receptive)	IG	0.06	0.13	[–0.20, 0.31]	.659
	CG	0.83***	0.13	[0.57, 1.09]	<.001
Meaning Recall (Receptive)	IG	–0.10	0.12	[–0.34, 0.14]	.424
	CG	0.49**	0.16	[0.17, 0.80]	.004
Meaning Recognition (Productive)	IG	0.79***	0.13	[0.52, 1.05]	<.001
	CG	0.57**	0.17	[0.24, 0.91]	.001
Meaning Recall (Productive)	IG	0.53*	0.20	[0.13, 0.94]	.011
	CG	0.14	0.10	[–0.06, 0.35]	.169
Association (Receptive)	IG	0.74***	0.09	[0.57, 0.93]	<.001
	CG	0.83***	0.08	[0.67, 0.98]	<.001
Association (Productive)	IG	0.47**	0.15	[0.17, 0.77]	.003
	CG	0.00	0.00	[0.00, 0.00]	–
Receptive (Total)	IG	0.57	0.33	[–0.07, 1.21]	.084
	CG	2.69***	0.19	[2.30, 3.07]	<.001
Productive (Total)	IG	2.26***	0.33	[1.60, 2.91]	<.001
	CG	0.46*	0.18	[0.09, 0.82]	.016

Note. I = immediate posttest; D = delayed posttest; CI = confidence interval. Bonferroni-adjusted *p* values reported. $p < .05$*, $p < .01$**, $p < .001$***.

(receptive) ($M_{difference}$ = 0.49, p = .004), meaning recognition (productive) ($M_{difference}$ = 0.57, p = .001), association (receptive) ($M_{difference}$ = 0.83, $p < .001$), as well as the receptive total score ($M_{difference}$ = 2.69, $p < .001$) and productive total score ($M_{difference}$ = 0.46, p = .016). No significant changes were found in meaning recall (productive) or association (productive). Overall, the CG displayed broad declines in both receptive and productive dimensions, with stronger losses evident in receptive knowledge.

The IG and the CG exhibited insignificant but dissimilar patterns of *vocabulary loss* throughout the study period. The IG experienced a major loss in productive knowledge. On the other hand, the CG experienced a major loss in both receptive and productive measures, with the steepest loss in the receptive knowledge area. Nevertheless, the IG group managed to keep a

Table 6. *Perceived mental effort and task difficulty*

Item	Scale	IG		CG		t (68)	p	d
		M	SD	M	SD			
Mental effort invested in playing the game	1–9	3.42	1.18	5.76	1.41	−7.56	< .001	1.80
Ease/difficulty of playing the game	1–7	2.47	1.02	4.39	1.08	−7.57	< .001	1.81
Mental effort invested in studying the language	1–9	3.71	1.21	5.02	1.34	−4.34	< .001	1.03
Ease/difficulty of understanding the language	1–7	2.62	1.05	2.94	1.10	−1.27	.208	0.30

Note. Lower scores indicate lower perceived mental effort or difficulty. Negative t values indicate lower scores for IG relative to CG. Cohen's d values of .20, .50, and .80 represent small, medium, and large effects, respectively.

large part of its productive vocabulary knowledge intact during the course of the study compared with the CG, implying that EL minimization of the DGBVL activity might have contributed to more durable gains, particularly in the productive dimensions of word knowledge.

Participants were asked about their mental efforts for completing their DGBVL activities (Table 6). IG participants reported investing significantly less mental effort when playing the game (t (68) = −7.56, p < .001, d = 1.80). Also, they found the game to be significantly easier to play (t (68) = −7.57, p < .001, d = 1.81). In other words, the integration format was effective in reducing the mental efforts for playing the game itself. Moreover, IG participants found the language significantly easier to process (t (68) = −4.34, p < .001, d = 1.03), meaning that the presentation of the language component did not add any extra loads on their working memories. However, there was no significant difference between IG and CG in understanding the language (t (68) = −1.27, p = .208, d = 0.30), suggesting that this design did not support text comprehension.

Discussion

The main aim of this study was to understand whether minimizing EL can enhance IVL from DGBVL activities and contribute to the mixed results in the literature. The results showed that by the EL-minimized integration format, almost all aspects and dimensions of word knowledge might possibly be enhanced more effectively. This supports the general findings of the studies by Chen et al. (2020) and Rasti-Behbahani et al. (2024). However, this outcome contradicts the majority of previous studies in the literature. A possible explanation can be found in the study's design and the level of control over contextual factors such as the degree of context informativeness, uniformity of target words, like their lexical classes, familiarity with, and frequency band of the target words. These factors have been frequently reported to be effective on IVL (Nation, 2024; Nation & Webb, 2011). This suggests that the difference in vocabulary knowledge gains in previous studies could be due to those factors rather than EL minimization.

Why IG had higher vocabulary knowledge gain can also be explained by reference to the availability of attentional sources. According to Brünken et al. (2010), the split-attention effect

stemming from information misplacement can overload attentional resources and impair the noticing mechanism, which are the first necessary cognitive resources (Truscott, 1998) for vocabulary learning. Therefore, the IG participants might have had enough attentional resources available to invest in noticing the target words' forms, to guess the meaning from the surrounding context, and to link both forms and meaning. However, the CG participants might have constantly switched between the game screen and the paper, which possibly led them to additional and unguided searches for meaning (Rasti-Behbahani, 2023). Therefore, they might not have processed new words deeply enough to generate robust links between their forms and meanings. Table 6 reports that IG invested less effort in playing and studying the language of the game and found the game easier than CG. This suggests that IG participants likely had freer working memory and might have devoted more attentional resources to noticing the target words. Therefore, the possibility of effective noticing might have been another advantage for IG, possibly leading to higher IVL gains.

Another explanation can be the significant role of the human-computer interactions (HCI) in enhancing vocabulary acquisition from digital games (Zheng et al., 2025). Unlike CG, IG participants might have a more convenient HCI because of the availability of sentences containing target words on the game screen. This possibility might have supported them in running all necessary processes for learning the target words with the least interruption. Therefore, their searches for meaning were guided, and the design did not demand keeping information in the short-term memory while switching between the game and paper. IG participants found the game easier to play and reported minimal effort in completing it (Table 6). This also suggests that the game and its design might have been very straightforward for them because HCI load might have been minimized effectively.

Moreover, activity design could have benefited IG because they could see the words twice, on paper and in the game, while CG had them only on paper. Although exposure to the target words only twice is superficial (Webb, 2020), it apparently matters in DGBVL contexts where EL is minimized. IG could just bring a hypothesis of the meaning from the paper to the game and have the word again in front of them in a shorter, more focused context. So, they could notice and process the form more deeply. Whereas the game screen for CG might have been just the next step and not another source of information. Overall, IG participants were likely more focused, had possibly more memory capacity available to process both form and meaning deeply, and consequently might have experienced a task that was possibly less cumbersome for their memory. Hence, they could gain more word knowledge incidentally. This explanation can be supported by reference to item 4 in Table 6. Both groups found the text easy to understand; that is, the text was equally informative for both groups. However, IG had an extra chance of exposure to the target words in the game context. Therefore, it can be valid to assume that the extra exposure, although trivial based on IVL literature (Webb, 2020), could have enhanced IG's IVL gains effectively.

Another explanation for the outperformance of IG might have been the alignment of their tasks with tests, which, according to the transfer-appropriate processing hypothesis, can lead to effective learning and superior performance in the tests. The transfer-appropriate processing hypothesis "emphasizes that the value of particular acquisition activities must be defined relative to particular goals and purposes. Furthermore, assumptions about the quality and durability of the resulting memory traces can only be determined relative to the appropriateness of testing situations" (Morris et al., 1977, p. 528). In the present study, the gameplay required participants to read short contexts, infer meanings, and select pictorial representations corresponding to those meanings. Several tests drew on comparable processes. The IG participants might have experienced better alignment due to the following two reasons.

Firstly, the game interface provided direct access to target words and contextual sentences, which enabled players to process words while making decisions in the same visual space. Secondly, in the same visual space, IG participants selected meanings from multiple options that resembled the testing methods used in vocabulary assessments through multiple-choice questions. However, the CG participants first read the text on paper before they used the game interface, which lacked any linguistic context, thus creating a mismatch between their learning and assessment methods.

Another interesting finding was that productive vocabulary knowledge showed a greater relative gain in effectiveness and retention, which supports the findings by Rasti-Behbahani and Shahbazi (2022), while receptive vocabulary knowledge was acquired at a higher absolute level. The results showed that the difference between the IG and CG was proportionally much larger for productive tests, especially in productive meaning recall (IG M = 2.23 vs. CG M = 0.26). This can be explained by referring to Vyvey et al.'s (2018) findings and the concept of ludic focus. The IG group's DGBVL activity had the words embedded in the game, or say, the words were an element of the core gameplay. Vyvey et al (2018) discussed that in a situation like this, due to memory constraints and cognitive load, focus will be diverted to the core gameplay and its elements. Then, IG's DGBVL activity might have promoted ludic focus, which resulted in more effective immediate gains and retention of forms by generating numerous and robust links between form and meaning after guessing the correct meaning. The quantity of links assisted with accurate retrieval of the Arabic meanings, and their quality, with meaning retention in posttests.

The abundance of tests might have been another reason for gaining more productive knowledge. Although the order of test administration was carefully controlled, it gave the participants more chances of exposure to word forms, especially for IG. In addition to the frequency of exposure, this increased the possibility of spaced repetition, which is known for its substantial contribution to supporting IVL (Nation, 2024). In other words, by administering 8 tests, participants had the opportunity to practice the target words indirectly and unintentionally. Also, since there was no context provided, participants could focus on the target word forms more deeply and save their mental pictures in their memory easily and recall them more effectively. Furthermore, since target words were met in a different context, i.e., tests, which is different from the DGBVL activity, the possibility of the generation effect (Nation & Webb, 2011) is inevitable. Generation, or exposure to target words in different contexts, has been widely reported as an influential factor in supporting incidental vocabulary knowledge gain.

Another finding was more effective gains and retention in a few scopes and aspects of word knowledge, such as orthography (receptive) and meaning recognition (receptive and productive) by IG participants, which partially contradicted results from Vyvey et al (2018). In contrast with the game used in their study, our game was not only less visually disturbing but also encouraging and pivotal in the IG participants' focus. Therefore, they might have conveniently concentrated on and deeply processed the target word forms as the main elements of core gameplay and retained the orthographical knowledge effectively. So, when participants saw the target words in orthography recognition (receptive) posttests, they might have retrieved the correct spelling more easily due to the likely durable and deeply encoded visual representations of the target word forms in their memories.

IG participants might have remembered and retained more effectively receptive and productive knowledge of meaning recognition, possibly due to their DGBVL activity design, which may have induced a higher level of involvement load (Rasti-Behbahani, 2023). Unlike CG, in the IG's DGBVL activity, target words were presented in both textual and game screen contexts, which triggers the effect of generation (Nation & Webb, 2011). With EL minimized and no need

for unguided searches (Rasti-Behbahani, 2023), the generation effect might have strengthened the form-meaning connections both qualitatively and quantitatively. So, IG participants might have had a durable visual presentation of target words' form-meaning carved in their memory; therefore, recognizing them in posttests was easier for them. On the other hand, CG participants' DGBVL activity did not provide possibilities for triggering the generation effect or inducing a higher level of involvement load because it encouraged continuous hypothesis making and testing, focus shifting between paper and the game, and unguided searches for finding target words' meaning. Hence, with such an overloaded memory, form-meaning links might have been formed weakly and minimally.

It was also found that IG experienced less mental effort than CG globally. In other words, manipulation might have functioned as it was intended. This can be explained with reference to the split-attention effect. IG participants could see the target words on the game screen surrounded by the context. That is, they might not have experienced split-attention because they did not need to roll their eyes from the game screen to the paper repeatedly for refining their guesses. This might have kept them focused and warranted a deeper processing of form and the generation of the form-meaning links. However, we can see that CG participants reported that the game was difficult to play, they invested a bigger mental effort in playing the game, and in the language component of the DGBVL activity. It suggests that splitting their attention between the paper and the game screen might have forced them to invest great mental effort in studying the language. Therefore, they found the game effortful to play and felt that they invested a great effort in playing the game. This could also have demotivated CG participants from completing the DGBVL activity and might have changed it to an unpleasant experience. However, the EL minimization did not help with enhancing the understanding of the DGBVL activity text for IG. This suggests that EL minimization through avoiding split-attention was mostly effective in working memory management because the minimization process did not add or remove anything from the text. Also, this suggests that the superior form–meaning processing opportunities in the IG were likely an unintentional byproduct of the design rather than a direct consequence of minimizing EL itself.

Finally, the results indicated different attrition rates between groups. CG participants' delayed posttest scores showed significant attrition in almost all aspects of word knowledge, while IG participants showed a more stable gain in receptive and a smaller decrease in productive knowledge. It suggests that although the DGBVL activity might have enhanced both receptive and productive knowledge, the instructional format and activity design might have played a more effective role in incidental vocabulary knowledge gain and retention over time (Rasti-Behbahani & Shahbazi, 2022). IG format, with minimized EL, higher involvement load, and proper integration of target words as an element of core gameplay, might have regulated the quantity of triggered working memory processes and kept them between the working memory limits, and promoted the quality of deep processing necessary for vocabulary retention. As a result, words were processed deeply enough to leave durable traces on participants' memories with robust and numerous cognitive connections between the form and meaning of the target words. It might have enabled a durable retrieval over time. However, a separate format overwhelmed the CG participants' working memory with a myriad of necessary and unnecessary processes like hypothesis making and testing, focus shifting, processing pictures, etc., which contributed to shallow memory traces and temporary and weak form-meaning links, resulting in high attrition rates.

Conclusion

Overall, it can be concluded that when learners engage with a DGBVL activity that is presented in a proper instructional format and its EL is minimized, the possibility of learning and retaining

more words for longer periods of time may increase. These findings correspond with those found in CLT (Sweller, 1999) and subsequent empirical research (Tuovinen & Sweller,1999), which also explained that minimizing EL is helpful for learning because it could decrease unnecessary effort by learners and prevent split attention. However, this study showed that EL minimization alone is not enough for durable IVL gains. To have more durable incidental word knowledge gains, informed by the outcomes, firstly, factors that are related to managing working memory limits and promoting deep vocabulary processing should be considered in combination when designing DGBVL activities. Moreover, the IG format can be more beneficial for boosting word knowledge gain *when* HCI of the DGBVL activity promotes frequency of exposure, induces a high level of involvement load, and minimizes EL simultaneously. Also, vocabulary items should be embedded into the game as one of the main elements of the core gameplay to benefit from the ludic focus (Vyvey et al, 2018) for maximum IVL gains and effective retention from even a cognitively mismanaged DGBVL activity. These findings can contribute to research on vocabulary learning, the CLT field, and pedagogical implementation of DGBVL activities.

Nevertheless, there are limitations to this study that can be addressed in future studies. First, the cognitive load measure was used later, subjectively and retrospectively, to avoid its intervention with the flow of the learning process, and the cognitive load optimization effect was considered globally. Next, the target words were restricted to nouns, whereas verbs, adjectives, and adverbs were ignored. After that, the target words were all concrete. Abstract nouns were also ignored. Then, the digital game was short and limited to just one genre: arcade style. Future studies can replicate this study across a range of game genres with extended gameplay. Finally, CLT is sensitive to proficiency levels (Yeung et al., 1997), and this study was restricted to a homogeneous participant group. If this study can be replicated with different levels of proficiency, we might develop a greater understanding of the minimization of EL in DGBVL activities.

Funding

This research was supported by Dhofar University (Grant Number: DU/AY/2024-25/DURG-006).

Acknowledgment

We would like to thank all the students who participated in this study, the esteemed reviewers, and the journal editor-in-chief, Professor Glenn Stockwell, whose constructive comments and stellar support increased the quality and availability of this paper.

References

Al-shehri, S. & Gitsaki, C. 2010. Online reading: a preliminary study of the impact of integrated and split-attention formats on L2 students' cognitive load. ReCALL, 22, 356–375. https://doi.org/10.1017/S0958344010000212.

Ashcroft, R. J., Garner, J., & Hadingham, O. (2019). IVL through watching movies. *Australian Journal of Applied Linguistics*, *1*(3), 135–147. https://doi.org/10.29140/ajal.v1n3.89

Awad, H. (2022). *The impact of integrated and split-attention online reading formats on building adult EAP learners' vocabulary knowledge* [Doctoral dissertation, Newcastle University]. Newcastle University. http://hdl.handle.net/10443/6426

Brünken, R., Plass, J. L., & Moreno, R. (Eds.). (2010). *Cognitive load theory*. Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.014

Chen, H. J. H., & Hsu, H. L. (2019). The impact of a serious game on vocabulary and content learning. *Computer Assisted Language Learning*, *33*(7), 811–832. https://doi.org/10.1080/09588221.2019.1593197

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*, 1–17. https://doi.org/10.1080/10494820.2020.1863233

deHaan, J., 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. http://dx.doi.org/10125/44215

File, K. A., & Adams, R. (2010). Should vocabulary instruction be integrated or isolated? *TESOL Quarterly, 44*(2), 222–249. https://doi.org/10.5054/tq.2010.219943

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

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*(2), 689–696. https://doi.org/10125/67046

Hung, H. C. M. (2007). Reducing extraneous cognitive load by using integrated format in reading comprehension for EFL/ESL. In C. Gitsaki (Ed.), *Language and languages: Global and local tensions* (pp. 130–146). Cambridge Scholars Publishing.

Karimi, M. N., & Nasouri, A. (2024). *EFL learners' flow experience and IVL during text-based game tasks: The moderating role of working memory capacity*. System, 124, Article 103398. https://doi.org/10.1016/j.system.2024.103398

Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R*. Routledge. https://doi.org/10.4324/9781138024571

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

Liu, D. (2024). The effects of segmentation on cognitive load, vocabulary learning and retention, and reading comprehension in a multimedia learning environment. *BMC Psychology, 12*(1), 4–12. https://doi.org/10.1186/s40359-023-01489-5

Lowell, R., & Morris, R. K. (2017). Impact of contextual constraint on vocabulary acquisition in reading. *Journal of Cognitive Psychology, 29(5), 551–569*. https://doi.org/10.1080/20445911.2017.1299155

Meara, P. (1997). Towards a new approach to modelling vocabulary acquisition. in N. Schmitt & M. McCarthy (eds.), *vocabulary: description, acquisition and pedagogy* (pp. 109–121). Cambridge: Cambridge University Press.

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. In I. Vedder, I. Bartning, & M. Martin (Eds.), *Communicative proficiency and linguistic development: intersections between SLA and language testing research* (pp. 211–232). Second Language Acquisition and Testing in Europe Monograph Series 1.

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

Mondria, J. A., & Wiersma, B. (2004). Receptive, productive, and receptive+ productive L2 vocabulary learning: What difference does it make. *Vocabulary in a second language: Selection, acquisition, and testing*, *15*(1), 79–100. http://dx.doi.org/10.1075/lllt.10.08mon

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Memory and Language*, *16*(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

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

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

Nation, I. S. P. (2024). Re-thinking the principles of vocabulary learning and their applications. *Languages*, *9*(5), 160. https://doi.org/10.3390/languages9050160

Outamgharte, B., Yeou, M., & Zyad, H. (2025). Electronic L1 and L2 glosses' impact on vocabulary acquisition of Moroccan English as a foreign language learners. *Australian Journal of Applied Linguistics*, *8*(1), 2236. https://doi.org/10.29140/ajal.v8n1.2236

Peterson, M., & Swier, R. (2025). Task-based learning and innovation in CALL: Digital game-based learning. *Language Teaching Research Quarterly, 51*, 376–391. https://doi.org/10.32038/ltrq.2025.51.10

Pradheepa, A., Gurusamy, K., & Pushpanathan, T. (2025). The role of language games in enhancing vocabulary acquisition: A meta-analysis. *Australian Journal of Applied Linguistics*, *8*(1), 2073. https://doi.org/10.29140/ajal.v8n1.2073

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. (2023). Involvement load and vocabulary acquisition in digital game-based tasks. In: Reynolds, B.L. (eds) Vocabulary learning in the wild. Springer, Singapore. https://doi.org/10.1007/978-981-99-1490-6_6

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

Rasti-Behbahani, A., Abugohar, M. A., & Khallaf, N. M. (2024). Beware of what lurks in the margin: Glosses, mobile digital games, and incidental vocabulary learning. *International Journal of Interactive Mobile Technologies, 18*(20), 1–17. https://doi.org/10.3991/ijim.v18i20.50247

Soyoof, A., Reynolds, B. L., Rassaei, E., & Vázquez-Calvo, B. (2026). The effects of mobile, gamified, and paper flashcards on vocabulary and collocation learning in young language learners. *International Journal of Applied Linguistics*. Advance online publication. https://doi.org/10.1111/ijal.70095

Sweller, J. (1999). *Instructional design in technical areas*. ACER Press.

Truscott, J. (1998). Noticing in second language acquisition: A critical review. second language Research, 14(2), 103–135. https://doi.org/10.1191/026765898674803209

Tuovinen, J., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*(2), 334–341. https://doi.org/10.1037/0022-0663.91.2.334

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

Vyvey, T., Núñez Castellar, E., & Van Looy, J. (2018). Loaded with fun? The impact of enjoyment and cognitive load on brand retention in digital games. *Journal of Interactive Advertising, 18*(2), 72–88. https://doi.org/10.1080/15252019.2018.1446370

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*(1), 33–52. https://doi.org/10.1017/S0272263105050023

Webb, S. (2008). The effects of context on IVL. *Studies in Second Language Acquisition*, *30*(1), 101–128. https://doi.org/10.1017/S0272263108080045

Webb, S. (2007). The effects of synonymy on second-language vocabulary learning. *Reading in a Foreign Language*, *19*(2), 120–136. Retrieved from: https://www2.hawaii.edu/~readfl/rfl/October2007/webb/webb.html

Webb, S. (2020). IVL. 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

Yeung, A. S., Jin, P. & Sweller, J. 1998. Cognitive Load and Learner Expertise: Split-Attention and Redundancy Effects in Reading with Explanatory Notes. Contemporary Educational Psychology, 23, 1–21. https://doi.org/10.1006/ceps.1997.0951

Zhang, R., Cheng, G., & Zou, D. (2025). Effects of digital game elements on engagement and vocabulary development. *Language Learning & Technology*, *29*(1), 1–36. https://doi.org/10.64152/10125/73618

Appendix A

Target words

English Word	Pseudo-word	Freq. Band	English Word	Pseudo-word	Freq. Band
Debris	*ancon*	5K	Portrait	*tasper*	3K
Shack	*cader*	7K	Closet	*gisham*	6K
Latch	*dangy*	6K	Ember	*ictay*	9K
Skull	*faddam*	4K	Pouch	*mesut*	7K
Shovel	*sagod*	6K	Drape	*nasin*	5K

Appendix B

Context rating criteria adopted from webb (2008, p. 236)

Extremely unlikely that the target word can be guessed correctly. The text contains no contextual clues and may be misleading.	1
Information in the context may lead to partial knowledge of the target word's meaning.	2
Information in the context may make it possible to infer the meaning of the target word. However, there are a number of choices. Participants may gain partial knowledge.	3
Participants have a good chance of inferring the meaning correctly. There are few meanings that are logical apart from the correct meaning. Participants should gain at least partial knowledge.	4

Appendix C

EL-minimized text for IG

Jump6: The tasper of his father was on the wall. His father was only 22 years old in that. James stood in front of it for a moment, eyes closed, and mind relaxed. He could feel his heart pounding in his chest and the blood rushing through his veins. It was like he was back at home, sitting by his father's bedside—only this time, his father wasn't there.	ةحول سورت روراج

Appendix D

uVLT, 1000 word level: Question 1

	choice	**computer**	**garden**	**photograph**	**price**	**week**
Cost						
Picture						
Place where things grow outside						