Content list available at http://ijltr.urmia.ac.ir

Iranian Journal of Language Teaching Research

ORIGINAL ARTICLE

Urmia University

Technique Feature Analysis Revisited: Enhancing Digital Game-Based Vocabulary Learning Design

Amin Rasti-Behbahani [a, *], Amir Mashhadi [b]

[a] *Dhofar University, Oman*

[b] *Shahid Chamran University of Ahvaz, Iran*

ABSTRACT

The technique feature analysis (TFA) checklist and digital game-based vocabulary learning (DGBVL) were introduced for enhancing incidental vocabulary learning (IVL) gains. However, TFA precision was found to be random owing to factors causing cognitive overload, especially in DGBVL activities. Therefore, this study aimed to revise TFA by examining the effect of cognitive load theory (CLT)-informed DGBVL activity design on IVL. CEFR A2-level Omani-Arabic speakers (N = 140) were randomly assigned to a control and three experimental groups, where they completed DGBVL activities with optimized intrinsic (IL), extraneous (EL), or germane (GL) cognitive load for learning 10 target words incidentally. Participants were tested immediately and three weeks later on 8 aspects of word knowledge. Results showed that the experimental groups outperformed, and the TFA checklist precision was enhanced, suggesting that CLT-informed designs were effective. Pedagogically, teachers, material developers, and researchers can use our enhanced TFA checklist for designing effective DGBVL activities.

Keywords: Digital Game-Based Vocabulary Learning (DGBVL); incidental vocabulary learning; Technique Feature Analysis (TFA); Cognitive Load Theory (CLT); Involvement Load Hypothesis (ILH); video games

ARTICLE HISTORY

Received: 14 June 2025 **Revised version received:** 21 Nov. 2025

Accepted: 12 Feb. 2026 **Available online:** 1 Mar. 2026

* Corresponding author: Center for Preparatory Studies (CPS), Dhofar University, Salalah, Oman
Email address: aminrasti@du.edu.om

10.30466/ijltr.2026.56272.3050

Introduction

Incidental vocabulary learning (IVL)—the acquisition of new words through language-learning activities not explicitly designed to teach vocabulary and tested without forewarning (Webb, 2020)—can significantly enhance the quality of learners' lexicons (Mirzaei et al., 2023). However, incidental knowledge gain of words happens in small increments, which makes the process very slow (Webb, 2008). To accelerate this process, various methods and techniques have been proposed, including digital game-based vocabulary learning (DGBVL) (Rasti-Behbahani & Shahbazi, 2022) and Techniques Feature Analysis (TFA) (Nation & Webb, 2011), which form the focus of this study.

DGBVL, defined as acquiring vocabulary items through digital game-based activities (Rasti-Behbahani & Shahbazi, 2022), has emerged as a widely reported and effective method for increasing the quantitative gains of IVL (Gouniband Shooshtari et al., 2015; Karimi & Nasouri, 2024; Mashhadi & Khazaie, 2018; Rasti-Behbahani et. al., 2024; Vnucko & Klimova, 2023). Similarly, the TFA checklist (Nation & Webb, 2011) was developed to account for factors that enhance IVL both qualitatively and quantitatively. The checklist provides a systematic framework for designing effective IVL activities and includes factors that have been frequently cited in the literature as critical to IVL success.

Nation and Webb (2011) argued that the TFA checklist can predict the effectiveness of vocabulary learning activities. In other words, an IVL activity designed according to the TFA checklist is expected to yield favorable IVL outcomes because it can induce a high level of involvement load (Laufer & Hulstijn, 2001). While numerous studies support the validity of the TFA checklist (Ehsani & Karimi, 2022; Hu & Nassaji, 2016; Jahangiri & Abilipour, 2014), others have questioned its reliability (Eskandari et al., 2024; Kamali et al., 2020), particularly in the context of DGBVL (Rasti-Behbahani, 2023; San Mateo-Valdehíta & Criado de Diego, 2021). Rasti-Behbahani (2023) showed that the TFA checklist fails to accurately predict the outcomes of DGBVL activities because it does not account for factors such as cognitive load, which significantly influences learning in multimedia contexts. Previous research has also shown that cognitive overload can even negate the benefits of IVL activities, inducing a high level of involvement load (Dai et al., 2019). Therefore, there is a need to revise the TFA checklist by incorporating factors that address the interaction between cognitive systems and learning environments. This study aims to address this gap.

Technique Feature Analysis (TFA)

TFA (Nation & Webb, 2011) (see Table 1) represents an elaborated iteration of the vocabulary-learning activity indexing method originally proposed by Laufer and Hulstijn (2001). Laufer and Hulstijn (2001) argued that richer qualitative and quantitative processing of vocabulary items leads to deeper memory traces and faster retrieval. They identified three factors—need, search, and evaluation—which can each be absent, moderate, or strong. They further posited that the combination of these factors and their degrees of prominence constitutes the task-induced involvement load, where higher involvement loads result in more effective vocabulary tasks (Khoshsima & Eskandari, 2018). This framework is known as the Involvement Load Hypothesis (ILH) in vocabulary learning literature.

Despite its influence, the ILH framework, with only three factors—need, search, and evaluation—and three levels of prominence—none, moderate (+), and high (++)—for each, was criticized for its lack of precision. Researchers subsequently sought to refine the indexing method, leading to the development of the TFA checklist (Jahangiri & Abilipour, 2014; Nation & Webb, 2011; Yanagisawa & Webb, 2022; Zou, 2017).

TFA has been widely recognized for its enhanced accuracy over ILH in indexing the success rate of vocabulary learning activities (Ehsani & Karami, 2022; Hu & Nassaji, 2016; Khoshsima & Eskandari, 2018). The TFA checklist includes 18 factors that are well-established in the literature as significant contributors to vocabulary learning but were overlooked in the ILH framework. Each factor is assigned a score of 1 or 0, depending on whether it is triggered by the activity in question. As with ILH, a higher TFA score suggests a higher likelihood of the activity's success in facilitating vocabulary acquisition. The broader coverage of TFA has been attributed to its ability to detect more variances, thereby increasing its predictive accuracy (Hu & Nassaji, 2016; Yanagisawa & Webb, 2022).

Despite its broader applicability, TFA has not been without its limitations. For instance, Kamali et al. (2020) compared oral reproduction and summary writing tasks for teaching 40 English words to 66 Persian speakers. While oral reproduction proved more effective for vocabulary acquisition and retention, both tasks yielded similar TFA scores. Similarly, San Mateo-Valdehíta and Criado de Diego (2021) found that TFA failed to differentiate among three tasks with varying involvement loads, each targeting the acquisition of 12 infrequent and 12 pseudo-Spanish words. Additionally, Eskandari et al. (2024) reported that TFA did not accurately predict the superior effectiveness of composition writing and rewarding tasks, as these were assigned scores equivalent to less effective tasks in their study.

While the limitations of TFA have been acknowledged, to our knowledge, only two recent proposals for its refinement have been introduced since its inception in 2011. Kamali et al. (2020) proposed adding a fourth criterion to the generation component, introducing levels of "no, low, reasonable, and high generation" to make TFA more sensitive to the degree of generation in vocabulary learning activities. Separately, Nation (2024) suggested reducing the number of factors in the checklist to 16, based on new insights into vocabulary learning. However, these proposals remain theoretical and have yet to undergo empirical validation. Consequently, further calibration of the TFA framework is necessary to enhance its predictive accuracy, especially in multimedia contexts such as DGBVL.

Table 1
TFA Score for the DGBVL Activity in This Study

	TFA Criteria	
	Motivation	
1	Is there a clear vocabulary learning goal?	0
2	Does the activity motivate learning?	1
3	Do the learners select the word?	0
	Noticing	
4	Does the activity focus attention on the target words?	1
5	Does the activity raise awareness of new vocabulary learning?	1
6	Does the activity involve negotiation?	0
	Retrieval	
7	Does the activity involve retrieval of the word?	1
8	Is it productive retrieval?	0
9	Is it recall?	1
10	Are there multiple retrievals of each word?	1
11	Is there spacing between retrievals?	0
	Generation	
12	Does the activity involve generative use?	0
13	Is it productive?	0
14	Is there a marked change in context that involves the use of other words?	0

	Retention	
15	Does the activity ensure successful linking of form and meaning?	0
16	Does the activity involve instantiation?	0
17	Does the activity involve imaging?	1
18	Does the activity avoid interference?	1
	Maximum score	8

Cognitive Load Theory

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). These processes are referred to as 'germane loads' (GL). However, poorly designed educational interventions—such as information misplacement or overly complex prompts—can impose additional, often unnecessary cognitive processes that lead to cognitive overload, thereby hindering learning. The processes triggered by information misplacement are referred to as 'extraneous'(EL) and the ones triggered by overly complex prompts, as 'intrinsic' (IL) loads (Brünken et al., 2010). In other words, CLT should be central to meaningful learning, as "the limitations of working memory are directly relevant to instructional design issues" (Sweller, 2017, p. 7).

CLT has been shown to be effective in vocabulary acquisition. Li et al. (2022) assigned 60 Chinese EFL learners to two groups. Participants in the experimental group were exposed to multimodal inputs, while the control group received monomodal inputs for the incidental learning of 48 English words. The results indicated that the experimental group outperformed the control group in the immediate posttest but performed poorly in the delayed posttest. The authors attributed these results to CLT, explaining that multimodal inputs increase IL and require more cognitive effort than monomodal inputs. As a result, due to cognitive overload, the inputs did not transfer to long-term memory, resulting in unsuccessful learning. In another study, Liu (2024) investigated the effect of segmentation on cognitive load, vocabulary acquisition, and retention. Segmentation involves exposing learners to instructional materials in segments that align with their cognitive capacity, which can reduce EL and enhance GL (Liu, 2024). In this quasi-experimental study, Liu assigned 90 Chinese teenagers to low- or high-segmentation groups. The low-segmentation group received a text with no input enhancement or control over the content, while the high-segmentation group received the same text, but with colored target words on PowerPoint slides containing the audio version of the text, graphics, and controlled content presentation. Liu found that segmentation can effectively reduce cognitive load and enhance vocabulary acquisition and retention.

The effects of CLT on the effectiveness of ILH have also been explored. Dai et al. (2019) found that CLT played a more significant role than ILH in acquiring metaphorical collocations from dictionaries. They assigned 70 Chinese participants to three groups with different types of dictionaries. In group 1, metaphorical information was available and prominent, in group 2, it was available but not prominent, and in group 3, it was absent. Group 1 showed the highest gains in knowledge. From the ILH perspective, this type of dictionary was less effective; however, through the lens of CLT, it proved superior in enhancing collocation knowledge. The importance of CLT over ILH was further clarified by Ansarian and Kazemipour Khabazi (2021). In their study, 204 Persian-speaking participants with varying working memory capacities learned 20 captioned target words while listening to an expository text. They completed sentence writing, close deletion, and paragraph writing activities of varying involvement loads. The results indicated that participants with low working memory did not benefit from activities that induced a high involvement load, resulting in low vocabulary gains. In contrast, participants with high working memory exhibited greater receptive and productive word knowledge gains. The authors concluded that the effectiveness of ILH on vocabulary gains can vary depending on the size of working memory and the number of cognitive processes required to complete an activity.

Overall, CLT plays a central role in vocabulary knowledge acquisition. While ILH and its iterations, such as TFA, aim to facilitate uptake and accelerate vocabulary expansion, their effects seem superficial and inconsistent if CLT is not considered in the design of vocabulary learning activities. Therefore, TFA is effective only if it elicits processes that align with the constraints of learners' working memory. However, current TFA approaches do not adequately account for CLT, and a revision to this framework appears necessary.

Digital Game-Based Vocabulary Learning

Digital game-based vocabulary learning (DGBVL)—the educational use of video or digital games to extend learners' vocabulary both qualitatively and quantitatively (Rasti-Behbahani, 2023)—has been widely reported as an effective method for enhancing both receptive and productive vocabulary knowledge. For instance, Chen and Hsu (2019) found that a serious game, *Slave Trade*, could enhance 66 Chinese participants' receptive knowledge acquisition of specific words and historical events after comparing their pre and posttests results. Rasti-Behbahani and Shahbazi (2022) found that DGBVL activities can enhance the incidental acquisition of nearly all dimensions and aspects of word knowledge. They recruited 124 Persian speakers and assigned them to experimental and control groups. The experimental group learned 10 target words through DGBVL activities, while the control group practiced the same words through a fill-in-the-blank paper activity. After testing eight aspects of word knowledge three weeks later, the experimental group retained more vocabulary knowledge, particularly productive-recognition knowledge of form-meaning. Rasti-Behbahani et al. (2024) showed that adding meaning-given glosses to DGBVL activities can further support incidental vocabulary acquisition. They assigned 54 Persian speakers to meaning-given, meaning-inferred, and no-gloss DGBVL activities and found that glosses enhanced IVL. However, multiple-choice glosses had a reverse effect. Zhang et al (2025) studied the effect on 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. Overall, DGBVL activities were found to be more supportive of IVL than traditional pen-and-paper methods (Vnucko & Klimova, 2023).

DGBVL activities have also been shown to support IVL by boosting motivation. For example, Sadeghi et al. (2022) found that gamified instruction increased Turkish participants' motivation, leading participants to perceive gamified vocabulary learning as an efficient method, despite a small increase in vocabulary knowledge. Karimi and Nasouri (2024) found that DGBVL activities can enhance motivation and vocabulary acquisition by fostering a state of 'flow,' in which learners are fully immersed and engaged in the task. They further found that game-induced flow can directly affect IVL, and language-induced flow can predict vocabulary learning outcomes.

Also, few studies have addressed the role of CLT in IVL through DGBVL activities. DeHaan et al. (2010) examined the role of interactivity in the recognition and recall of English word forms. They assigned 80 Japanese teenagers to two groups: one that played a *rhythm game*, and another that watched the game. The players were required to press a specific button to keep the game character rapping along, while the watchers simply observed. 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 induced by high interactivity, which hindered vocabulary acquisition in the playing group. Mohsen (2016) replicated this study using a less *interactive simulated* surgery game 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 vocabulary learning.

Finally, a few studies have explored the effect of involvement load induced by DGBVL activities on vocabulary acquisition. Reynolds (2017) asked 92 Taiwanese university students to play *Draw Something* (*Casual*) and report their attitudes toward the game as a vocabulary learning tool. Participants reported three stages of involvement: searching for words, recognizing their needs, and evaluating their understanding. While Reynolds concluded that involvement load could enhance the effectiveness of DGBVL activities, he did not empirically measure its impact. Later, Rasti-Behbahani (2023) empirically studied the effect of ILH on IVL through DGBVL activities (*Adventure*). He recruited 33 B1-level Persian teenagers and assigned them to three DGBVL activities, inducing low, moderate, and high levels of involvement load (LIL) for learning 20 low-frequency English nouns, or the names of inanimate objects. Notably, he indexed the activities with TFA. In a quasi-experimental design, he conducted the study and found that DGBVL-induced involvement load can *partially* enhance IVL gains because the results showed that moderate load was less effective than low load. His think-aloud data revealed that the induced loads were less effective because of inappropriate activity designs, which had led to cognitive overload. He explained further that his activity designs encouraged an intense and unguided *search* that was previously found as a major factor causing cognitive overload and interference with learning (Brünken et al., 2010). Notably, in another study, the same results were found. Rasti-Behbahani (2025) compared TFA and ILH in predicting the success of DGBVL activities. He recruited 105 Arabic speakers with A1 proficiency and assigned them to three DGBVL activities (*Point-and-click*) that induced low, moderate, and high involvement loads. Although TFA was better than ILH in prediction, its predictions were partially accurate because moderate loads were more effective than high loads.

Evidently, CLT and ILH highlight key factors that can shape the effectiveness of DGBVL activities regardless of digital game genres and context. Also, CLT determines the extent to which ILH can be effective.

However, current frameworks, such as TFA, are not sensitive to factors that may cause cognitive overload, resulting in mixed and sometimes random effects. This issue has yet to be addressed in depth. Therefore, the primary aim of this study is to revise TFA to make it sensitive to factors that can cause cognitive overload. To do this, it is necessary first to test if DGBVL designs consistent with CLT can enhance the effect of ILH and consequently IVL. Next, informed by the results, it can be decided on the necessary revision of the current TFA checklist. Hence, this study addresses the following questions:

1. Does optimizing cognitive load enhance the effectiveness of LIL in promoting greater word knowledge acquisition and retention?
2. What revisions are necessary to improve the TFA checklist sensitivity?
3. Which sources of cognitive load need to be optimized to significantly enhance the effectiveness of LIL and TFA precision?

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). The rationale behind the study is that if optimizing cognitive load can enhance the effect of DGBVL activity-induced LIL, the TFA checklist could be revised by incorporating a new category dedicated to cognitive load optimization.

Participants

Participants in this study were 140 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. Their general English language proficiency level was at CEFR A2, measured by the online Cambridge English Proficiency Test (CEPT; Scores between 20 and 29). The test is explained in more detail in the instruments section below. To guarantee that participants had enough vocabulary knowledge for proper comprehension of the texts used in the DGBVL activities, the first and second bands of most frequent English words on the Updated Vocabulary Levels Test (uVLT) (Webb et al., 2017) were administered. According to the vocabulary profile of the text (M_{1K} = 82.8%; M_{2K} = 12.3%; $M_{\leq 3K}$ = 4.9%; 1K and 2K text coverage: 95%), participants needed to demonstrate mastery in the 1K band. Therefore, the cut-off point was set to 29 (Webb et al., 2017) for 1K. However, because only 12.3% of the whole text comprised 2K word families, no cut-off point was defined for the 2K band.

Participants' first band score was 95.42% (M = 28.63, SD = 1.77), second band score was 70.04% (M = 21.01, SD = 4.23), and overall score was 82.73% (M = 24.81, SD = 3.59). Because participants had a high enough score in 1K and an acceptable mean in both 1k and 2K, we could assume that they had the mastery of the level (Webb et al., 2017) and a high enough vocabulary knowledge (González-Fernández & Schmitt, 2020) to understand the texts in DGBVL activities. Notably, the scores of eight participants were found to be unsatisfactory, and they were subsequently 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

The Digital Game

An arcade-style digital game was developed by the researchers on a web platform (www.educaplay.com). This platform was selected for designing the game because it is accessible on any internet-enabled device, provides game elements such as a leaderboard, time tracking, and scoring, helps teachers easily convert any content into games that are short and do not take up class time, and is a popular game-based learning platform in the context of the study. In the game, players must help a frog cross a lake by jumping onto specific lotus leaves. Each jump presents three options (lotus leaves), and selecting the incorrect leaf results in the frog drowning, causing the player to lose a life and points, after which they have another opportunity to choose. The ability to add user-selected texts, pictures, and audio to each game mode enabled researchers to optimize cognitive load. Two versions of the game were created, as explained in later sections (see Figures 1 and 2).

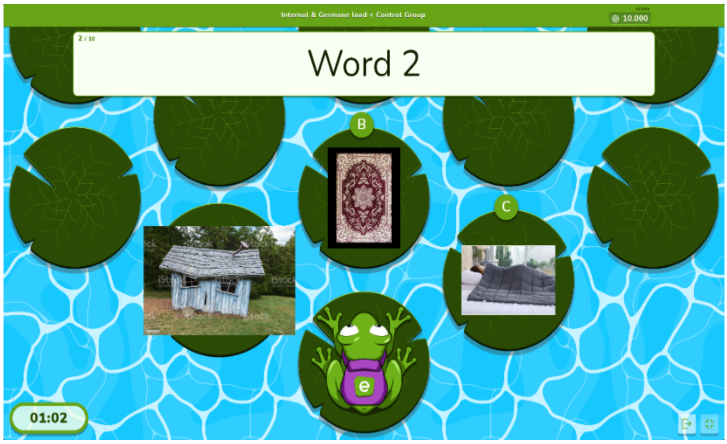

Figure 1. The Game Designed for The Intrinsic Load, Germane Load, and Control Groups

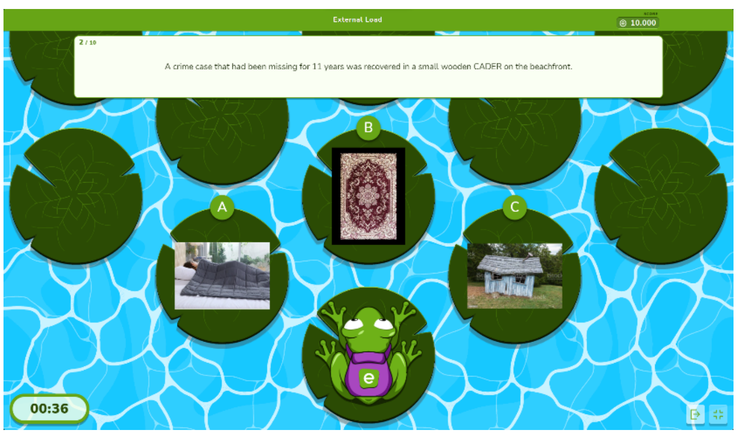

Figure 2. The Game Designed for the Extraneous Load Group

Target Words

Ten nouns representing inanimate objects were selected for the study (see Table 2), 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 VocabProfiler 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 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.

Table 2
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

DGBVL Activity Texts

Since cognitive load 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 the surrounding texts to be consistent with CLT. 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 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 to 2500. The overall vocabulary profile of the text (M_{1K} = 82.8%; M_{2K} = 12.3%; $M_{\leq 3K}$ = 4.9%; 1K and 2K text coverage: 95%) showed that the mastery of 1K band was critical to guarantee proper understanding of the text.

Additionally, three native English-speaking teachers reviewed and revised the texts according to the scale defined by Webb (2008) (see Table 3) to ensure all paragraphs were moderately informative. Prior research indicates that moderately informative contexts are more effective for learning novel words (Lowell & Morris, 2017). By following this approach, the cognitive load was standardized across all participants (see Table 5).

In total, 10 short, moderately informative paragraphs were composed. Each paragraph consisted of three to four sentences and contained one target word.

Table 3
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

DGBVL Activity Design and TFA Indexing

In the general design of DGBVL activities, there were 10 short paragraphs including a target word each. The target words were presented in bold and underlined within each paragraph. Three Arabic definitions were provided alongside each paragraph for each target word. It was ensured that no definition appeared more than twice in the margins (see Tables 4–6 for a sample). Paragraphs were later printed on paper and handed over to the participants.

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 & 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.*"

All DGBVL activities were scored by the TFA checklist to ensure that they induce the same LIL and have the same TFA scores. All DGBVL activities were scored 8 (see Table 1) because they activated the following components in the TFA checklist: 2, 4, 5, 7, 9, 10, 17, and 18. In other words, component 2 across all activities was triggered and received 1 score because digital games enhance motivation (Sadeghi et al., 2022). The use of bold text to highlight target words triggered component 4. Employing vocabulary learning techniques ensured the activation of component 5. Components 7, 9, and 10 were engaged, as participants encountered each target word at least twice, facilitating recall. Component 17 was activated through the inclusion of object images in the game. Finally, component 18 was triggered by ensuring that words with opposite meanings, such as hot and cold, were not introduced simultaneously (Nation & Webb, 2011). The scoring process was reviewed and rechecked by an applied linguistics professor.

CLT-Informed Optimizations

To design DGBVL activities that are consistent with CLT, four versions of the DGBVL activity were designed.

Managing Intrinsic Load (IL)

To manage the IL, first, the sentences surrounding the target words were translated into Arabic (see Table 4). According to Brünken et al. (2010), background knowledge—such as a learner's first language (L1)—can reduce IL and facilitate the mental representation of a portion of the text, leading to a better understanding of other parts. In this context, since not all portions of the text are unfamiliar, the IL is less likely to hinder learning (Tuovinen & Sweller, 1999).

In the game, only images were added, and no additional information was provided (see Figure 1).

Table 4
IL Text

Jump2: لقد تم الكشف عن جريمة ترجع لإحدى عشرة سنة مضت داخل cader خشبي صغير على شاطي البحر A policeman found three dead bodies buried there. It was a well-hidden place, although old and broken. The police chief stated that this discovery could help solve many of their unsolved cases.	موقد سجادة كيس

Minimizing Extraneous Load (EL)

To minimize the EL, the text itself was not modified (see Table 5). However, the sentences containing the target words were incorporated into the game. 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 (see Figure 2). This design was based on two 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 reduce EL.

Table 5
EL Text

Jump2: A crime case that had remained for 11 years was recovered in a small wooden cader on the beachfront. A policeman found three dead bodies buried there. The police chief stated that this discovery could help solve many of their unsolved cases. It was a well-hidden place, although old and broken.	كوخ موقد سجادة

Maximizing Germane Load (GL)

To maximize GL, concrete appositives were added in parentheses immediately after the target words (see Table 6). Appositives provide additional information about the target words, and in this study, adjectives were used to describe the objects in greater detail. This approach was intended to offer limited guidance by loosely indicating the end goal (Paas et al., 2001)—that is, selecting the correct meaning—while also introducing contextual interferences (Kirschner, 2002). Both strategies have been shown to effectively maximize GL, facilitate schema acquisition, and enhance retention.

In the game, only images were added, and no additional information was provided (see Figure 1).

Table 6
GL Text

Jump2: A crime case that had remained for 11 years was recovered in a small wooden cader *(it is like a small house in a jungle)* on the beachfront. A policeman found three dead bodies buried there. The police chief stated that this discovery could help solve many of their unsolved cases. It was a well-hidden place, although old and broken.	كوخ موقد سجادة

Control Group (CG)

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

Instruments

Cambridge English Proficiency Test (CEPT)

CEPT is a standardized proficiency test administered online and comprises two modules: reading and listening (Cambridge English, 2025). The test uses adaptive technology, meaning that the difficulty of each question is adjusted automatically based on the participants' performance. Therefore, no two test takers receive the same set of questions. The highest score in this test is 50, and each score corresponds to a specific level on the CEFR, as shown in Table 7.

Table 7
CEPT and CEFR (Cambridge English, 2025)

CEFR Level	50-Point Scale
Below A1	0 – 9
A1	10 – 19
A2	20 – 29
B1	30 – 39
B2	40 – 49
C1 or above	50

Updated Vocabulary Level Test

On uVLT (reliability for this study = 0.87), 1000- and 2000-word levels, which encompass 60 receptive recognition questions, were administered. These levels were selected because a vocabulary size of nearly 2K was necessary to understand the text in the DGBVL activities. There are three definitions and six words in each question. Participants had to match one definition with one word (see Table 8). With no time limit, a paper-based format was administered.

Table 8
uVLT, 1000 Word Level - 1st Question

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

Achievement Tests

Given the multidimensional nature of word knowledge (Rasti-Behbahani & Shahbazi, 2022), a comprehensive assessment of vocabulary acquisition across multiple dimensions was essential to evaluate the effect of the CLT-informed DGBVL designs. 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 ensures that the information encountered in the first test does not provide the answers for the next one. For instance, test A did not provide the orthographical forms of the target words to be helpful for test B, or test B did not provide useful information for choosing the definition in test C, and so on.

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.

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. **كيس** ☑ *b.* بطانية *c.* لوحة *d.* الستارة

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 حطام, 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 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.

The tests were administered without prior notice, with no time constraints, immediately following the activity and again three weeks later in the same setting. Participants completed the tests in approximately one hour during each session.

Procedure

This study employed an experimental design with three experimental groups and a control group (see Figure 3). The experimental groups engaged in three distinct DGBVL activities that either controlled or enhanced cognitive load. Group names reflected their respective activities. The vocabulary learning approach was incidental, as participants were unaware of the immediate and delayed posttests or that the DGBVL activities were designed to teach vocabulary (Webb, 2020).

In accordance with the university's research ethics policy, participants were informed of their rights and signed consent forms written in Arabic, which provided detailed information about their participation. A total of 140 CEFR A2-level (informed by their CEPT results) English learners were selected. Out of 148 volunteers, 8 were excluded due to insufficient uVLT scores. The remaining participants were randomly assigned to the IL, EL, GL, and CG groups, with 35 participants in each. In designated classrooms, participants of each group gathered and completed the DGBVL activities individually on their mobile phones or tablets. To increase motivation through competition, a game leaderboard was displayed, and prizes (three packs of chocolate) were awarded to the top three performers in each group. The average time to complete the activities was approximately 8 minutes.

Following the activities, immediate posttests were administered without prior notice. The same set of tests was re-administered three weeks later in the same classrooms to assess vocabulary retention. Participants completed all tests within approximately one hour on both occasions. After the delayed posttests, participants were informed that the words were not real and the real words with their Arabic definitions were projected on the projector screen for them to read.

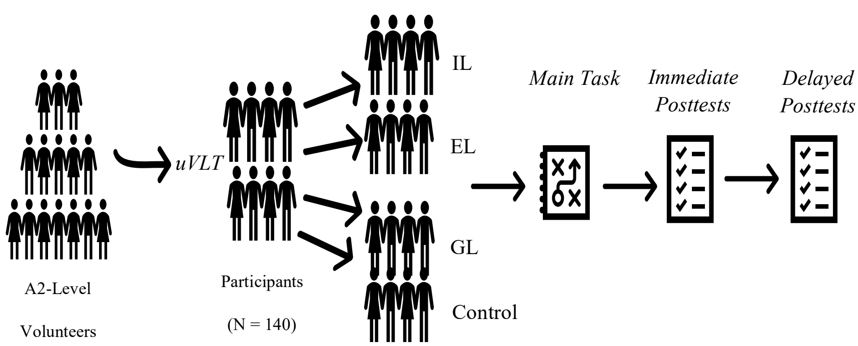

Figure 3. A Graphic Illustration of the Study Process

Data Analysis

The data were analyzed using MANOVA and Bonferroni-adjusted posthoc tests in SPSS 26. All statistical assumptions (e.g., multivariate normality, homogeneity of covariance matrices) necessary for these analyses were met, and the significance level was set at $p < .05$ (Larsen-Hall, 2015).

Results

The results are explained in two parts. Since research questions 1 and 2 are closely related, they are addressed together. After that, question 3 is answered elaborately.

Q1 and Q2: Optimizing sources of cognitive load enhances the effectiveness of LIL in promoting greater word knowledge acquisition and retention, and there is a need to revise the TFA checklist by incorporating a category based on CLT.

Table 9 and figure 4 illustrate that the overall performance of the IL, EL, and GL groups surpassed that of the CG group in all immediate and delayed posttests. Additionally, all groups demonstrated greater effectiveness in acquiring and retaining receptive knowledge compared to productive knowledge.

Regarding the components of word knowledge, Table 9 indicates that receptive knowledge of orthography and meaning recognition, as well as productive knowledge of meaning recognition, were acquired and retained more effectively than other aspects of word knowledge by the experimental groups. Notably, receptive knowledge of orthography was acquired (M = 7.46) and retained (M = 8.49) most effectively by GL participants. In contrast, participants in the IL group achieved the highest immediate ($M_{productive}$ = 8.23; $M_{receptive}$ = 8.20) and delayed ($M_{productive}$ = 6.89; $M_{receptive}$ = 6.74) gains in productive and receptive meaning recognition. Overall, receptive and productive word knowledge retention was most effective for GL participants, followed by IL, EL, and CG participants.

Table 9
Descriptive Statistics

	Germane Load				Internal Load				External Load				Control Group			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
1st Orthography (Productive) (A)	2.83	1.44	0.00	6.00	3.51	1.46	1.00	6.00	2.09	1.87	0.00	7.00	0.89	0.63	0.00	2.00
1st Meaning Recall (Productive) (F)	1.89	2.73	0.00	10.00	1.54	1.99	0.00	6.00	2.23	2.86	0.00	10.00	0.26	0.44	0.00	1.00
1st Orthography (Receptive) (B)	7.46	2.02	2.00	10.00	7.26	2.47	0.00	10.00	6.40	2.33	2.00	9.00	4.60	2.30	2.00	9.00
1st Association (Productive) (H)	2.69	3.03	0.00	10.00	1.31	2.77	0.00	10.00	1.40	2.48	0.00	10.00	0.32	0.10	0.00	0.00

1st Meaning Recognition (Productive) (E)	6.20	2.60	1.00	10.00	8.23	1.96	4.00	10.00	5.77	3.23	0.00	10.00	4.17	2.38	0.00	8.00
1st Meaning Recall (Receptive) (D)	4.09	2.94	0.00	10.00	6.94	2.83	3.00	10.00	3.54	3.90	0.00	10.00	1.46	1.62	0.00	4.00
1st Association (Receptive) (G)	4.66	1.57	0.00	7.00	4.06	2.24	0.00	8.00	3.43	2.00	0.00	7.00	1.91	1.36	0.00	6.00
1st Meaning Recognition (Receptive) (C)	6.51	2.85	0.00	10.00	8.20	2.18	2.00	10.00	5.46	3.00	0.00	10.00	4.66	1.71	3.00	7.00
2nd Orthography (Productive) (A)	2.74	2.15	0.00	8.00	3.94	3.24	1.00	10.00	2.17	2.20	0.00	7.00	1.14	1.06	0.00	3.00
2nd Meaning Recall (Productive) (F)	2.60	3.13	0.00	10.00	1.00	0.94	0.00	3.00	2.26	2.98	0.00	10.00	0.11	0.32	0.00	1.00
2nd Orthography (Receptive) (B)	8.49	2.28	1.00	10.00	7.86	1.82	2.00	10.00	5.91	2.97	1.00	10.00	4.06	2.46	1.00	8.00
2nd Association (Productive) (H)	2.49	3.47	0.00	10.00	1.03	1.60	0.00	5.00	1.60	2.80	0.00	10.00	0.20	0.13	0.00	0.00
2nd Meaning Recognition (Productive) (E)	6.11	3.39	0.00	10.00	6.89	2.25	2.00	10.00	4.77	3.27	0.00	10.00	3.60	2.86	0.00	8.00
2nd Meaning Recall (Receptive) (D)	4.11	3.23	0.00	10.00	4.06	3.11	0.00	10.00	2.97	3.68	.00	10.00	0.97	1.20	0.00	3.00
2ndAssociation (Receptive) (G)	4.37	1.68	2.00	7.00	4.23	1.66	1.00	6.00	2.77	2.02	0.00	7.00	1.09	1.09	0.00	4.00

2nd Meaning Recognition (Receptive) (C)	*6.29*	2.82	0.00	10.00	*6.74*	3.38	0.00	10.00	*5.34*	3.16	0.00	10.00	*3.83*	1.48	2.00	6.00
1st Productive (Total)	*13.60*	7.35	4.00	33.00	*14.60*	5.69	7.00	25.00	*11.49*	8.34	2.00	33.00	5.31	3.10	0.00	10.00
1st Receptive (Total)	*22.71*	7.73	7.00	36.00	*26.46*	7.05	15.00	37.00	*18.83*	9.33	5.00	35.00	12.63	5.79	6.00	21.00
2nd Productive (Total)	*13.94*	9.67	1.00	38.00	*12.86*	5.21	4.00	21.00	*10.80*	9.64	1.00	37.00	4.86	3.53	0.00	10.00
2nd Receptive (Total)	*23.26*	8.01	9.00	36.00	*22.89*	8.32	3.00	34.00	*17.00*	10.42	3.00	37.00	9.94	5.38	3.00	17.00

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

To ensure that vocabulary acquisition was solely the result of CLT-informed activity design, first an independent sample *t*-test was run to compare the combined means of modified to the mean of baseline DGBVL activities (see Table 10).

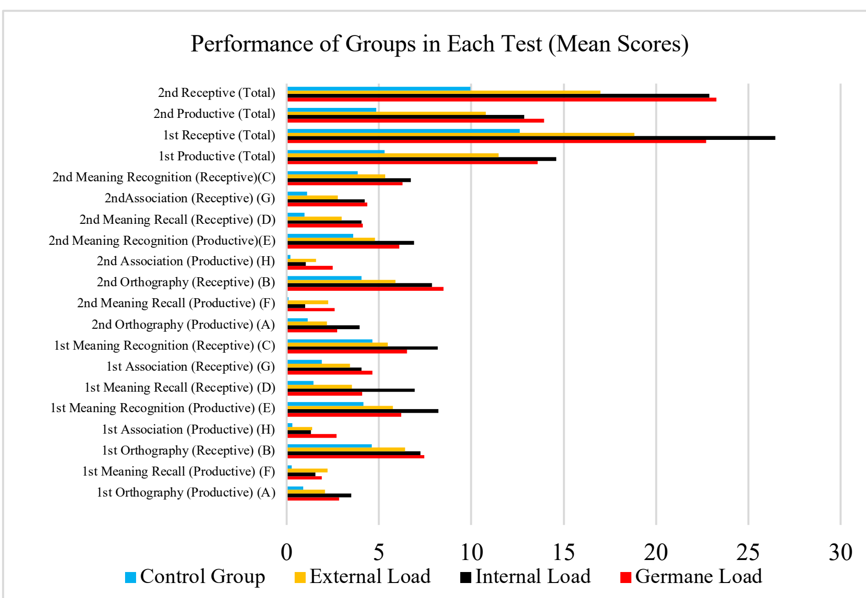

Figure 4. Visual Summary of Descriptive Statistics

The *t*-test results showed significant differences between modified and baseline DGBVL activities in both receptive and productive immediate and delayed posttests with large effect sizes for total scores and small-to-medium - strongest 2nd Association (Receptive) (d = 0.67) and 2nd Orthography (Receptive) (d = 0.57) - for individual tests. It suggests that higher gains in word knowledge for the GL, EL, and IL were not due to the digital game but cognitive load optimization.

Table 10
Results of t-test: Understanding the Effect of the Digital Game

	Modified DGBVL		Baseline DGBVL		T(138)	*p*	Cohen's d
	M	SD	M	SD			
1st Orthography (Productive) (A)	2.80	1.69	0.88	0.6	6.55	0.00	0.56
1st Meaning Recall (Productive) (F)	1.88	2.54	0.25	0.44	3.75	0.00	0.32
1st Orthography (Receptive) (B)	7.03	2.30	4.60	2.30	5.42	0.00	0.46
1st Association (Productive) (H)	1.80	2.81	0.00	0.00	3.77	0.00	0.32
1st Meaning Recognition (Productive) (E)	6.73	2.83	4.17	2.38	4.81	0.00	0.41
1st Meaning Recall (Receptive) (D)	4.85	3.55	1.45	1.61	5.45	0.00	0.46
1st Association (Receptive) (G)	4.04	2.00	1.91	1.35	5.86	0.00	0.50
1st Meaning Recognition (Receptive) (C)	6.72	2.90	4.65	1.71	3.97	0.00	0.34
2nd Orthography (Productive) (A)	2.95	2.66	1.14	1.06	3.91	0.00	0.33
2nd Meaning Recall (Productive) (F)	1.95	2.62	0.11	0.32	4.12	0.00	0.35
2nd Orthography (Receptive) (B)	7.41	2.62	4.05	2.46	6.66	0.00	0.57
2nd Association (Productive) (H)	1.70	2.77	0.00	0.00	3.62	0.00	0.31
2nd Meaning Recognition (Productive) (E)	5.92	3.11	3.60	2.86	3.90	0.00	0.33
2nd Meaning Recall (Receptive) (D)	3.71	3.35	0.97	1.20	2.74	0.00	0.23
2nd Association (Receptive) (G)	3.79	1.92	1.08	1.09	7.90	0.00	0.67
2nd Meaning Recognition (Receptive) (C)	6.12	3.15	3.82	1.48	4.14	0.00	0.35
1st Productive (Total)	13.22	7.25	5.31	3.10	6.25	0.00	1.22
1st Receptive (Total)	22.66	8.60	12.62	5.79	6.42	0.00	1.26
2nd Productive (Total)	12.53	8.45	4.85	3.53	5.21	0.00	1.24
2nd Receptive (Total)	21.04	9.34	9.94	5.38	6.66	0.00	1.36

Next, a MANOVA test was run to test whether optimizing different sources of cognitive load in DGBVL significantly enhance vocabulary knowledge gain differently. The MANOVA results (Table 11) revealed a statistically significant difference between the four groups on the combined dependent variables, Pillai's Trace = .685, $F(48, 369) = 5.26$, $p < .001$, confidence interval = 95% [14.35, 15.79]. These results provide sufficient evidence to reject the null hypothesis, concluding that designing DGBVL activities consistent with CLT significantly enhanced the effectiveness of LIL in promoting greater word knowledge acquisition and retention, and there is a need to revise the TFA checklist by incorporating a category based on CLT. The effect size was large (partial η^2 = .40) (Larsen-Hall, 2015), and the observed power was 1.00.

Table 11
Results of MANOVA Test

Effect	*Value*	*F*	*df*		*P*	*Partial* η^2	*Noncent. Parameter*	*Observed Power*
			Hypothesis	*Error*				
Pillai's Trace	1.22	5.26	48.00	369.00	0.00	0.40	252.59	1.00

Q3: *The optimization of specific sources of cognitive load will significantly enhance the effectiveness of involvement load and TFA precision.*

To understand the effect of cognitive load optimization better, a Bonferroni-adjusted comparison test was conducted to learn optimization of which source of cognitive load can result in more effective incidental word knowledge gain and retention.

The Bonferroni-Adjusted Comparisons (Table 12) indicate that overall GL participants had the best performance, followed by IL and EL. Specifically, it revealed that the experimental groups showed no significant differences in the immediate acquisition of receptive knowledge of orthography. However, in the delayed posttest, EL participants performed significantly worse than GL and IL participants in recognizing the forms of target words. All experimental groups significantly outperformed the CG group in both tests.

For the immediate productive meaning recognition posttest, Table 12 shows that IL participants significantly outperformed all other groups. In the delayed posttest, however, their performance was not significantly different from GL participants, although they still outperformed EL and CG participants. Differences between GL and EL performances in both tests were insignificant.

Similarly, IL participants outperformed all other groups in the immediate receptive meaning recognition posttest, while their differences from GL participants were insignificant in the delayed posttest.

Lastly, while the experimental groups exhibited differences in individual tests, they showed no significant differences in the acquisition and retention of productive knowledge of target words. However, significant differences were observed in receptive knowledge gains. Table 12 shows that IL participants achieved the highest immediate receptive knowledge gains (M = 26.46), significantly outperforming all other groups. GL (M = 22.71) and EL (M = 18.83) achieved the second and third-highest gains, with significant differences between them and the CG group. For retention, however, the IL group's advantage (M = 22.89) diminished, as their difference from GL participants became insignificant. GL participants exhibited the highest retention performance in the delayed receptive posttest (M = 23.26).

Table 12
Bonferroni-Adjusted Comparisons

Dependent Variable	Group	Group	Mean Difference	SE	*p*	Dependent Variable	Group	Group	Mean Difference	SE	*p*
1st Orthography (Productive) (A)	GL	IL	-.6857		0.27	2nd Orthography (Productive) (A)	GL	IL	-1.2000		0.18
		EL	.7429		0.18			EL	.5714		1.00
		CG	1.9429*		0.00			CG	1.6000*		0.02
	IL	EL	1.4286*	0.34	0.00		IL	EL	1.7714*	.549	0.00
		CG	2.6286*		0.00			CG	2.8000*		0.00
	EL	CG	1.2000*		0.00		EL	CG	1.0286		0.37
1st Meaning Recall (Productive) (F)	GL	IL	.3429		1.00	2nd Meaning Recall (Productive) (F)	GL	IL	1.6000*		0.01
		EL	-.3429		1.00			EL	.3429		1.00
		CG	1.6286*		0.01			CG	2.4857*		0.00
	IL	EL	-.6857	.531	1.00		IL	EL	-1.2571*	.530	.019
		CG	1.2857		0.10			CG	.8857		.097
	EL	CG	1.9714*		0.00		EL	CG	2.1429*		0.00
1st Orthography (Receptive) (B)	GL	IL	.2000		1.00	2nd Orthography (Receptive) (B)	GL	IL	.6286		1.00
		EL	1.0571		0.33			EL	2.5714*		0.00
		CG	2.8571*		0.00			CG	4.4286*		0.00
	IL	EL	.8571	.546	0.71		IL	EL	1.9429*	.578	0.00
		CG	2.6571*		0.00			CG	3.8000*		0.00
	EL	CG	1.8000*		0.00		EL	CG	1.8571*		0.01
1st Association (Productive) (H)	GL	IL	1.3714		0.10	2nd Association (Productive) (H)	GL	IL	1.4571		0.06
		EL	1.2857		0.15			EL	.8857		0.71
		CG	2.6857*		0.00			CG	2.4857*		0.00
	IL	EL	-.0857	.573	1.00		I L	EL	-.5714	.565	1.00
		CG	1.3143		0.14			CG	1.0286		0.42
	EL	CG	1.4000		0.09		EL	CG	1.6000*		0.03
1st Meaning Recognition (Productive) (E)	GL	IL	-2.0286*		0.00	2nd Meaning Recognition (Productive) (E)	GL	IL	-.7714		1.00
		EL	.4286		1.0			EL	1.3429		0.36
		CG	2.0286*		0.00			CG	2.5143*		0.00
	IL	EL	2.4571*	.617	0.00		IL	EL	2.1143*	.711	0.02
		CG	4.0571*		0.00			CG	3.2857*		0.00
	EL	CG	1.6000		0.06		EL	CG	1.1714		0.61
1st Meaning Recall (Receptiv e) (D)	GL	IL	-2.8571*		.000	2nd Meanin g Recall (Recept ive) (D)	GL	IL	.0571		1.00
		EL	.5429	.701	1.00			EL	1.1429	.708	0.65
		CG	2.6286*		0.00			CG	3.1429*		0.00

	IL	EL	3.4000*		.000		IL	EL	1.0857		0.76
		CG	5.4857*		.000			CG	3.0857*		0.00
	EL	CG	2.0857*		0.02		EL	CG	2.0000*		0.03
1st Association (Receptive) (G)	GL	IL	.6000		1.00	Association (Receptive) (G)	GL	IL	.1429		1.00
		EL	1.2286*		0.03			EL	1.6000*		0.00
		CG	2.7429*		0.00			CG	3.2857*		0.00
	IL	EL	.6286	.436	0.91		IL	EL	1.4571*	.393	0.00
		CG	2.1429*		0.00			CG	3.1429*		0.00
	EL	CG	1.5143*		0.00		EL	CG	1.6857*		0.00
1st Meaning Recognition (Receptive) (C)	GL	IL	-1.6857*		0.03	2nd Meaning Recognition (Receptive) (C)	GL	IL	-.4571		1.00
		EL	1.0571		0.47			EL	.9429		0.97
		CG	1.8571*		0.01			CG	2.4571*		0.00
	IL	EL	2.7429*	.595	0.00		IL	EL	1.4000	.671	0.23
		CG	3.5429*		0.00			CG	2.9143*		0.00
	EL	CG	.8000		1.0		EL	CG	1.5143		0.15
1st Productive (Total)	GL	IL	-1.0000		1.00	2nd Productive (Total)	GL	IL	1.0857		1.00
		EL	2.1143		1.00			EL	3.1429		0.49
		CG	8.2857*		0.00			CG	9.0857*		0.00
	IL	EL	3.1143	1.53	0.26		IL	EL	2.0571	1.79	1.00
		CG	9.2857*		0.00			CG	8.0000*		0.00
	EL	CG	6.1714*		0.00		EL	CG	5.9429*		0.00
1st Receptive (Total)	GL	IL	-3.7429		0.24	2nd Receptive (Total)	GL	IL	.3714		1.00
		EL	3.8857		0.20			EL	6.2571*		0.01
		CG	10.0857*		0.00			CG	13.3143*		0.00
	IL	EL	7.6286*	1.81	0.00		IL	EL	5.8857*	1.96	0.02
		CG	13.8286*		0.00			CG	12.9429*		0.00
	EL	CG	6.2000*		0.00		EL	CG	7.0571*		0.00

In summary, these findings demonstrate that CLT-informed DGBVL activity design can enhance various aspects, dimensions, and scopes of receptive and productive word knowledge. Specifically, managing cognitive load sources, such as IL, can improve immediate receptive knowledge gains from DGBVL activities. Furthermore, depending on the design of the DGBVL activities, maximizing GL processes can lead to long-term retention of word knowledge.

Moreover, these results suggest that CLT-informed DGBVL activity design enhances the effects of involvement loads induced by DGBVL activities. Thus, the hypothesis of this study is supported. Revising the TFA checklist by incorporating a cognitive load optimization category—comprising IL, GL, and EL—can significantly improve its precision.

Discussion

This study investigated whether designing DGBVL activities consistent with CLT can enhance the effectiveness of levels of involvement load and ultimately IVL gains. The results reinforced previous findings (Ansarian & Kazemipour Khabazi, 2021; Dai et al., 2019; deHaan, 2010; Li et al., 2022; Liu, 2024; Mohsen, 2016) and demonstrated that in addition to motivation (Sadeghi et al., 2022), internal elements of digital games (Rasti-Behbahani et al., 2024; Zheng et al., 2025), and inducing a high level of involvement load (Rasti-Behbahani, 2023), managing IL and minimizing EL, along with maximizing opportunities for GL, can significantly improve the effectiveness of DGBVL activities in facilitating vocabulary acquisition.

Furthermore, it was found that gains in certain components of word knowledge, such as productive knowledge of meaning recognition, were higher, aligning with previous findings (Rasti-Behbahani & Shahbazi, 2022). However, overall gains in receptive knowledge were both higher and more stable than gains in productive knowledge, which diverges from the findings of Rasti-Behbahani (2023) and Sundqvist (2019). A possible explanation for this discrepancy is the nature of activity completion: Rasti-Behbahani's (2023) study utilized pair work, while this study adopted an individual task structure. De la Fuente (2002) reported that pair work can enhance productive knowledge more effectively than individual work. Another explanation can be individual differences, such as mother tongue, gaming experience, gameplay style, and time spent playing games. Unlike this study, Sundqvist (2019) recruited participants who were Swedish speakers, familiar with many English words due to the abundance of cognates in their mother tongues, and experienced gamers who were playing massively multiplayer online adventure games weekly for long hours. However, in this study, participants were Arabic speakers with very few English cognates in their language, writing, and reading in a different alphabet and writing system, and learned with an arcade-style game.

Another plausible explanation could be the effect of previous tests. Although tests were administered carefully to control for the learning-from-the-test factor, visiting new word forms multiple times could have enhanced form acquisition (Rasti-Behbahani & Shahbazi, 2022).

The second and most significant finding was that optimizing each source of cognitive load can differentially enhance various dimensions, aspects, and scopes of word knowledge while also influencing the longevity of new gains. For instance, the results revealed that managing IL significantly supported immediate gains in receptive knowledge of orthography and both productive and receptive knowledge of meaning recognition. In contrast, maximizing GL not only enhanced immediate gains in similar dimensions but also facilitated retention over time. This can be explained by the effects of unguided search, as discussed by Sweller (1999), who argued that unguided searches for meaning significantly reduce meaningful learning because they do not leave sufficient working memory capacity for learners to engage in essential schema development processes. In the context of DGBVL, this phenomenon was corroborated by recent studies (Rasti-Behbahani, 2023; Reynolds, 2017). Managing IL through the use of L1 translations freed participants from the need for extensive searching, enabling them to rely on background knowledge and focus exclusively on the target words. As a result, they had sufficient cognitive resources to activate essential processes, such as connecting target word forms with meanings, background knowledge, and their existing lexicon, thereby generating meaningful schemas. Evidence for this was observed in the IL participants' gains in association receptive knowledge, as they could recall relevant words. However, after 3 weeks, the lack of extensive information about the target words resulted in weaker links between new knowledge and existing knowledge, making retrieval less efficient.

In contrast, maximizing GL reduced the need for searching by providing more information about the target words, thus leading to the availability of more cognitive resources for processing the target words. This likely enabled participants to make better guesses, add detailed information about the target words to their lexicon, and develop robust schemas that could be retrieved promptly. The lower rate of word knowledge attrition in the GL group supports this interpretation. Furthermore, the GL activity may have induced higher LIL that were not detected by TFA, possibly due to its limited sensitivity (Rasti-Behbahani, 2025). When paired with reduced cognitive load, the GL condition likely facilitated the formation of strong connections between target word forms, meanings, and existing lexical knowledge, leading to substantial learning gains and more reliable retrieval.

While minimizing EL also reduced cognitive load by minimizing distractions, it was not as effective as IL and GL optimizations. Minimizing EL mainly prevented participants from splitting their attention between multiple sources of information (e.g., the text and the game), but it did not sufficiently support schema development. Consequently, EL participants acquired and retained target words more effectively than the control group but lagged behind the IL and GL groups.

The interaction between different sources of cognitive load may also explain these findings. According to Tuovinen and Sweller (1999), these loads can either enhance or counteract one another. IL participants, for example, may have processed and understood the text more easily due to L1 text, freeing up cognitive resources for additional processes such as GL. However, because there was insufficient information about the target words, participants developed numerous but weak links between new word knowledge and their existing lexicon. While these links facilitated faster retrieval immediately after activity completion, their fragility resulted in diminished retention over time. On the other hand, GL participants likely formed stronger and more durable links between new vocabulary knowledge and their lexicon, thanks to the availability of additional information about the target words. This not only reduced IL but also enhanced retention. However, due to the creation of fewer links overall, GL participants retrieved the newly learned words more slowly immediately after activity completion compared to IL participants. Nevertheless, their retention was superior over time due to the robustness of these links. Minimizing EL was less effective because it may not have sufficiently reduced cognitive load to allow for efficient GL or IL processing, thereby risking cognitive overload (Tuovinen & Sweller, 1999).

Conclusion

The ultimate goal of this study was to refine the TFA for designing more effective DGBVL activities. The findings demonstrated the importance of optimizing sources of cognitive load in DGBVL activities. Moreover, the results indicated that optimizing cognitive load without simultaneously supporting schema development and the generation of numerous, high-quality links between new and existing knowledge can result in superficial and transient vocabulary acquisition. Therefore, it can be concluded that inducing appropriate LIL while optimizing cognitive load may enhance both vocabulary acquisition and retention. Consequently, it is justifiable to propose the addition of a new category to the TFA framework (see Appendix 1 for the full version):

Table 13
Suggested Category for TFA Checklist

Cognitive Load Optimization		
Is intrinsic cognitive load managed?	1	-1
Is extraneous cognitive load minimized?	1	-1
Is germane cognitive load maximized?	1	-1

With the inclusion of this category, the maximum TFA score could range from 15 to 21. The addition of a -1 score reflects the detrimental effect of neglecting cognitive load on vocabulary acquisition, whereas optimizing cognitive load, as demonstrated in this study, can boost involvement load and ultimately improve vocabulary acquisition and retention.

From a pedagogical perspective, the outcomes of this study offer valuable insights for both researchers and language teachers. It is recommended that they first ensure their DGBVL activities induce high LIL and subsequently design them to be consistent with CLT. Teachers and researchers should also note that DGBVL activities inducing high LIL are more effective for immediate vocabulary acquisition when IL is managed and for vocabulary retention when GL is maximized. So, if the DGBVL activities are followed by more drills focusing on new vocabulary items, managing IL suffices. However, if the DGBVL activity is the only activity that addresses new vocabulary items, maximizing GL is necessary.

Nevertheless, caution should be exercised when implementing the findings of this study due to several limitations. First, this study focused exclusively on nouns, neglecting other lexical categories. Second, the target words consisted solely of concrete nouns, excluding abstract nouns. Third, the repeated administration of tests may have contributed to incidental learning. Fourth, the study was conducted exclusively within a DGBVL context. Fifth, the game genre was limited to arcade and educational games, while other genres may yield different results. The sixth limitation was that although participants were homogeneous in their lexical knowledge, they had different levels of gaming experience. Sevent, the number of females was higher than that of male participants. Eighth, participants were low-proficient English learners. The ninth limitation was that the magnitude of cognitive load was not measured directly. Finally, the combined effects of cognitive load optimizations were not considered, leaving room for future studies to address these gaps.

Future research could replicate this study using other lexical categories and more number of participants with higher proficiency levels. Also, they can compare the newly proposed TFA with other indexing methods to measure its precision and reliability in DGBVL activity design. The new TFA checklist is also recommended for indexing IVL activities designed beyond digital games and multimedia contexts, in order to examine its broader applicability.

Acknowledgment:

We would like to acknowledge Ms. Erica Bailey for her support in improving the language and clarity of this manuscript. Also, we would like to thank all the students who participated in this study, the esteemed reviewers, and the journal editor-in-chief, Professor Karim Sadeghi, whose constructive comments increased the quality of this paper.

Funding:

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

References

Ansarin, A. A. & Kazemipour Khabbazi, S. (2021). Task-induced involvement load and working memory: Effects on active and passive vocabulary knowledge of EFL learners in a multimedia learning environment. *Eurasian Journal of Applied Linguistics, 7*(1), 277–302. https://doi.org/10.32601/ejal.911288

Brünken, R., Plass, J. L., & Moreno, R. (Eds.). (2010). *Cognitive load theory*. Cambridge University Press.

Cambridge English. (2025, October 8). *Cambridge English Placement Test (CEPT) – FAQs*. Cambridge English Support Site. https://support.cambridgeenglish.org/hc/en-gb/articles/210044206-Cambridge-English-Placement-Test-CEPT-FAQs

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

Dai, Y., Wu, Z., & Xu, H. (2019). The effect of types of dictionary presentation on the retention of metaphorical collocations: Involvement load hypothesis vs. cognitive load theory, *International Journal of Lexicography*, *32*(4), 411–431, https://doi.org/10.1093/ijl/ecz010

De La Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary: The roles of input and output in the receptive and productive acquisition of words. *Studies in second language acquisition*, *24*(1), 81-112. https://doi.org/10.1017/S0272263102001043

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

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.

Eskandari, Z., Khatin-Zadeh, O., Farsani, D., & Banaruee, H. (2024). The effect of type of task on EFL learners' vocabulary learning. *Frontiers in Psychology, 15*. https://doi.org/10.3389/fpsyg.2024.1306306

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

Gouniband Shooshtari, Z., Khazaie, S., & Mashhadi, A. (2015). An investigation into the impact of didactic games on L2 English pedagogy. *Journal of Educational Sciences*, *22*(1), 165-192. https://doi.org/10.22055/edu.2015.13831

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

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.

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

Kamali, M., Behjat, F., & Bagheri, M. S. (2020). Examining the effects of oral reproduction and summary writing vocabulary tasks on L2 word learning: Technique feature analysis on trial. *Cogent Education, 7*(1). https://doi.org/10.1080/2331186X.2020.1795966

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

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://doi.org/10.22059/ijeap.2018.66578.

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction*, *12*(1), 1-10. https://doi.org/10.1016/S0959-4752(01)00014-7

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., & 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

Li, W., Yu, J., Zhang, Z., & Liu, X. (2022). Dual coding or cognitive load? Exploring the effect of multimodal input on English as a foreign language learners' vocabulary learning. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.834706

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

Mashhadi, A., & Khazaie, S. (2018). Familiar or unfamiliar context?: Application of M-games in the blended module of L2 learning. In *Online course management: Concepts, methodologies, tools, and applications* (pp. 29–47). IGI Global. https://doi.org/10.4018/978-1-5225-5472-1.ch027

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.

Mirzaei, A., Farhang, M., & Eslami, Z. (2023). Reading-based incidental vs. intentional focus on lexis and development of L2 phraseological competence. *Iranian Journal of Language Teaching Research*, *11*(2), 33-53. https://doi.org/10.30466/ijltr.2023.121329

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

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

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

Paas, F., Camp, G., & Rikers, R. (2001). Instructional compensation for age-related cognitive declines: Effects of goal specificity in maze learning. *Journal of Educational Psychology, 93*(1), 181–186. https://doi.org/10.1037/0022-0663.93.1.181

Rasti-Behbahani, A. (2023). Involvement load and vocabulary acquisition in digital game-based tasks. In: Reynolds, B.L. (eds) *Vocabulary learning in the wild* (pp. 163-189). Springer, Singapore. https://doi.org/10.1007/978-981-99-1490-6_6

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

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

Reynolds, B. L. (2017). Evidence for the task-induced involvement construct in incidental vocabulary acquisition through digital gaming. *The Language Learning Journal*, *45*(4), 466-484.

Sadeghi, K., Sağlık, E., Mede, E., Samur, Y., & Comert, Z. (2022). The effects of implementing gamified instruction on vocabulary gain and motivation among language learners. *Heliyon, 8*(11), e11811. https://doi.org/10.1016/j.heliyon.2022.e11811

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

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

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

Sweller, J. (2017). Cognitive load theory and teaching English as a second language to adult learners. *Contact Magazine*, *43*(1), 10-14.

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

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. (2008). The effects of context on incidental vocabulary learning. *Studies in Second Language Acquisition*, *30*(1), 101-128. https://doi.org/10.1017/S0272263108080045

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

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. Doi:10.1017/s0272263121000577

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

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

Appendix 1
Enhanced Technique Feature Analysis Checklist

	Enhanced TFA Criteria	
	Motivation	
1	Is there a clear vocabulary learning goal?	0 1
2	Does the activity motivate learning?	0 1
3	Do the learners select the word?	0 1
	Noticing	
4	Does the activity focus attention on the target words?	0 1
5	Does the activity raise awareness of new vocabulary learning?	0 1
6	Does the activity involve negotiation?	0 1
	Retrieval	
7	Does the activity involve retrieval of the word?	0 1
8	Is it productive retrieval?	0 1
9	Is it recall?	0 1
10	Are there multiple retrievals of each word?	0 1
11	Is there spacing between retrievals?	0 1
	Generation	
12	Does the activity involve generative use?	0 1
13	Is it productive?	0 1
14	Is there a marked change in context that involves the use of other words?	0 1
	Retention	
15	Does the activity ensure successful linking of form and meaning?	0 1
16	Does the activity involve instantiation?	0 1
17	Does the activity involve imaging?	0 1
18	Does the activity avoid interference?	0 1
	Cognitive Load Optimization	
19	Is intrinsic cognitive load managed?	1 -1
20	Is extraneous cognitive load minimized?	1 -1
21	Is germane cognitive load maximized?	1 -1
	Maximum Score	15 to 21

Amin Rasti-Behbahani has a PhD in Applied Linguistics and currently works as an English lecturer at Dhofar University – Centre for Preparatory Studies (CPS). His main areas of research interest are digital game-based learning, out-of-school language learning activities, Extramural English (EE), informal English learning, Cognitive Load Theory, and vocabulary learning/acquisition.

Amir Mashhadi is an Associate Professor of Applied Linguistics, specializing in Technology-Mediated Second Language Acquisition at Shahid Chamran University of Ahvaz, Iran. He has published numerous articles in esteemed domestic and international journals, and has actively participated in both national and international conferences. In addition to his research activities, he serves as the Managing Editor of the *Journal of Research in Applied Linguistics* and sits on the editorial boards of several other domestic and international journals.