System 96 (2021) 102396

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

# The effect of visuospatial designing elements of zoomable user interfaces on second language vocabulary acquisition

Saad Alzahrani *, Leah Roberts

*The University of York, UK*

ARTICLE INFO

*Article history:*
Received 12 June 2019
Received in revised form 6 October 2020
Accepted 25 October 2020
Available online 28 October 2020

*Keywords:*
Computer-assisted language learning
Multimedia
Second language acquisition
Vocabulary learning



## 1. Introduction

This study explores the differential influences of targeted features of instructional software for intentional multimedia vocabulary learning. To achieve this, it investigates multimedia coding elements with potential for a differential impact on second language vocabulary acquisition (SLA) by connecting computer-assisted language learning (CALL) software design to theoretically grounded SLA research. The majority of SLA studies involving adults (with several exceptions, e.g. Klein & Dimroth, 2009; Gullberg, Roberts, & Dimroth, 2012) have investigated intermediate L2 learner's abilities in terms of processing or developing linguistic representations. Unlike that majority of research, the current study controls for the influence of pre-existing linguistic knowledge, by choosing participants with no background in the L2. The intention was being to capture the abilities of learners when they encounter an L2 that is typologically distant. The participants were exposed to Arabic vocabulary as a foreign language (AFL) via the Computer-Assisted Vocabulary Acquisition software (CAVA) via three interactive interfaces in a multimedia CALL environment: a verbal-based menu driven interface (L2-L1: MDI), a visual-based graphical user interface (L2-Picture: GUI) and a visuospatial-based zoomable user interface (L2-Context: ZUI). To the best of the researchers' knowledge, ZUI is used here for the first time in field of vocabulary CALL, and is presented alongside CAVA as an instructional software tool for intentional multimedia vocabulary learning designed by the researchers. Therefore, this study will add to the body of knowledge by exploring the effective features of three multimedia representations to provide further insight to improve CALL design.

* Corresponding author.
*E-mail addresses:* saad.alzahrani@york.ac.uk (S. Alzahrani), leah.roberts@york.ac.uk (L. Roberts).

https://doi.org/10.1016/j.system.2020.102396
0346-251X/

## 2. Literature review

### 2.1. Teaching vocabulary via multimedia CALL programs

Multimedia CALL designers have been influenced by the assumptions of SLA researchers, a combination of intentional and incidental vocabulary learning can support L2 vocabulary learning (e.g., Cobb, 2007; Godwin-Jones, 2010; Min, 2008; Nation, 2013; Schmitt, 2010). Incidental vocabulary learning refers to non-deliberate learning of vocabulary, which occurs when learners perform tasks, such as communication, reading or listening. Conversely, intentional vocabulary learning refers to the deliberate learning of vocabulary by directing explicit attention towards a lexical item. Laufer (2006) argued that if a teacher cannot create an L2 learning environment with similar input conditions to those associated with first-language acquisition then intentional vocabulary learning becomes crucial.

The use of multimedia CALL is a relatively more effective tool for incidental L2 vocabulary learning (e.g., Sato, Matsunuma, & Suzuki, 2013) and intentional learning (e.g., Esit, 2011; Hirschel & Fritz, 2013; Khezrlou, Ellis, & Sadeghi, 2017; Rusanganwa, 2015), when compared with traditional pen and paper learning methods. For example, Sato et al. (2013) emphasized automatized L2 word recall with their subjects by exposing them to incidental vocabulary learning and multimedia software with a time-control function, for comparison with other subjects using printed text. They concluded that CALL facilitated the decoding of word forms, and freeing up working memory to support better comprehension of the text. The advantage of CALL over non-CALL vocabulary learning was also observed in L2 intentional learning. For instance, Esit (2011) reported a positive effect from intentional vocabulary CALL software that overlapped traditional methods of teaching English as a foreign language to Turkish learners. More recently, Hirschel and Fritz (2013) compared the use of notebook and CALL in the context of L2 vocabulary learning; their 140 Japanese students participated in pre-, post-, and delayed post-tests. Despite no significant differences between the two groups on the immediate post-test, the multimedia CALL group performed slightly better on the delayed post-test. In Rusanganwa's (2015) study, multimedia presentation, including visual stimuli in combination with aural presentation, enabled EFL students to learn L2 vocabulary more effectively, outperforming their peers learning by chalk and talk. In a recent study, however, Khezrlou et al. (2017) found that if learners explicitly learned a non-CALL L2-L1 list of the target words before reading computer-based texts including multimedia glosses, they made greater gains in terms of L2 vocabulary in word recognition tests, and maintained long-term knowledge of the target words better than participants who did not.

Other studies moved away from comparing vocabulary CALL with non-CALL interventions. Researchers investigated the impact of incidental vocabulary learning of multimedia annotations on different gloss types (e.g., Boers, Warren, Grimshaw, & Siyanova-Chanturia, 2017; Chen, 2016; Ramezanali & Faez, 2019; Çakmak & Erçetin, 2018), and considered the effects when using different multimedia during intentional communication (e.g., Dubois & Vial, 2000; Fagehi, 2013; Kaplan-Rakowski & Loranc-Paszylk, 2019; Kim, 2019; Kim & Gilman, 2008).

The majority of previous studies on multimedia vocabulary CALL programs have investigated incidental vocabulary acquisition. A comparatively smaller body of research on multimedia vocabulary CALL relates to intentional learning. One of the striking findings of these studies is that communicating the meaning of a word through verbal and pictorial annotations leads to more effective learning than presenting either one of the two annotations; i.e., findings that have often been related to Paivio's (1991) dual-coding theory.

Plass, Chun, Mayer, and Leutner (1998) aimed to establish "For whom is multimedia instruction effective?" by looking at individual differences between verbalizers and visualizers, based on the participant's choices during interactions with multimedia representations. They presented learners with an English language story to teach them L2 words through translation, spoken words, pictures, and video clips. Of the 103 participants, reading comprehension was found to improve if new vocabulary was annotated with both verbal input and images, rather than with only one of these. Plass et al. (1998) obtained results supporting multiple annotated representations of learning efficacy, reporting significant effects on vocabulary acquisition when presenting multiple annotations. Similar results were noted by Plass, Chun, Mayer, and Leutner (2003). Yeh and Wang (2003) also explored the preference for these types of vocabulary annotations within the context of vocabulary learning when reading an English passage. The results of the study revealed the presence of the text and a picture together was the most effective.

More recently, Çakmak and Erçetin (2018) reported no effect from different types of multimedia glosses on facilitating recognition and producing vocabulary while listening to a story with a mobile phone. Mohsen (2016), in an incidental multimedia listening learning environment, found learners performed slightly better on delayed vocabulary recall tests when provided with captions rather than transcripts during the learning phase (including animations and annotations). Türk and Erçetin (2014) found presenting both pictorial and verbal annotations in the glosses led to better L2 incidental vocabulary learning than alternatives. Yanguas' (2009) study revealed recognition of target L2 words was significantly better when using multimedia glosses to incidentally communicate meaning visually and verbally, when compared with presenting only one medium. Rouhi and Mohebbi (2013) also noted a positive effect from multimedia glosses on L2 vocabulary learning, with both high and low spatial ability learners. Sun and Dong (2004) argued that L2 vocabulary multimedia tools should be accompanied by additional learning support, i.e. sentence-level translation or target warming-up to boost efficiency in young learners.

Despite the variety of multimedia technologies produced to date, previous research in the field of L2 multimedia vocabulary CALL has been largely restricted to limited comparisons of the effects of different annotations, such as texts, audio,

pictures and videos. In contrast with previous vocabulary CALL studies, the present research aims explicitly to employ technology as a way to teach visuospatial aspects of L2 word meaning through ZUI.

### 2.2. The zoomable user interface (ZUI)

There is potential added value for ZUI over GUI and MDI, due to its richer and more advanced design. ZUI representations can communicate additional aspects of target L2 word knowledge. Showing the whole view of an office, for instance, will concretely represent the spatial relationship between the chair and the desk, imitating what the learner would observe in real-life situations. Many aspects of scenes can be focused on individual objects, by zooming in for a closer look, or offering a holistic overview by zooming out.

By positioning words in close proximity to their related objects, ZUI can be exploited to employ the spatial contiguity principle as proposed by Moreno and Mayer (1999). They found participants learned better when texts and related visuals were in close proximity on the screen. This linking of the word with its location could afford an extra memory cue which is more likely to be formed when multimedia lessons contain both words, corresponding pictures, and locations together to explain to-be-learned content.

ZUIs are engaging (Ware, 2004), and all learning requires behavioural and psychological engagement. *Psychological engagement* describes the cognitive processing of content in ways that can support the acquisition of new knowledge and skills. Some cognitive processes that result in learning include paying attention to relevant material, and mentally organizing it for coherence (Clark & Mayer, 2011). Clark and Mayer (2011) discovered that adding relevant graphics to words can be a powerful way to support learners to engage in active learning.

In the current study, ZUI was designed to serve as a platform to view layouts that stimulate the human spatial memory. The learner will be able to view words together in a single layout such that the relationship between words can be seen and associations can be easily made between items. This could facilitate construction of a spatial model to describe information as an internal mental representation.

As discussed above, the majority of the multimedia vocabulary CALL experimental studies mentioned focused on assessing the impact of technology on strengthening the connection between the form of a word and its meaning (mainly comparing different multimedia annotations). Additionally, to date intentional language-focused vocabulary learning interventions have failed to effectively employ multimedia to present and practice other aspects of word knowledge beyond form-meaning connections.

### 2.3. The study

The current study aims to determine the veracity of the hypothesis that L2 learners' long and short-term vocabulary acquisition varies significantly when employing different multimedia representations (i.e. MDI, GUI, ZUI) to acquire L2 word meaning. This study aims to contribute to developing our understanding of the vocabulary acquisition process, in turn, the field of second language acquisition: specifically, how zoomable visuospatial representations can enhance vocabulary learning for learners in the initial phase of the learning process. To achieve this, the study answers the following research question:

What is the effect on second language vocabulary acquisition of the three different interactive representation types; namely verbal-based Menu Driven Interface (L2-L1: MDI), visual-based Graphical User Interface (L2-Picture: GUI), and visuospatial-based Zoomable User Interface (L2-Context: ZUI)?

## 3. Methodology

### 3.1. Participants

The study participants in this research were fifty monolingual British undergraduate students at the University of York (30 females and 20 males). The participants' ages ranged between 18 and 23 years old, with a mean of 20. None had previously learned Arabic, and they only learned it for this experiment. All the participants avoided studying the target vocabulary between the immediate post-test and the delayed post-test.

### 3.2. Design of the experiment

The experiments investigated the efficacy of three multimedia representations (i.e., verbal-based Menu Driven Interface (L2-L1: MDI), visual-based Graphical User Interface (L2-Picture: GUI), and visuospatial-based Zoomable User Interface (L2-Context: ZUI)), for vocabulary acquisition. The Computer-Assisted Vocabulary Acquisition (CAVA) software presented the participants with 24 words divided into six lessons (four words in each lesson).

The independent variable considered in this study was multimedia representation type. The dependent variable was vocabulary acquisition. The students' acquisition of target words was operationalized based on the scores they received in the immediate and delayed post-tests.

**Table 1**
Multimedia representation distribution in the Latin Square design of the study.

| Lesson | Exposure Group 1 (CAVA 1) | Exposure Group 2 (CAVA 2) | Exposure Group 3 (CAVA 3) |
|---|---|---|---|
| Office | ZUI | MDI | GUI |
| Food | GUI | ZUI | MDI |
| Occupations | MDI | GUI | ZUI |
| Clothes | ZUI | MDI | GUI |
| Kitchen | GUI | ZUI | MDI |
| Class | MDI | GUI | ZUI |

### 3.3. Intervention

To answer the research question using CAVA, it was important to ensure distribution of the three representations to balance all the words, not only in the study phases, but also in the practice exercises phase. There were two main objectives here: first, all the learners had to be exposed to all three multimedia representations, to allow a fair comparison of the efficiency of each representation for each learner. Second, each multimedia representation had to be assessed alongside all the target words, to control for any potentially confounding effects arising from the learning burden imposed on the words through representational efficiency. To realize these objectives, a Latin square within-subjects experimental design was employed. It is a type of partial counterbalancing design that controls for the effects of practice and fatigue.

In a Latin square experimental design, the measurements in any condition are repeated relative to the other conditions for each participant. In the current study, the experimental design for the Latin square was employed to represent each treatment (multimedia representation) equally in each column, and in each row as a way to control for the two sources of extraneous variation (word and learner) (see Table 1).

In the Latin square design implemented here, each lesson was studied by different participants in all the multimedia representations. For each participant, the multimedia representation of the word was consistent across the phases of CAVA (study phases and practice phase). Moreover, 24 words were split thematically into 6 lessons within the Latin square design (see Table 2).

To clearly measure the effect, and minimize extraneous differences between the three conditions of CAVA (MDI, GUI, and ZUI), all the target L2 words were chosen from concrete nouns. This avoided potential confounders and concrete nouns were more easily represented pictorially. However, using words to communicate an idea might convey the distinctive characteristics of an object. Neither a picture of a mouse, for instance, nor the word *mouse* communicates all its characteristics. Thus, three CAVA conditions were designed to convey different distinctive characteristics associated with the target L2 word, with the expectation of a greater effect from the new multimedia design ZUI over MDI and GUI.

### 3.4. Apparatus

With regard to delivery methods for the different aspects of L2 word knowledge targeted in CAVA, the researchers took into consideration Nation's (2001) list of aspects involved in the receptive and productive knowledge of words. In CAVA, information is provided pertaining to all three aspects of word knowledge, including form, meaning, and the use of the target words (see Table 3). In addition, CAVA utilized three user interfaces (MDI, GUI, and ZUI) to teach different levels of *meaning*, whereas *form* and *use* aspects were consistent across all three conditions.

#### 3.4.1. Three different sources of information of the word's meaning

In addition to the word's written and spoken forms, CAVA provided three different sources of information: L1 translation (MDI), pictures (GUI), and visuospatial word associations (ZUI). The target word يسرك *(chair)* for example had to be covered in CAVA with different aspects of the word knowledge involved. Default information in CAVA (i.e., L2 written and spoken form, L2 example sentence) was presented to all students for every word. However, representational information (i.e., L1 translation, picture, word associations) was only provided to learners when studying the word with a user interface that includes this type of representation.

**Table 2**
The six four-word lessons in CAVA.

| Lessons | Words |
|---|---|
| بتكم (Office) | ةلواط (Desk), يسرك (Chair), وبساح (Laptop), باودل (Bookcase) |
| امعط (Food) | ضيب (Eggs), بيلح (Milk), زبخ (Bread), ءام (Water) |
| فئاظو (Occupations) | يئاطفإ (fireman), يطرش (Policeman), يعاس (postman), ىهان (Painter) |
| سبالم (Clothes) | لاطنب (Trousers), ةعبق (Hat), صيمق (Shirt), ءاذح (Shoes) |
| خبطم (Kitchen) | رد ق (Pot), ناخس (Kettle), نرف (Oven), ةلسغم (Sink) |
| لصف (Class) | حول (Board), باتك (Book), ةبيقح (Bag), ملق (Pen) |

**Table 3**
Aspects of word knowledge covered by CAVA.

| Aspect of word knowledge | Information | Delivery method in CAVA |
|---|---|---|
| Form | Written form | Default information (all methods) |
| | Spoken form | Default information (all methods) |
| Meaning | L1 translation | Representation type (MDI) |
| | Picture | Representation type (GUI) |
| | Word association | Representation type (ZUI) |
| Use | L2 example sentence | Default information (all methods) |

*3.4.2. Program flow*

CAVA program flow was created to establish clear links between learning objectives and course content. The program did not contain a pre-test since all the learners were complete beginners. The program was comprised of five phases: two study phases, a practice phase, an immediate post-test (*follow-up test 1*), and a delayed post-test (*follow-up test 2*).

*3.4.3. Developing the menu-driven interface (MDI)*

The first type of representation in CAVA was MDI, which presented a list of target L2 words, with their equivalent translations (see Fig. 1). In this representation, L1 translations were delivered as the meaning information pertaining to target Arabic words.

Following receipt of the instructions, the learner was required to click on the identified word, which was hyperlinked to another slide (see in Fig. 2.). This new slide presented the selected target word, translated with a single word that expresses its most common meaning (e.g., *chair* as the translation of كرسي).

*3.4.4. Developing the graphical user interface (GUI)*

Pictures were used to communicate information related to the target words in the second type of multimedia representation. The GUI displayed the list of target L2 words alongside illustrative pictures. In this representation, a set of coloured photographs was presented to illustrate the meaning of target nouns (see Fig. 3). To eliminate any possible impact from the cultural dimension on the meaning and use of the vocabulary (Nation, 2001), the pictures were taken from the participants' culture.

The learners then were instructed to click on the directed word. Each pair of word and its picture was linked via an unseen background box, hyperlinked to the related study slide (see Fig. 4.). After clicking on the word the learner was directed to the study slide to study the word at their own pace.

*3.4.5. Developing the zoomable user interface (ZUI)*

The third type of multimedia representation in CAVA was ZUI, which presented a visuospatial layout of target L2 words affording environmental context. In this interface, the words' visuospatial associations were presented to the learner as semantic information in addition to the visuals as a part of a larger visual representation involving pictures of other related words. Word association is encouraged in L2 vocabulary learning as "understanding these relations is useful for explaining the meanings of the words and for creating activities to enrich learner's understanding of the words" (Nation, 2013, p. 79).


**Fig. 1.** The home page of the lesson (My office) in MDI.

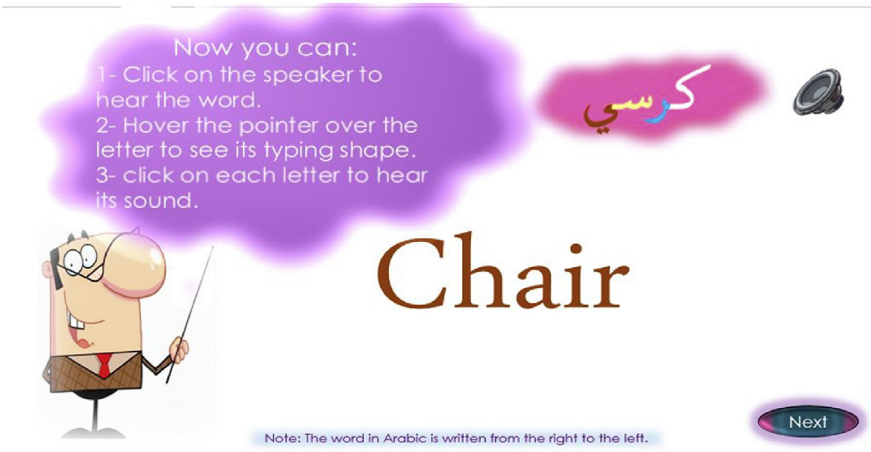


**Fig. 2.** CAVA screenshot of the target word كرسي *(chair) in MDI representation.*


**Fig. 3.** The home page for the lesson (My office) in GUI.

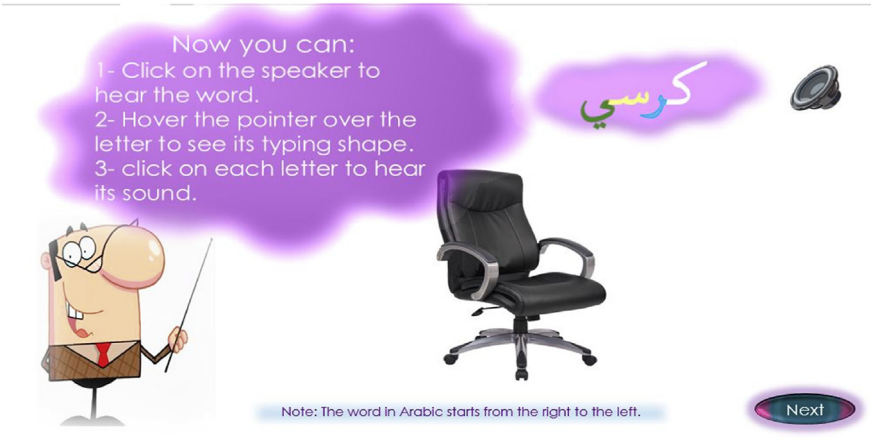


**Fig. 4.** CAVA screenshot of the target word كرسي *(chair) in GUI representation.*


**Fig. 5.** An overview of an environment (the office) using related words.

Each picture in this representation presented four target Arabic words, which are thematically related, and exist in a common environment. For example, the office picture showed a desk, chair, laptop and a bookcase (see Fig. 5.). The Arabic target word was written near the picture of the item to apply the spatial contiguity principle. This designing principle is strongly recommended with low-knowledge learners (Mayer, Steinhoff, Bower, & Mars, 1995) to help minimize extraneous cognitive load.

The learner was required to click on a target word. The program then zoomed in to enlarge the size of the selected object to fill a new screen to enable a separate study of the word (see Fig. 6.).

As shown above, the three multimedia representations were developed to deliver different aspects of the word's meaning; however, the representation was consistent in all the phases of CAVA; i.e., a consistent background appears for each condition, even in the slides that present default information.

*3.4.6. The two post-tests (phases four and five)*

As planned in the storyboard, the post-tests in CAVA were developed using a receptive and productive recall format. The same design elements employed when developing the study phases were used to create test slides. The immediate and delayed post-tests were identical and consisted of five questions. The first question tested the participant's ability to link the target meaning of the Arabic word (L1 translation, picture, or visuospatial representation) to its written form. The correct


**Fig. 6.** CAVA screenshot of the target word كرسي *(chair) in ZUI representation.*

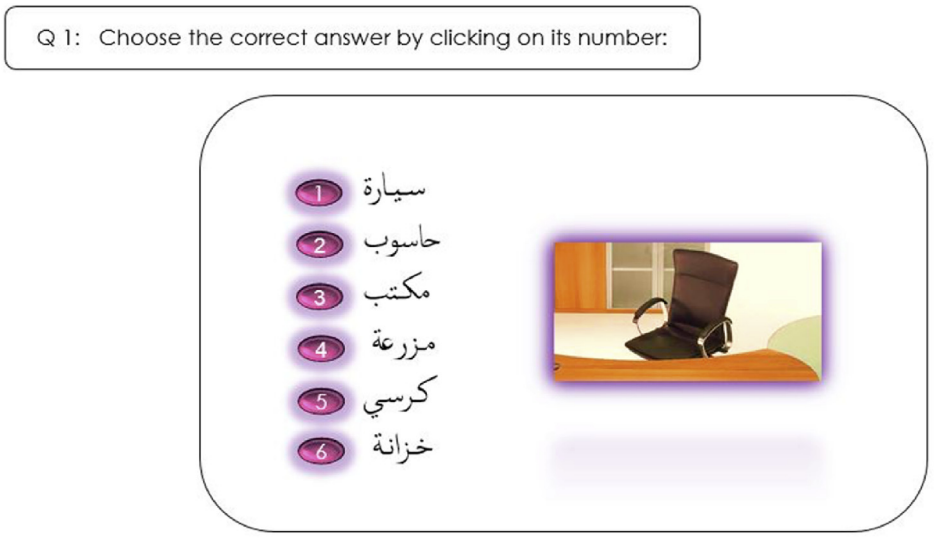


**Fig. 7.** The first type of questions in the post-test (meaning-written form) in the ZUI interface.

answer was presented along with five distractors to attract more learners with partial knowledge, who failed to identify a correct answer. The number of distractors was chosen to be five to avoid potential disadvantages of multiple-choice testing, such as guessing effects.

The learner then was required to identify the meaning of the written form of the Arabic target word as displayed on the screen (see Fig. 7.).

Once the learners responded to the question, they were automatically moved to the next question. In the second question format on the post-test, the same process was repeated, but this time it was used to test the learners' ability to associate meaning with the spoken form (see Fig. 8.).

The third, fourth and fifth types of questions on the post-test were association based. The third type was developed to assess the learners' ability to select an odd word from a list of related words. The fourth assesses the learners' ability to find the hidden target Arabic word among the distractors to complete a list of semantically related Arabic words. Finally, for the fifth type of question, the learners were required to write down an Arabic word to complete a list of four semantically related words (see Fig. 9 and 10, and 11).

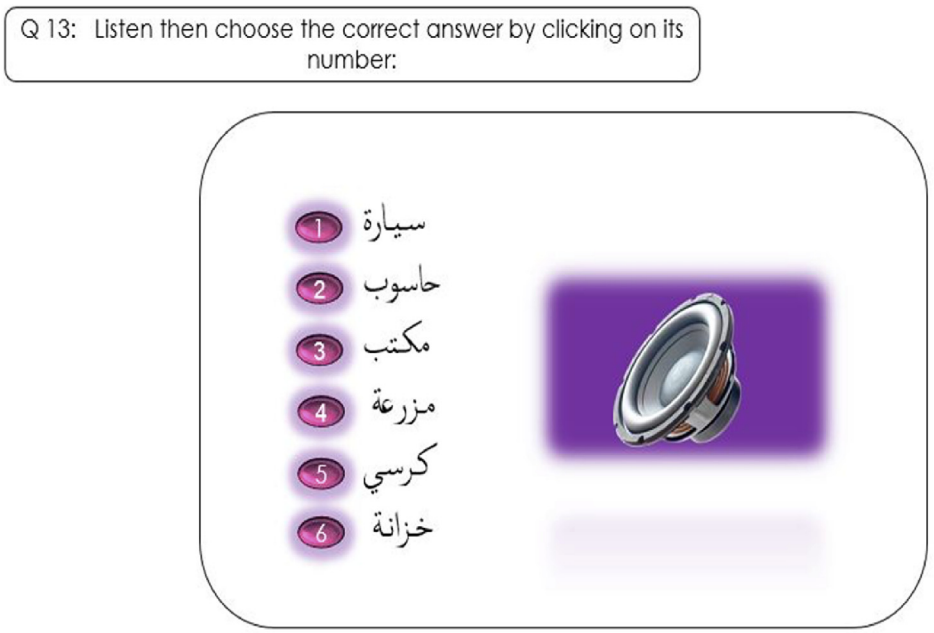


**Fig. 8.** The first type of questions in the post-test (meaning-spoken form).

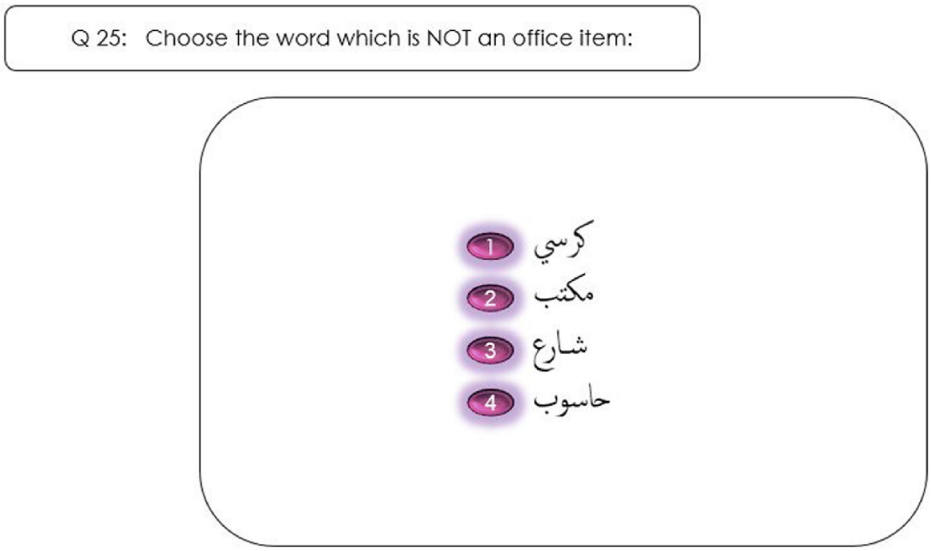


**Fig. 9.** The final type of questions in the post-test (word association - the odd word out).

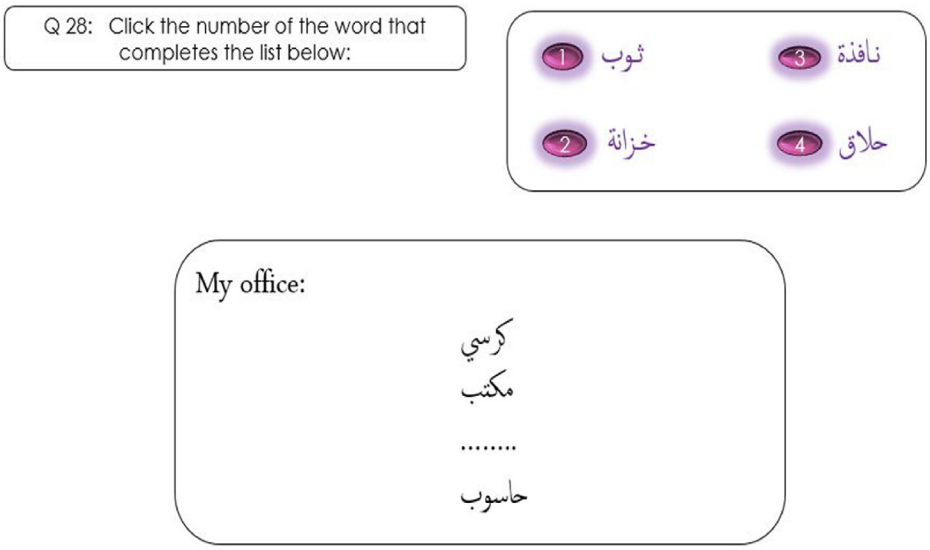


**Fig. 10.** The final type of questions in the post-test (word association - the hidden word).

The test proposed was also presented to a group of experts to ensure construct validity. Considering the experts' comments, the test was then modified to focus on measuring the word knowledge aspects it was anticipated the learners would have acquired.

### *3.5. Procedure*

All the sessions were conducted in a private study room in the University of York library. The researcher met each participant individually and introduced them to the experiment, describing the sessions clearly. The researcher's personal laptop (which hosts CAVA as the apparatus of the study and an Arabic/English keyboard) was used for the experiment. At the meeting, participants were exposed to CAVA and completed the lessons. Then, immediately after they completed the immediate post-test. A month later, the participants took the delayed CAVA post-test to measure the long-term effects of the three teaching methods.

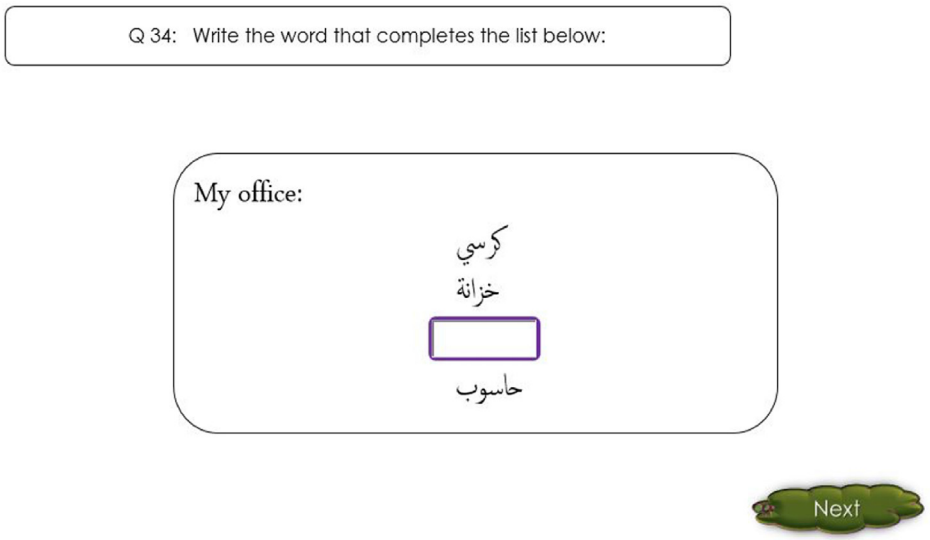


**Fig. 11.** The final type of questions in the post-test (word association - the missing word).

### 3.6. Scoring and analyses

The immediate and delayed post-test were identical and consisted of receptive recall, and recognition of multiple-choice questions, in addition to one productive recall question to which the participants responded by typing the correct word. All receptive recall, and recognition questions were scored dichotomously with 0 for an incorrect answer and 1 for a correct answer. When scoring the productive question, the participants received 0 points, 0.5 points or 1 point for each word. 0 points were received if the participant did not write the word down, or if there were more than two errors in the string of items. An error counted as any omission or substitution of a single letter in the target word. An example of a 0 point response would be when a participant writes **(فكيت)** for the target response **(مكتب)** (desk). If the participant writes the target word with one or two changes, they receive a score of 0.5 points. For example, the student response is **(مكيب)** for the target response **(مكتب)** *(desk).* If the response is identical to the target item, it is counted as correct and receives 1 point.

Descriptive and inferential analyses were run to answer our research question using Microsoft Excel and statistical package for social science program (SPSS) version 22. With multimedia representation as the within-subject factor, a one way repeated analysis of variance (ANOVA) and the Bonferroni post-hoc multiple pairwise comparison were carried out for the immediate and delayed tests respectively. The level for alpha was set to 0.05, to establish the statistical significance of all the tests. Since a repeated measures ANOVA was used, Mauchly's Test of Sphericity was conducted to evaluate the homogeneity of variances between conditions. In addition, the eta-squared ($\eta^2$) measure was used to calculate the effect size.

Below, the results from the data analysis for the immediate and delayed post-tests are presented and discussed in relation to the findings from previous studies and the research gap.

## 4. Results and discussion

The effectiveness of multimedia representations on L2 vocabulary acquisition were investigated by examining the main effects across the three interactive user interfaces (i.e., MDI, GUI, ZUI) on vocabulary acquisition in the immediate and delayed post-tests.

### 4.1. Immediate post-test

#### 4.1.1. Descriptive statistics

The learners, as shown in Table 4 below, achieved a mean score of 48.39 out of 78, corresponding to an average vocabulary acquisition rate of 62.03%.

The use of ZUI multimedia representation, comprising a zoomable interface and layout showing an overview of all the target words in the lesson, resulted in the highest mean score on the immediate post-test. Thus, it was found to support vocabulary acquisition the most in the immediate post-test. The second best multimedia representation was GUI followed by MDI.

**Table 4**
Immediate post-test results for the three multimedia representation types.

| Multimedia representations | Rank | Mean | Maximum score | Percentage | SD |
|---|---|---|---|---|---|
| MDI | 3 | 15.46 | 26 | 59.46% | 18.6 |
| GUI | 2 | 16.2 | 26 | 62.31% | 17.79 |
| ZUI | 1 | 16.73 | 26 | 64.35% | 17 |
| All multimedia representations | | 48.39 | 78 | 62.03% | 17.83 |

#### 4.1.2. Inferential statistics

The results revealed that the assumption of sphericity had not been violated. There was no statistically significant difference in the immediate post-test $\chi 2\,(2) = 0.175$, $p = 0.91$, indicating that ANOVA would be a valid testing procedure. ANOVA results showed that the effectiveness of the multimedia representation types on the acquisition of L2 vocabulary did not differ significantly in the immediate post-test. The Bonferroni post-hoc multiple pairwise comparison results were in line with the descriptive statistics. These results suggest a trend towards multimedia representation ZUI being more effective for vocabulary acquisition than the alternatives. The differences, however, were not statistically significant $F\ (2,\ 147) = 0.9$, $p = 0.3 > 0.05$ (see Table 5). The measure for effect size was slightly over the Cohen (1969) guidelines for a small effect ($\eta^2 < 0.01$) for the immediate post-test ($\eta^2 = 0.013$).

The analysis of the immediate post-test suggested that including visuospatial design elements in multimedia learning enhances learners' acquisition of L2 words as evidenced by the higher scores for the ZUI condition than for the GUI or MDI conditions. The multi-channel combination of presentations (verbal, visual and spatial) in ZUI, which were intended to communicate L2 words, resulted in enhanced learner performance in the immediate post-test. The spatial contiguity principle (Moreno & Mayer, 1999) was reflected in the short term ZUI advantage over GUI, while also enhancing the effectiveness of the learning material. The combination of multiple channels of presentations proved effective, in line with assertions about multimedia learning effectiveness literature; namely, the Cognitive Theory of Multimedia Learning (Mayer, 2005). ZUI was also followed by GUI, which outperformed MDI. Arguably the influence of ZUI and GUI proved greater as they delivered the meaning of L2 words in two modes (verbal and visual) in line with Dual-Code Theory (Paivio, 1991). These findings were evidenced in various studies which were designed to investigate multimedia instruction in CALL environments (e.g., Jones, 2009; Jones & Plass, 2002). ZUI and GUI displayed visual information about concrete objects, which is functionally similar to that which is visualized in the mind's eye. This seemed to facilitate the process of mental visualization, which is thought to be strongly analogous, although not identical, to visual perception (Farah, 1988; Finke, 1989; Jolicoeur & Kosslyn, 1985). Consequently, ZUI and GUI support better encoding and retrieval of target L2 words in the immediate post-test relative to MDI.

MDI proved less effective as a mode of facilitating learning, and for recalling a target L2 word when answering the CAVA immediate post-test questions. The low vocabulary recall scores in the MDI condition might imply that using only written words as a stimuli limits this particular teaching method. Our findings regarding the value of combined verbal and pictorial annotation in intentional vocabulary learning in multimedia CALL environments supports those of Dubois and Vial (2000), Fagehi (2013), Kim and Gilman (2008), and Rimrott's (2010) studies. These previous research findings highlighted the fact that less information can be delivered via the written form of the word. Thus, teaching L2 meaning through MDI proved inadequate to communicate multiple aspects of word knowledge when compared with ZUI and GUI.

### 4.2. Delayed post-test

To ensure the participants did not deliberately study the target vocabulary between the CAVA immediate and delayed post-tests, they were not told there would be an additional test.

#### 4.2.1. Descriptive statistics

The mean word retention on the delayed post-test of all multimedia representations was 28.79 out of a possible 7 8 points, that is 37%. This was lower than that for the immediate post-test, which was 62.03%.

The differences across the three multimedia representations for the delayed post-test were less pronounced than for the immediate post-test. Nevertheless, as shown in Table 6, GUI seemed to be the most effective multimedia representation. MDI and ZUI were less effective. The difference in mean vocabulary recall scores between the most effective multimedia representation in the delayed post-test (GUI) and the least effective one (ZUI) was 2.3%.

**Table 5**
Results of repeated measures ANOVA (Pairwise) comparing the scores for the three conditions, MDI, GUI, and ZUI on the CAVA immediate post-test.

| Multimedia representation (I) | Multimedia representation (J) | Mean difference (I − J) | Standard error | Sig. |
|---|---|---|---|---|
| ZUI | MDI | 4.88 | 3.5 | 0.17 |
| | GUI | 2 | 3.5 | 0.57 |
| GUI | MDI | 2.88 | 3.5 | 0.41 |

**Table 6**
Delayed post-test results for the three multimedia representation types.

| Multimedia representations | Rank | Mean | Maximum score | Percentage | SD |
|---|---|---|---|---|---|
| MDI | 2 | 9.5 | 26 | 36.57% | 11 |
| GUI | 1 | 9.9 | 26 | 38.46% | 13.8 |
| ZUI | 3 | 9.39 | 26 | 36.15% | 13.4 |
| All multimedia representations | | 28.79 | 78 | 37% | 12.7 |

*4.2.2. Inferential statistics*

The condition of sphericity was satisfied (i.e., the null hypotheses of sphericity were not rejected) for the delayed post-test $\chi 2\ (2) = 4.89$, $p = 0.086$. The results of ANOVA revealed that the effect of multimedia representation type on learners' scores was again notable in the delayed post-test as it had been for the immediate post-test. However, the effectiveness of the multimedia representations showed no significant difference. The results for the Bonferroni post-hoc multiple pairwise comparisons replicated the descriptive statistics. The results suggested a trend towards GUI being more effective for long term vocabulary acquisition than MDI and ZUI. The differences, however, were not found to be statistically significant $F\ (2, 147) = 0.4$, $p = 0.6 > 0.05$ (see Table 7). The measure for effect size for the delayed post-test ($\eta^2 = 0.006$) was under the Cohen (1969) guidelines for a small effect ($\eta^2 < 0.01$).

In the delayed post-test, the ZUI scores fell dramatically. Although, the use of spatial elements in vocabulary educational software is scarce, making it difficult to compare these results with the previous body of CALL literature, ZUI results concur with earlier findings suggesting spatial memory decays rapidly (Fagan et al., 2013; Hole, 1996). Conversely, GUI proved to be the most effective teaching method over the long-term, followed by MDI. Thus, the CAVA delayed post-test analyses highlighted the long-term effects of using an illustrated picture in a GUI design. Furthermore, over the long term, L2 vocabulary learners might be better able to acquire L2 concrete words via pictures, compared with visuospatial representations or L1 translation equivalents. In line with immediate post-test findings, the long-term impact of GUI compared with MDI corroborates evidence in Dual Coding research by Paivio (1991), suggesting that encoding information through verbal and visual channels results in better retention than employing a single information modality. In addition, the present findings might relate to Mayer's (2005) multimedia principle. The multimedia principal supports the use of GUI, observing that words and pictures together produced better results than pictures or words on their own as learners analyse verbal and visual communications differently. Certainly, the study participants tended to be better at processing both pictures and texts when acquiring word meaning in GUI. On the other hand, the absence of visual information in MDI seemed to weaken the encoding process, and thereby reduce the long-term retrieval of L2 words. This conclusion correlates with previous research focused on vocabulary CALL, presented in both incidental and intentional programs (see for example, Akbulut, 2007; Lin & Tseng, 2012; Yeh & Wang, 2003; Yoshii, 2006; Yoshii & Flaitz, 2002). As for the rapid decay of ZUI representation's impact, it appears that more studies are required to investigate the influence of zoomable visuospatial organization on the acquisition of L2 vocabulary since the current study represents the first empirical attempt to do so.

## 5. Conclusion

This study explored the effects of multimedia user interfaces on successful acquisition of Arabic vocabulary as a foreign language (AFL) via three different interactive interfaces; namely MDI, GUI and ZUI. There was an observable effect on both short- and long-term vocabulary learning when using the various multimedia representations to communicate L2 words. Nevertheless, the differences between the three teaching methods were insignificant which may have been related to the three main limitations of the current study. First, the number of target L2 words in CAVA was 24, and this might need to be increased to attain greater differences in learners' achievements. Second, CAVA seemed to provide rich default content in all three teaching methods across all three learning phases, including the practice exercises, which may have affected the results, minimizing differences across the multimedia representations. Third, presenting such a content in a short duration of intentional vocabulary teaching time might have also reduced the differences between the impact of MDI, GUI and ZUI on vocabulary learning.

On a practical level, however, the findings of the current study have implications for teaching L2 words in multimedia environments. First, the ZUI was designed to represent authentic situations associated with each specific target concrete word so the knowledge communicated through this particular e-learning module seemed to be richer and more effective on a short-term basis. When reviewing cutting-edge technology, it is apparent that it is almost possible to represent the majority

**Table 7**
Results for repeated measures ANOVA (Pairwise) comparing the scores for the three conditions, MDI, GUI, and ZUI on the delayed post-test.

| Multimedia representation (I) | Multimedia representation (J) | Mean difference (I − J) | Standard error | Sig. |
|---|---|---|---|---|
| ZUI | MDI | −0.42 | 2.57 | 0.87 |
| | GUI | −2.3 | 2.57 | 0.37 |
| GUI | MDI | 1.88 | 2.57 | 0.46 |

of situations in the target language use domains to allow the language learner to interact virtually within a simulated real life situation. Presenting such situations, in which target language learning can take place, might render the learner more able to see and acquire aspects of the target word's knowledge.

Therefore, more studies are probably required to explore the impact of the inclusion of visuospatial components in multimedia CALL designs. ZUI's high impact in the immediate post-test implies employment of spatial design elements is essential. However, given that the participants in the current study were exposed to the target words over a relatively short time frame, long lasting effects might be more observable in future studies when providing subsequent learning phases to increase retrieval efforts through recycling exercises. Second, ZUI learners were exposed to various aspects of scenes, by zooming in for a closer look, or accessing a holistic overview when zooming out. Through this visuospatial interaction, and by zooming in and out, a *spatial memory* of where things are arose. It can be implied that, through such an interaction, a mental image of the layout view was stored and a script for word retrieval encoded in the space in which the item is located. Therefore, the use and design of advanced educational technologies, such as the visuospatial zoomable interfaces, should always be compatible with learners' cognitive abilities, and render them responsive to individual differences between users. Thus, future CALL makers are advised to base their technology on Cognitive Theories of Learning such as Working Memory Models, which assume an essential role of working memory in second language vocabulary acquisition (e.g. Atkins & Baddeley, 1998; Gathercole & Baddeley, 1990; Martin & Ellis, 2012; Papagno & Vallar, 1992). Relating multimedia CALL design to teaching vocabulary to SLA and memory research has also been encouraged by some researchers (e.g. Handley, 2014; Pederson, 1987), but few convincing empirical attempts have been made to support this. Consequently, more research is needed to investigate the impact of individual differences in cognitive abilities on intentional learning via visuospatial representations.

## CRediT authorship contribution statement

**Saad Alzahrani:** Conceptualization, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **Leah Roberts:** Conceptualization, Methodology, Validation, Resources, Data curation, Writing - original draft, Supervision, Project administration.

## References


Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science, 35*(6), 499–517.

Atkins, P. W., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied PsychoLinguistics, 19*(4), 537–552.

Boers, F., Warren, P., Grimshaw, G., & Siyanova-Chanturia, A. (2017). On the benefits of multimodal annotations for vocabulary uptake from reading. *Computer Assisted Language Learning, 30*(7), 709–725.

Çakmak, F., & Erçetin, G. (2018). Effects of gloss type on text recall and incidental vocabulary learning in mobile-assisted L2 listening. *ReCALL, 30*(1), 24–47.

Chen, I.-J. (2016). Hypertext glosses for foreign language reading comprehension and vocabulary acquisition: Effects of assessment methods. *Computer Assisted Language Learning, 29*(2), 413–426.

Clark, R. C., & Mayer, R. E. (2011). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. San Francisco: Pfeiffer.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language, Learning and Technology, 11*(3), 38–63.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Dubois, M., & Vial, I. (2000). Multimedia design: The effects of relating multimodal information. *Journal of Computer Assisted Learning, 16*(2), 157–165.

Esit, Ö. (2011). Your verbal zone: An intelligent computer-assisted language learning program in support of Turkish learners' vocabulary learning. *Computer Assisted Language Learning, 24*(3), 211–232.

Fagan, W. F., Lewis, M. A., Auger-Méthé, M., Avgar, T., Benhamou, S., Breed, G., et al. (2013). Spatial memory and animal movement. *Ecology Letters, 16*(10), 1316–1329.

Fagehi, A. Y. (2013). *Strategies for defining and understanding critical technology integration terms*. Doctoral dissertation, University of Kansas.

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95*(3), 307.

Finke, R. A. (1989). *Principles of mental imagery*. The MIT Press.

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*(4), 439–454.

Godwin-Jones, R. (2010). Emerging technologies from memory palaces to spacing algorithms: Approaches to second language vocabulary learning. *Language, Learning and Technology, 14*(2), 4–11.

Gullberg, M., Roberts, L., & Dimroth, C. (2012). What word-level knowledge can adult learners acquire after minimal exposure to a new language? *International Review of Applied Linguistics (IRAL), 50*, 239–276.

Handley, Z. (2014). Vocabulary CALL for Young ESL/EFL Learners: A systematic review of the research evidence. In *Engaging Language Learners through Technology Integration: Theory, Applications, and Outcomes* (pp. 299–325). New Jersey, US: IGI Global.

Hirschel, R., & Fritz, E. (2013). Learning vocabulary: CALL program versus vocabulary notebook. *System, 41*(3), 639–653.

Hole, G. J. (1996). Decay and interference effects in visuospatial short-term memory. *Perception, 25*(1), 53–64.

Jolicoeur, P., & Kosslyn, S. M. (1985). Is time to scan visual images due to demand characteristics? *Memory & Cognition, 13*(4), 320–332.

Jones, L. C. (2009). Supporting student differences in listening comprehension and vocabulary learning with multimedia annotations. *Calico Journal, 26*(2), 267–289.

Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal, 86*(4), 546–561.

Kaplan-Rakowski, R., & Loranc-Paszylk, B. (2019). The impact of verbal and nonverbal auditory resources on explicit foreign language vocabulary learning. *System, 85*, 1–10.

Khezrlou, S., Ellis, R., & Sadeghi, K. (2017). Effects of computer-assisted glosses on EFL learners' vocabulary acquisition and reading comprehension in three learning conditions. *System, 65*, 104–116.

Kim, S. H. (2019). The effects of using images and video clips for vocabulary learning: Single words vs. phrasal verbs. *STEM Journal, 20*(2), 19–42.

Kim, D., & Gilman, D. A. (2008). Effects of text, audio, and graphic aids in multimedia instruction for vocabulary learning. *Educational Technology & Society, 11*(3), 114–126.

Klein, W., & Dimroth, C. (2009). Untutored second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 503–522). New York: Academic Press.
Laufer, B. (2006). Comparing focus on form and focus on forms in second-language vocabulary learning. *Canadian Modern Language Review, 63*(1), 149–166.
Lin, C. C., & Tseng, Y. F. (2012). Videos and animations for vocabulary learning: A study on difficult words. *Turkish Online Journal of Educational Technology-TOJET, 11*(4), 346–355.
Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*(3), 379–413.
Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. New York: Cambridge University Press.
Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research & Development, 43*(1), 31–41.
Min, H. T. (2008). EFL vocabulary acquisition and retention: Reading plus vocabulary enhancement activities and narrow reading. *Language Learning, 58*(1), 73–115.
Mohsen, M. A. (2016). Effects of help options in a multimedia listening environment on L2 vocabulary acquisition. *Computer Assisted Language Learning, 29*(7), 1220–1237.
Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358–368.
Nation, I. S. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.
Nation, I. S. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press.
Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*(3), 255–287.
Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. *The Quarterly Journal of Experimental Psychology Section A, 44*(1), 47–67.
Pederson, K. (1987). Research on CALL. In W. F. Smith (Ed.), *Modern media in foreign language education: Theory and implementation* (pp. 99–131). Lincolnwood, IL: National Textbook.
Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 90*(1), 25–36.
Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*(2), 221–243.
Ramezanali, N., & Faez, F. (2019). Vocabulary learning and retention through multimedia glossing. *Language, Learning and Technology, 23*(2), 105–124.
Rimrott, A. (2010). *Computer-assisted vocabulary learning: Multimedia annotations, word concreteness and individualized instruction*. Burnaby, BC, Canada: Doctoral dissertation, Simon Fraser University.
Rouhi, A., & Mohebbi, H. (2013). Glosses, spatial intelligence, and L2 vocabulary learning in multimedia context. *3L: The Southeast Asian Journal of Language Studies, 19*(2), 75–87.
Rusanganwa, J. A. (2015). Developing a multimedia instrument for technical vocabulary learning: A case of EFL undergraduate physics education. *Computer Assisted Language Learning, 28*(2), 97–111.
Sato, T., Matsunuma, M., & Suzuki, A. (2013). Enhancement of automatization through vocabulary learning using CALL: Can prompt language processing lead to better comprehension in L2 reading? *ReCALL, 25*(1), 143–158.
Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan.
Sun, Y., & Dong, Q. (2004). An experiment on supporting children's English vocabulary learning in multimedia context. *Computer Assisted Language Learning, 17*(2), 131–147.
Türk, E., & Erçetin, G. (2014). Effects of interactive versus simultaneous display of multimedia glosses on L2 reading comprehension and incidental vocabulary learning. *Computer Assisted Language Learning, 27*(1), 1–25.
Ware, C. (2004). *Information visualization: Perception for design* (2nd ed.). San Francisco, CA: Morgan Kaufman.
Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language, Learning and Technology, 13*, 48–67.
Yeh, Y., & Wang, C. W. (2003). Effects of multimedia vocabulary annotations and learning styles on vocabulary learning. *Calico Journal, 21*, 131–144.
Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language, Learning and Technology, 10*(3), 85–101.
Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of picture and annotation types. *CALICO Journal, 20*(1), 33–58.