<AH>The effect of textual and textual-pictorial glosses on incidental vocabulary learning in mobile-assisted listening<AH>

<AU>Arman Shahipanah<AU>

<AD>University of Bojnord, Iran ([info@armanshahipanah.ir](mailto:info@armanshahipanah.ir))<AD>

<AU>Gholam Hassan Khajavy<AU>

<AD>University of Bojnord, Iran ([hkhajavy@ub.ac.ir](mailto:hkhajavy@ub.ac.ir))<AD>

<AU>Majid Elahi Shirvan<AU>

<AD>University of Bojnord, Iran ([m.elahi@ub.ac.ir](mailto:m.elahi@ub.ac.ir))<AD>

<ABS><ABH>Abstract<ABH>

This study investigated how multimedia glossing affects incidental vocabulary learning from a listening task on mobile devices. A total of 118 English language learners were asked to listen to a story with 25 glossed target words on their mobile phones. In order to examine the effects of different types of glossing, participants were divided into four groups with access to four glosses during their listening: L1 textual, L2 textual, L1 textual and pictorial, and L2 textual and pictorial. Two vocabulary tests (i.e. definition-supply test and meaning-recognition test) were administrated immediately after treatment and two weeks later to measure vocabulary gain for target words. The results indicated that participants who had access to L1 textual and pictorial glosses had significantly higher vocabulary gains than other conditions, especially in meaning-recall word knowledge. Finally, a detailed discussion of the findings was provided to explain the results based on the theoretical framework of the study.

**Keywords:** multimedia gloss; incidental L2 vocabulary learning; L1 and L2 textual gloss; mobile-assisted language learning (MALL)<ABS>

<A>1. Introduction<A>

Mobile learning has provided many opportunities for language learning (McQuiggan, Kosturko, McQuiggan & Sabourin, 2015) due to its accessibility outside the classroom. One area that has attracted researchers’ attention is using mobiles for learning vocabulary (see Lin & Lin, 2019). Limited time in classrooms urges teachers to find alternative approaches to teaching vocabulary. An excellent strategy to compensate for the problem is to empower learners with autonomous abilities to learn new vocabulary incidentally outside the classroom by extensive listening or reading (Rashtchi & Pourmand, 2014; van Zeeland & Schmitt, 2013; Waring & Nation, 2004). Incidental vocabulary learning is defined as “learning of vocabulary as the by-product of any activity not explicitly geared to vocabulary learning” (Hulstijn, 2001: 270). However, learning new words outside the classroom has some drawbacks (Huckin & Coady, 1999). For instance, the lack of understandable contextual information could make correct inferencing difficult for learners. Using glosses was proposed as a remedy to aid the correctness of inferencing, avoiding misleading guessing, and consequently effective incidental vocabulary learning (Hu, Vongpumivitch, Chang & Liou, 2014; Laufer, 2000). Using gloss as a vocabulary enhancement technique can be regarded as an opportunity to be implemented in mobiles.

Many studies have investigated the effects of using glosses, mostly in computer-assisted language learning (CALL) context (e.g. Boers, Warren, He & Deconinck, 2017; Mohsen & Balakumar, 2011; Ramezanali & Faez, 2019). However, most of these studies have focused on the effects of glosses on L2 reading comprehension, while incidental vocabulary learning through L2 listening has received little attention (Çakmak & Erçetin, 2018; Mohsen & Balakumar, 2011), though some studies have addressed incidental vocabulary learning through viewing (Çekiç, 2024; Hsieh, 2020; Montero Perez, Peters & Desmet, 2018). In addition, the effectiveness of the language of glossing (L1 or L2) has been investigated by several studies in reading comprehension (Ramezanali, Uchihara & Faez, 2021; Yanagisawa, Webb & Uchihara, 2020). A recent meta-analysis by Yanagisawa *et al*. (2020) found that L1 glosses led to more vocabulary learning than L2 glosses in reading comprehension. Furthermore, previous studies have shown that when a text gloss is combined with a picture, its effectiveness increases for incidental vocabulary acquisition in reading comprehension (Ramezanali *et al*., 2021; Yanagisawa *et al*., 2020; Yeh & Wang, 2004; Yoshii & Flaitz, 2002). However, it is not clear whether the same patterns emerge when there is a listening task on mobile phones. Previous studies on listening comprehension focused on L1 glosses only (e.g. Çakmak & Erçetin, 2018; Jones & Plass, 2002; van Zeeland & Schmitt, 2013; Vidal, 2011), and further research is needed to compare the effectiveness of L1 and L2 glosses (both alone and combined with pictures) in incidental vocabulary learning through listening (see Mohsen & Balakumar, 2011). Therefore, this study seeks to address these gaps by examining the effectiveness on incidental vocabulary learning of using L1 and L2 glosses with and without pictures while listening to a story on mobile phones.

<A>2. Literature review<A>

<B>2.1 Theoretical foundations<B>

The present paper begins with a review of the theoretical foundations of glossing. One of the theoretical bases for implementing multimedia glosses for vocabulary learning is the dual coding theory (DCT) proposed by Paivio (1990). The DCT focuses on the cognitive dimensions of processing new information. According to the DCT, there are two interrelated systems in information processing: “one specialized for the representation and processing of information concerning nonverbal objects and events, the other specialized for dealing with language” (Paivio, 1990: 53).

Paivio (1990) claims that new information presented to learners leaves a trace in each system and is coded differently. For example, abstract and verbal information leaves verbal traces, while concrete concepts and pictorial information leave both visual and verbal traces, which are coded differently in two systems (<AQ1>Boers *et al*., 2017). Paivio believed that information that is coded dually or has compound traces (received both verbally and nonverbally) leads to richer interactions between two systems and, consequently, more effective learning.

Building upon the foundations of the DCT, one of the most comprehensive theories of multimedia learning is Mayer’s (2014) cognitive theory of multimedia learning (CTML). The CTML describes the underlying processes of learning in a multimedia environment. It consists of three main assumptions: the dual channels assumption, the limited capacity assumption, and the active processing assumption. The assumption of dual channels is based on Paivio’s DCT (1990) just described. The assumption of limited capacity states that verbal and visual channels have their limited capacity, meaning that learners can only process a finite amount of information at a given time through each channel. Finally, the assumption of active processing proposes that learning involves actively selecting relevant visual or verbal information, filtering them through each channel, organizing information in separate systems, and integrating verbal and visual representations with each other and with prior knowledge (Mayer, 2014). Like Paivio (1990), Mayer noted that “people learn more deeply from words and pictures than from words alone” (Mayer, 2009: 47), and thus it has been argued that by using words and appropriate pictures together, more connections between mental representations and meanings are created (Çakmak & Erçetin, 2018).

Finally, Kroll and Stewart (1994) proposed a revised hierarchical model (RHM) to explain language processing among bilinguals whose L1 is the dominant language. The RHM states that at the early stages of bilingualism, since learners’ L1 is dominant and has more substantial access to concepts, L2 words require L1 translation to access concepts. This process happens until the L2 proficiency level increases to the point where learners can directly associate concepts with L2 words without mediating via the L1 (Kroll, Van Hell, Tokowicz & Green, 2010). Based on this model, it is expected that L1 glosses may lead to better vocabulary learning when L2 proficiency is low (Hu *et al*., 2014; Yoshii, 2006).

<B>2.2 Multimedia gloss in vocabulary learning<B>

Traditional glosses were mainly used on the margins of pages to explain complex terms of texts or provide examples in relation to the main content of a text. Nowadays, an electronic gloss (e-gloss) consists of a hypertext embedded within texts that does not take up any space on the page or distract readers and listeners. Advances in technologies have allowed additional forms of glosses (e.g. pictures, videos, and audio), which can be presented through different modalities (visual, textual, and auditory, respectively). The term multimedia gloss refers to textual glosses that are accompanied by sounds or pictures, whether static or animated (Elekaei, Faramarzi & Koosha, 2015). In a recent review of the role of multimedia glosses in L2 skills development, Zhang and Zou (2022) reported on 41 studies between 2009 and 2020, among which 24 studies (58.5%) were devoted to vocabulary knowledge, with a smaller proportion focusing on listening, reading, and grammar knowledge. It should be noted that very few studies focused on incidental vocabulary learning by listening comprehension on a mobile device (see Çakmak & Erçetin, 2018, for an exception). Moreover, of the 24 studies devoted to lexical learning, most were conducted on computers: only six studies used mobile devices. This shift towards e-glosses and multimedia resources reflects a broader trend in language education, which increasingly values the integration of technology in learning environments.

Glosses can be used for different purposes. Glossing complex texts enables learners to read and listen above their current level (Widdowson, 1978) and reduces inaccurate guessing of meanings for low-ability readers when they encounter unknown words (Khezrlou, Ellis & Sadeghi, 2017). Moreover, glossing can attract learners’ attention to difficult words that might be missed (<AQ2>Boers *et al*., 2017). Due to the rapid developments in computer technologies, different multimedia forms of glossing can be used. Several studies have investigated the effectiveness of implementing different types of glosses on L2 vocabulary learning through reading (Akbulut, 2007; Hu *et al*., 2014; Lin & Lin, 2019; Yeh & Wang, 2004; Yoshii, 2006; Yoshii & Flaitz, 2003), listening (Çakmak & Erçetin, 2018; Jones & Plass, 2002), and viewing (Çekiç, 2024; Hsieh, 2020; Montero Perez *et al*., 2018). Moreover, several review studies have indicated that multimedia glosses can be effective in improving vocabulary learning, listening comprehension, reading comprehension, and grammar knowledge (see Ramezanali *et al*., 2021; Yanagisawa *et al*., 2020; Zhang & Zou, 2022).

<B>2.3 L1 and L2 glosses in vocabulary learning<B>

As noted earlier, a primary concern in multimedia glossing is the choice of language (Mohsen & Balakumar, 2011; Ramezanali *et al*., 2021; Yanagisawa *et al*., 2020). Previous studies that compared the effectiveness of L1 and L2 glosses showed mixed results. Some studies found an advantage for L1 over L2 glosses (e.g. Arpaci, 2016; Choi, 2016; Kim, Lee & Lee, 2024; Xu, 2010), while others found no significant difference (Jacobs, Dufon & Hong, 1994; Yoshii, 2006). Yoshii (2006) found that the retention rate was better for L2 plus picture glosses than for L1 plus picture glosses in the definition-supply test (DST), a measure where participants are asked to provide the definitions of words without any cues, thus assessing their ability to recall vocabulary actively. However, the results of Yanagisawa *et al*.’s (2020) meta-analysis showed that L1 glosses were generally more efficient than L2 glosses in vocabulary learning. They proposed that this result can be explained by learners’ greater familiarity with L1 words, meaning that connections between L1 words and target words can be made more easily than between L2 words and target words. Another meta-analysis by Kim *et al*. (2024) also reported the superiority of L1 glosses over L2 glosses for vocabulary learning. This advantage was greater in immediate post-test conditions than in delayed post-tests.

Previous research has also indicated that proficiency level might affect the effectiveness of L1 and L2 glosses (Hu *et al*., 2014; Ko, 2017, Yoshii, 2006). For instance, Hu *et al*. (2014) investigated the effects of online L1 and L2 glosses on incidental vocabulary learning in learners with lower and higher L2 proficiency using the RHM (Kroll & Stewart, 1994) as their theoretical foundation. They found that “L1 conceptual links are stronger than L2 conceptual links in the early stages of L2 acquisition (i.e. lower proficiency level)” (Hu *et al*., 2014: 95). They tested this hypothesis by conducting experiments that measured the strength of conceptual links through tasks that assessed the recall and recognition of L2 vocabulary, comparing the performance of participants at different proficiency levels. In their meta-analysis, in contrast, Yanagisawa *et al*. (2020) found no significant interaction between proficiency and glossing languages, explaining that regardless of proficiency level of the participants, L1 glosses were more effective than L2 glosses. This issue was also investigated by Kim *et al*. (2024), who found L1 glossing to be more effective for beginners than for higher-proficiency language learners. However, as mentioned by Yanagisawa *et al*. (2020), because very few studies have reported learner proficiency, the results of these studies cannot be considered sufficient support for this claim.

<B>2.4 Gloss mode and type in vocabulary learning<B>

As mentioned, glosses can be presented in different modes: textual, pictorial, auditory, and video (Ramezanali *et al*., 2021; Yanagisawa *et al*., 2020). Previous studies have examined single gloss mode (i.e. only one type of gloss), dual gloss mode (i.e. a combination of two glosses), multimodal glossing (i.e. three different glosses). Results of the meta-analysis by Ramezanali *et al*. (2021) indicated that dual glossing had more effect on vocabulary learning than single glossing, but no difference was found between dual and triple glossing modes. They further found that the addition of another mode to a single L2 textual gloss is more beneficial for vocabulary learning than the addition of another mode to a single L1 textual gloss. Previous studies have also indicated the superiority of pictorial-textual glosses over textual-only or pictorial-only glosses in vocabulary learning (Çakmak & Erçetin, 2018; Jones & Plass, 2002; Yanguas, 2009; Yeh & Wang, 2004; Yoshii & Flaitz, 2003). Moreover, glosses can be presented in two different types: noninteractive and interactive glosses (see Yanagisawa *et al*., 2020). The former refers to glosses put in a fixed place (e.g. in the margin of a text, next to the word, or at the end of the text). Interactive glosses require an action on the part of the learners, such as hyperlinked glosses, which need clicking on the word, or multiple-choice glosses, where learners have to choose the correct meaning from different options (Yanagisawa *et al*., 2020).

<B>2.5 Using gloss in a listening context<B>

In this section, we delve into the application of glosses in a listening context, examining how different types of glosses (i.e. textual, pictorial, and a combination of both) affect vocabulary acquisition and text recall, starting with reviewing a study that has utilized mobile devices for glossing (e.g. Çakmak & Erçetin, 2018; Jones & Plass, 2002). In one study, Çakmak and Erçetin (2018) examined how different gloss types affected text recall and vocabulary learning while participants listened to a story on their mobile phones. Participants were assigned to one of three different glossing conditions or a control group without glossing treatment. The gloss conditions consisted of two single-channel groups (textual-only and pictorial-only) and one dual-channel group (pictorial plus textual glosses). The findings indicated a positive effect of glossing on incidental vocabulary learning, with no significant differences across glossing conditions. The authors speculated that glossing new words both visually and verbally might increase the cognitive load on low-ability learners, preventing an expected advantage for the dual-channel group.

In a similar study, Jones and Plass (2002) investigated the outcomes of using multimedia glossing on listening comprehension and vocabulary learning in a CALL environment and obtained differential results. Participants were randomly assigned into four groups: (a) listening without any glosses, (b) listening with only textual glosses, (c) listening with only pictorial glosses, and (d) listening with both textual and pictorial glosses. The findings in both immediate and delayed post-tests indicated that participants in gloss conditions had significantly better performance in aural recall and immediate and <AQ3>recall vocabulary tests compared to the no-gloss group. Participants who had access to both textual and pictorial glosses while listening had significantly higher scores on the vocabulary test than either textual-only or pictorial-only groups in the immediate post-test, with no significant difference between the single-mode conditions. At delayed post-test, the dual-mode group outperformed the textual gloss group, but not the pictorial gloss group. These findings support Mayer’s (2014) CTML for listening comprehension and vocabulary learning.

To conclude this review of recent glossing research, it is essential to recognize the limitations and potential areas for further investigation that have emerged from the existing body of work. Both Çakmak and Erçetin (2018) and Jones and Plass (2002) used only L1 glosses. They did not compare L1 textual glosses with L2 textual glosses and how their combination with pictures might affect incidental vocabulary learning. As mentioned by Mohsen and Balakumar (2011) in their review of glosses in CALL research, “in an L2 listening context, the comparison between L1 and L2 annotation types has not yet been explored. All the studies reviewed used native language annotations … and this issue should be considered in future research studies” (p. 151). The present study thus sought to address this issue by comparing L1 and L2 textual glosses. It also builds on recent research on dual-mode glossing to investigate how L1 textual-pictorial glosses might differ from L2 textual-pictorial glosses in incidental vocabulary learning through listening comprehension.

<A>3. The present study<A>

The first aim of this study is to examine the possible differences of using glosses in participants’ L1 (Persian) and L2 (English) to see which condition has better outcomes for incidental vocabulary learning while listening to a story. The second goal was to investigate the role of adding pictures to L1 and L2 textual glosses in the immediate post-test and in the delayed post-test. Taken together, we had four different conditions: L1 text-only (Group A), L1 textual-pictorial (Group B), L2 text-only (Group C), and L2 textual-pictorial (Group D). We adopted a pre-/post-/delayed post-test design among a sample of teaching English as a foreign language (TEFL) students at a university in Iran. We sought to answer two research questions (RQs):

<LN>

How does gloss language (L1 and L2) affect incidental vocabulary learning?

How does adding pictures to L1 and L2 textual glosses affect incidental vocabulary learning?<LN>

<A>4. Research method<A>

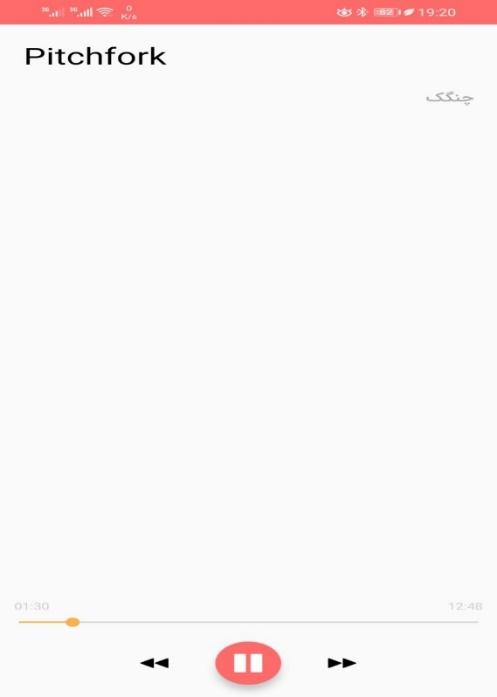
<B>4.1 Participants<B>

Participants in this study were 118 undergraduate students (55% female) of TEFL from four intact classes at a university in north-east Iran. Their ages ranged from 18 to 22 years old (*M* = 19.5, *SD* = 1.14). Before this study, all participants took the second version of the Oxford Quick Placement Test of Proficiency. Participants’ proficiency scores were between 16 and 35, including elementary (A2, 46%), pre-intermediate (B1, 38%), and upper-intermediate (B2, 16%) levels of proficiency based on the Common European Framework of Reference for Languages. Participants did not have any prior listening activity using glosses in their courses. The experiment and data gathering took place in the classrooms during regular class hours, and the environment and setting for all participants were similar.

<B>4.2 Instrumentation<B>

<C>4.2.1 Designing and developing the mobile application<C>

We developed a listening application using Java and Kotlin for Android mobile phones. Special attention was given to the user interface design to ensure that the glosses were properly displayed, making efficient use of the diverse screen sizes of the participants’ phones. A separate application was developed to display each of the four types of glosses to be compared in the study. The target words with their corresponding glosses appeared on screen as they were narrated in the story. The time of exposure of each gloss was set to four seconds. Participants could adjust the listening file by pausing and going back and forth, while simultaneously being able to see the related gloss on the screen. The pictures used in textual-pictorial glosses were taken from freely available online materials. To ensure the pictures were an appropriate representation of the target words and unambiguous, we consulted two experts in teaching English and two native English speakers for their opinions. Figure 1 shows screenshots for the four glossing conditions. In the L1 text-only condition (Figure 1A), when participants listen to the story, and the relevant target word is mentioned, they could see the target word with its Persian translation on the screen. In the L1 textual-pictorial condition (Figure 1B), in addition to the Persian translation of the target word, participants could see the relevant picture on the screen. In the L2 text-only condition (Figure 1C), participants could see the L2 definitions of the target words on the screen. Finally, in the L2 textual-pictorial condition (Figure 1D), participants could see the target word with its L2 definition and the picture, which was also used for the L1 textual-pictorial condition. In order to accommodate the various screen sizes of participants’ mobile phones, certain written glosses were adjusted slightly in terms of font size and position to ensure optimal fit on the screen.

**** 

**A B**

**** 

**C D**

**<FC>Figure 1.** Screenshot from the L1 text-only condition (A), L1 textual-pictorial condition (B), L2 text-only condition (C), and L2 textual-pictorial condition (D)<FC>

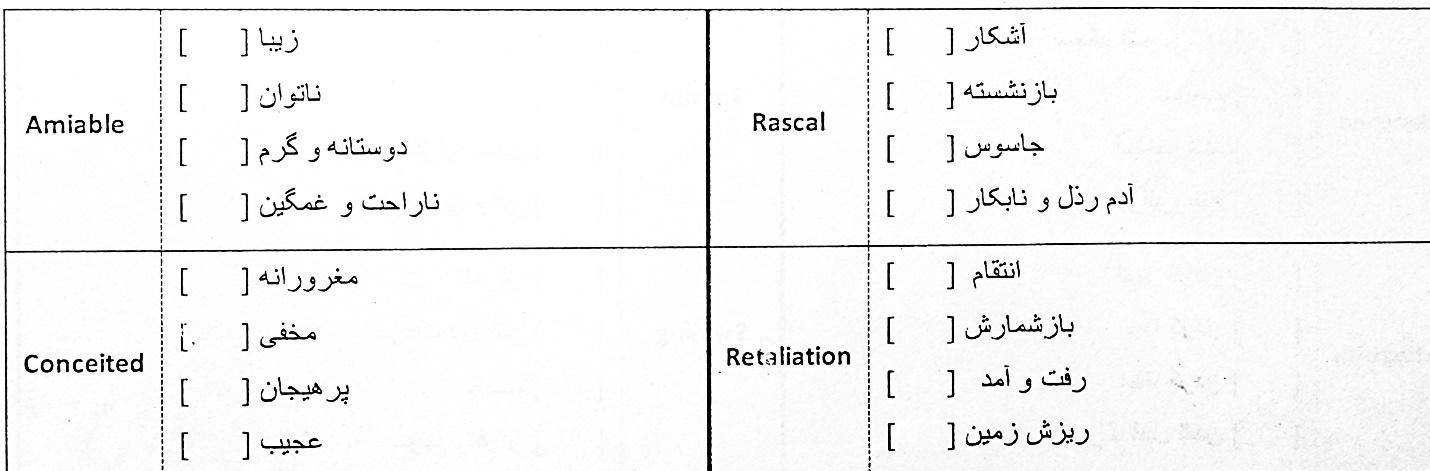
<C>4.2.2 Listening material<C>

Participants listened to a story from stage five of the Oxford Bookworm Series. The story’s title was “Wuthering Heights” and the total length of the audio file was 12:48 minutes, including 1,927 words. The story consisted of narrational descriptions of events and dialogues between fictional characters, with some minor sound effects. The readability index from the Flesch Reading Ease Formula (<AQ4>Flesch, 1948) found to be 87.9 for the story’s text. To establish the vocabulary profile of the listening text, we used BNC/COCA word family lists (Nation, 2017). Respectively, 74.56% and 12.08% of word types belonged to K-1 and K-2 frequency levels, 8.54% belonged to K-3 to K-5 frequency levels, and only 4.82% belonged to higher frequency levels. With regard to the target vocabulary items, 57.7% of the words belonged to levels K2-5, 26.93% to K6-8, and 15.39% to K10-12. The listening task had an average sentence length of 12.69 words. The average number of characters per word was 4.4. The speech rate was 109 words per minute, while the articulation rate was 137 words per minute. The speech rate refers to the average number of words spoken per minute during the entire audio file, including pauses and other non-speaking intervals. On the other hand, the articulation rate measures the number of words spoken per minute, but only during the periods when speech is actually occurring. The articulation rate is typically higher than the speech rate because it only considers the time when words are being spoken and not the silent intervals. Additionally, phrase pauses were 0.71 seconds on average.

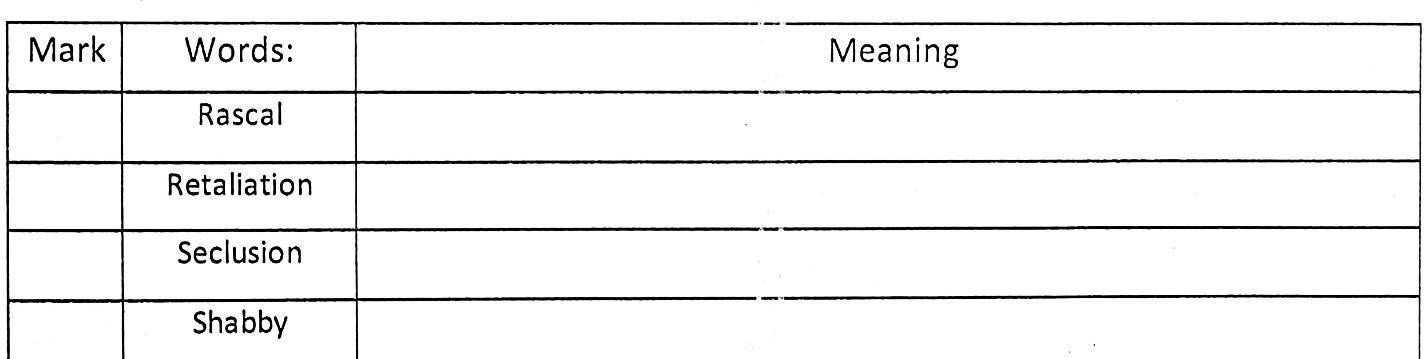
<C>4.2.3 Vocabulary tests<C>

Selecting target lexical items began with identifying the most appropriate words for glossing in the story. Forty-eight words suitable for glossing, both textually and visually, and six distractors were selected for the pre-test. The pre-test was a DST that required participants to write the meaning of the words either in Farsi or English. To choose target words, 25 words unknown to over 90% of the participants were selected for glossing. The target lexical items consisted of five verbs, eight nouns, and 12 adjectives (see online supplementary materials).

After choosing 25 lexical items, two unannounced vocabulary tests were used immediately after treatment (i.e. immediate post-test) and two weeks after treatment (i.e. delayed post-test). These post-tests consisted of a DST and a meaning-recognition test (MRT). The DST was designed to evaluate meaning-recall vocabulary knowledge.The items consisted of all 25 target words, and the participants had to provide the meaning either in their L1 (Farsi) or L2 (English). The MRT was designed to measure participants’ meaning-recognition vocabulary knowledge. Each item in the MRT consisted of four multiple-choice alternatives. Moreover, we used two versions of MRT, one with alternatives in Persian for L1 gloss groups and one with alternatives in English for L2 gloss groups, to ensure that the test did not favor a particular group. Participants had to select the most appropriate and closest definition to the given word (sample items have been provided in Figure 2; see supplementary materials for the complete list of items for tests).

****

**A**



**B**

**<FC>Figure 2.** Example items of meaning-recognition (A) and definition-supply (B) tests<FC>

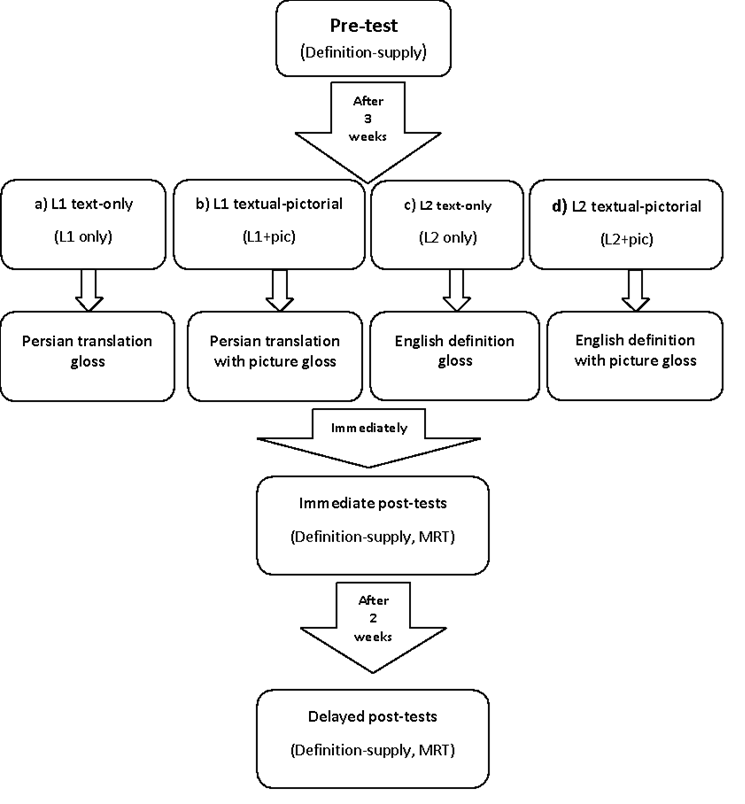
<B>4.3 Procedure<B>

As we used non-random sampling in this study, a quasi-experimental design was used in four intact classes. All data in pre-test, immediate post-test, and delayed post-test were gathered during regular classroom hours. Participation in the study was voluntary, with students awarded extra credit for participation; all students took part in the study.

First, students were asked to write the meaning of the 48 potential target words and six distractors selected for the pre-test. The treatment concerning the same 25 target words was conducted in a single session. Participants were randomly assigned to four different groups: (a) L1 text-only (*n* = 29), (b) L1 textual-pictorial (*n* = 29), (c) L2 text-only (*n* = 30), and (d) L2 textual-pictorial (*n* = 30). Then, we explained the task (title of the story and its duration) and how to download and install the applications on their mobile devices. Participants were already informed to bring their earphones to the class. Moreover, they were informed that they could see definitions of some unknown words on their mobile screens and could use pause or back-and-forth buttons. Furthermore, we did not tell them they would take vocabulary tests.

Immediately after the treatment, the immediate post-tests took place in the same session. The test began with DST to reduce the possibility of learning from MRT. Each correct answer in the DST scored one point, and no point was given for blank, wrong, or incomplete answers. For the MRT, correct responses received one point, and blank or inaccurate responses, none. The maximum possible score for each test was therefore 25.

Finally, delayed post-tests took place two weeks after immediate post-tests; participants took the same vocabulary tests with a changed order of items. The setting, time, and situations for the four groups were similar to the immediate post-tests. The design of the study is summarized in Figure 3.

****

**<FC>Figure 3.** Research design of the present study<FC>

<B>4.4 Data analysis<B>

We used SPSS Version 25.0 for data analysis. There were two dependent variables (DST and MRT) that can be explained by one between-group variable (treatment) with four levels (A-D) and one within-group variable (time). There is also a pre-test score serving as a control variable. Therefore, a mixed multivariate analysis of covariance (MANCOVA) was utilized.

For the pre-test DST, the Cronbach’s alpha reliability coefficient was .90. In the immediate post-test, the Cronbach’s alphas of DST and MRT were 0.84 and 0.88, respectively. Moreover, in the delayed post-test, the Cronbach’s alphas of DST and MRT were 0.83 and 0.88, respectively. Hence, all vocabulary tests had high-reliability estimates.

<A>5. Results<A>

<B>5.1 Preliminary data analysis<B>

We first checked the assumptions of the mixed MANCOVA (see online supplementary materials). In what follows, we report the results of immediate and delayed post-tests based on participants’ performance in DST and MRT tests. To make the interpretation more straightforward, we answer RQ1 and RQ2 for immediate and delayed post-tests separately.

<B>5.2 The results of immediate post-test for DST and MRT<B>

The results of MANCOVA showed the differences in between-subject design was statistically significant, Wilks’s lambda = 0.643, *F*(6, 170) = 7.012, *p* < .001, partial *η*2 = 0.198. Moreover, there was a significant interaction effect between conditions and time, Wilks’s lambda = 0.842, *F*(6, 170) = 2.551, *p* = .022, partial *η*2 = 0.083. Table 1 shows descriptive statistics for immediate post-tests.

**<TH>Table 1.** Descriptive statistics for DST and MRT vocabulary tests in the immediate post-test<TH>

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Measure | Group | *M* | *SE* | 95% CI | |
| Lower bound | Upper bound |
| DST | Group A | 11.210 | .987 | 9.248 | 13.171 |
| Group B | 14.022 | .969 | 12.095 | 15.949 |
| Group C | 6.763 | .859 | 5.057 | 8.470 |
| Group D | 9.203 | .826 | 7.561 | 10.845 |
| MRT | Group A | 17.960 | 1.162 | 15.650 | 20.270 |
| Group B | 19.342 | 1.141 | 17.073 | 21.611 |
| Group C | 12.052 | 1.011 | 10.043 | 14.062 |
| Group D | 16.838 | .973 | 14.904 | 18.771 |

*Note*. DST = definition-supply test; MRT = meaning-recognition test; Group A = L1 text-only; Group B = L1 textual-pictorial; Group C = L2 text-only; Group D = L2 textual-pictorial; CI = confidence interval.

The result of MANCOVA in immediate post-tests revealed a significant main effect of treatment on dependent variables, Wilks’s lambda = 0.658, *F*(6, 170) = 6.601, *p*< .001, partial *η*2 = 0.189*.* A univariate ANCOVA was conducted for a more detailed distinction in each dependent variable. The results revealed that there was a significant difference among the four groups in both DST, *F*(3, 86) = 11.266, *p* < .001, partial *η*2 = 0.282, and MRT, *F*(3, 86) = 9.039, *p* < .001, partial *η*2 = 0.240. Table 2 provides the results of the Tukey post-hoc analysis of pairwise comparisons, which were conducted to compare the different performances of each paired condition based on the RQs.

Concerning the differences between L1 and L2 textual glosses (RQ1) in immediate post-tests, Group A had a higher mean than Group B in both immediate DST and MRT tests. Regarding the role of pictures (RQ2) in immediate post-tests, participants who received glosses with pictures showed higher vocabulary scores in the immediate DST in comparison with groups without pictorial glosses. However, for immediate MRT, while Group D showed improved performance in vocabulary scores in comparison with Group C, Group B did not show any significant difference with Group A.

Taken together, in immediate DST, Group B had a significantly higher mean than the other three groups. Moreover, Group A and Group D had significantly higher means than that of Group C. For the immediate MRT test, Group A, Group B, and Group D had significantly higher means than that of Group C.

**<TH>Table 2.**The results ofpairwise comparisons in relation to DST and MRT vocabulary measures in immediate post-test<TH>

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Measure | Groups | | *MD* | *SE* | *p* | 95% CI | |
| Lower bound | Upper bound |
| DST | Group A | Group C | 4.446 | 1.306 | .001 | 1.849 | 7.043 |
| Group B | Group D | 4.819 | 1.276 | .000 | 2.281 | 7.356 |
| Group A | Group B | 2.813 | 1.391 | .046 | .048 | 5.577 |
| Group C | Group D | 2.440 | 1.191 | .043 | .073 | 4.807 |
| Group A | Group D | 2.006 | 1.286 | .122 | −.549 | 4.562 |
| Group B | Group C | 7.259 | 1.298 | .000 | 4.678 | 9.840 |
| MRT | Group A | Group C | 5.907 | 1.538 | .000 | 2.850 | 8.965 |
| Group B | Group D | 2.504 | 1.503 | .099 | −.484 | 5.492 |
| Group A | Group B | 1.382 | 1.637 | .401 | −1.873 | 4.637 |
| Group C | Group D | 4.785 | 1.402 | .001 | 1.998 | 7.572 |
| Group A | Group D | 1.122 | 1.514 | .461 | −1.887 | 4.131 |
| Group B | Group C | 7.289 | 1.529 | .000 | 4.251 | 10.328 |

*Note*. DST = definition-supply test; MRT = meaning-recognition test; Group A = L1 text-only; Group B = L1 textual-pictorial; Group C = L2 text-only; Group D = L2 textual-pictorial; CI = confidence interval.

<B>5.3 The results of delayed post-tests for DST and MRT<B>

Table 3 shows descriptive statistics of the four groups in DST and MRT for the delayed post-test. The result of MANCOVA in delayed post-tests revealed a significant main effect of treatment on dependent variables, Wilks’s lambda = 0.656, *F*(6, 170) = 6.657, *p*< .001, partial *η*2 = 0.190*.* A univariate ANCOVA was conducted for a more detailed distinction in each dependent variable. The result revealed that there was a significant difference among four groups in both DST, *F*(3, 86) = 11.266, *p* < .001, partial *η*2 = 0.282, and MRT, *F*(3, 86) = 9.039, *p* < .0005, partial *η*2 = 0.240. Table 4 provides the results of Tukey post-hoc analysis of pairwise comparisons, which were conducted to compare the different performances of each paired condition based on the RQs.

**<TH>Table 3*.***Descriptive statistics for vocabulary tests in the delayed post-test<TH>

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Measure | Group | *M* | *SE* | 95% CI | |
| Lower bound | Upper bound |
| DST | Group A | 5.974 | .721 | 4.542 | 7.407 |
| Group B | 8.068 | .708 | 6.661 | 9.475 |
| Group C | 3.335 | .627 | 2.089 | 4.581 |
| Group D | 6.806 | .603 | 5.607 | 8.005 |
| MRT | Group A | 15.585 | 1.101 | 13.397 | 17.773 |
| Group B | 16.156 | 1.081 | 14.006 | 18.305 |
| Group C | 9.148 | .958 | 7.245 | 11.052 |
| Group D | 13.410 | .921 | 11.578 | 15.241 |

*Note*. DST = definition-supply test; MRT = meaning-recognition test; Group A = L1 text-only; Group B = L1 textual-pictorial; Group C = L2 text-only; Group D = L2 textual-pictorial; CI = confidence interval.

Concerning the differences between L1 and L2 textual glosses (RQ1) in delayed post-tests, Group A had a higher mean than Group B in both the delayed DST and MRT. Regarding the role of pictures (RQ2) in delayed post-tests, participants who received glosses with pictures showed higher vocabulary scores in the delayed DST in comparison with groups without pictorial glosses. Moreover, for the delayed MRT, participants in Group D demonstrated enhanced vocabulary retention in comparison with Group C, whereas Group B did not exhibit a significant improvement in scores in comparison with Group A.

Taken together, in the delayed DST, Group B had a significant higher mean than that of Group A and Group C. Group A and Group D had significantly higher means than that of Group C. In the delayed MRT test, Group A, Group B, and Group D had significant higher means than that of Group C.

**<TH>Table 4.** The results of pairwise comparisons in relation to DST and MRT vocabulary measures in delayed post-test<TH>

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Measure | Groups | | *MD* | *SE* | *p* | 95% CI | |
| Lower bound | Upper bound |
| DST | Group A | Group C | 2.640 | .954 | .007 | .743 | 4.536 |
| Group B | Group D | 1.262 | .932 | .179 | −.590 | 3.115 |
| Group A | Group B | 2.094 | 1.015 | .042 | .075 | 4.112 |
| Group C | Group D | 3.471 | .869 | .000 | 1.743 | 5.199 |
| Group A | Group D | −.831 | .939 | .378 | −2.697 | 1.034 |
| Group B | Group C | 4.733 | .948 | .000 | 2.849 | 6.617 |
| MRT | Group A | Group C | 6.437 | 1.457 | .000 | 3.541 | 9.334 |
| Group B | Group D | 2.746 | 1.424 | .057 | −.084 | 5.576 |
| Group A | Group B | .571 | 1.551 | .714 | −2.513 | 3.654 |
| Group C | Group D | 4.262 | 1.328 | .002 | 1.621 | 6.902 |
| Group A | Group D | 2.176 | 1.434 | .133 | −.675 | 5.026 |
| Group B | Group C | 7.008 | 1.448 | .000 | 4.130 | 9.886 |

*Note*. DST = definition-supply test; MRT = meaning-recognition test; Group A = L1 text-only; Group B = L1 textual-pictorial; Group C = L2 text-only; Group D = L2 textual-pictorial; CI = confidence interval.

<A>6. Discussion<A>

The first research question examined whether L1 and L2 glosses had differential effects on vocabulary learning. The immediate post-test results showed that participants who had access to L1 textual glosses during listening had significantly higher vocabulary gains than participants who received L2 textual glosses. Similarly, after two weeks of delayed post-tests, participants in the L1 text-only group outperformed participants in the L2 text-only group in both vocabulary tests. These findings are congruent with several empirical studies that found an advantage of L1 glosses over L2 glosses in vocabulary acquisition (Arpaci, 2016; Choi, 2016; Xu, 2010) as well as two recent meta-analyses (Kim *et al*., 2024; Yanagisawa *et al*., 2020). An explanation for these results is that the majority of participants in this study were elementary and lower-intermediate proficiency learners, and based on Kroll and Stewart’s (1994) RHM, learners who have weaker abilities in L2 are more likely to associate new concepts and meanings of unknown words through L1 words and learn better from glosses that are provided in their L1. This can also explain why some previous studies found L2 glossing to be more helpful than L1 glossing or found no significant difference in L1 and L2 glossing in reading comprehension (Cheng & Good, 2009; Jacobs *et al*., 1994; Yoshii, 2006). Therefore, familiarity with L1 words rather than L2 words leads to better connections with target words for lower-proficiency learners.

The second research question examined whether adding pictures to glosses has a beneficial effect on vocabulary learning. Since the findings differed in the DST and the MRT, in what follows, the differences between textual and textual-pictorial groups in each vocabulary test are explained. In the DST that assessed meaning-recall aspects of learning target words, the findings showed that supplementing textual glosses with pictures was beneficial. When pictures were added to both L1 and L2 text-only glosses, the resulting dual glossing modes were more effective in vocabulary learning than their corresponding text-only glosses, both in immediate and delayed tests. These findings confirm the “picture superiority” in Mayer’s (2014) CTML as well as Ramezanali *et al*.’s (2021) meta-analysis. In the DST, participants had to rely on their memory to remember the meaning of words without any clue or assistance from the test itself (Yoshii, 2006). Therefore, based on the CTML, incorporating pictures in textual glosses could further assist learners in future recall processes by providing more memory traces. Mayer (2014) mentioned that providing visual input alongside verbal information helps to form more connections in learners’ minds, since words and pictures are processed through different channels and more interactions between verbal and visual systems lead to better understanding.

In the MRT, while comparing L2 textual-pictorial glosses and L2 text-only glosses in both immediate and delayed tests, it was found that the L2 textual-pictorial group obtained higher scores than the L2 text-only glosses. However, this pattern was not found while comparing L1 textual-pictorial glosses with L1 text-only glosses. Therefore, the presence of pictures did not have any significant effect when glosses were provided in participants’ first language. This finding can be explained based on the results of Ramezanali *et al*.’s (2021) meta-analysis that adding another gloss mode to L2 textual glosses was more effective than adding the same mode to L1 textual glosses. They explain that the reason for this finding might be that L1 glosses by themselves are so effective that adding another mode (i.e. pictures) does not make much difference, suggesting a ceiling effect of L1 glosses. Taken together, based on the results of our study, adding pictures to both L1 and L2 textual glosses was effective in a recall test, but in a recognition test, adding pictures was only effective for L2 textual glosses.

Additional support for this interpretation is provided by Yoshii (2006). According to this author, this phenomenon could be attributed to the nature of visual and textual supplementary input in glosses, which are intended to help learners make conceptual links between the form and meaning of unknown words. In a gloss that is provided in L2, low-ability learners might have difficulties making a connection between the form of an unknown word and its meaning by L2 explanation alone. Therefore, providing an additional picture as a visual cue could assist them. However, in L1 glosses, translations are clear and easy to understand, since the learners’ L1 conceptual repertoire is rich, and conceptual links could be established firmly between the forms and meanings of unknown words. In this situation, providing further visual cues might not improve learners’ understanding of the meaning of unknown words.

Although the results of this study are in line with those of the previous studies in favor of dual-modality glosses (Babaie Shalmani & Khalili Sabet, 2010; Kost, Foss & Lenzini, 1999; Warren, Boers, Grimshaw & Siyanova-Chanturia, 2018), some studies have produced contradictory findings (e.g. Boers, Warren, He & Deconinck, 2017; Çakmak & Erçetin, 2018; Yanguas, 2009). The studies mentioned above found no difference between pictorial-textual glosses and textual-only glosses. One possible explanation for this contradictory result might be due to the study design, method, and participants. As Çakmak and Erçetin (2018) mentioned, using excessive pictorial glosses might distract learners, especially in a listening context where there are some limitations in controlling the flow of listening and using glosses simultaneously. Some studies showed that lower-proficiency learners are more likely to use glosses (e.g. Babaie Shalmani & Khalili Sabet, 2010). However, when the target word is abstract, using pictures is unsuitable for conveying the meaning and could lead to cognitive overload in these learners (Ariew & Ercetin, 2004; Hu *et al*., 2014). When comparing all four groups in both immediate and delayed tests, students who had access to L1 pictorial glosses had the best performance, and those who had access to L2 text-only glosses had the weakest performance in recall and recognition tests.

<A>7. Pedagogical implications and limitations<A>

Some implications regarding the use of glossing in EFL listening for incidental vocabulary learning can be drawn from the findings of this study. First, the findings serve as a reminder to pay more attention to the beneficial role of multimedia annotation in L2 vocabulary learning, especially considering implementing glossing in listening materials that are designed for mobile devices. Therefore, future listening materials could be augmented with different types of annotations to help learners improve their vocabulary knowledge as a by-product of engaging in listening activities. Second, materials developers can consider using pictures in glosses instead of presenting textual information alone, since the findings of this study supported the superiority of textual-pictorial glosses in L2 dual-modality glosses for incidental vocabulary learning (Paivio, 1990). However, care must be taken, since providing excessive visual supportive cues could lead to overload in underachieving learners’ cognitive capacity (Çakmak & Erçetin, 2018; Mayer, 2014; <AQ5>Varol & Erçetin, 2019). Third, as we found that L1 textual glossing worked better than L2 textual glossing for our participants who were mainly at elementary and lower-intermediate levels, we recommend that teachers use L1 textual glossing for these proficiency levels rather than L2 textual glossing. Finally, given that adding pictures to L2 texts was more effective than adding to L1 texts in recognition tests, it can be concluded that using L1 text-only glosses was as effective as L1 textual-pictorial glosses for knowledge of recognition. Therefore, if teachers aim to develop language learners’ receptive knowledge of vocabulary, L1 text-only glosses are sufficient.

There are some limitations to the current study that should be considered. First, participants of this study were Iranian EFL students and mostly low proficiency learners. The generalizability of the findings is limited to this group, and more research is needed to generalize to other contexts. The second limitation was the setting in which the experiment was conducted. Mobile devices are designed to access information anytime and anywhere without any restriction, and most mobile-assisted vocabulary learning takes place outside the classroom (Ma, 2017). The setting of this study was a formal classroom, which is not ideal for participants to exploit the full potential of MALL. Third, it might be of interest to investigate the effect of other gloss types (e.g. aural and video glosses) on incidental vocabulary learning with high-level proficiency learners. Mayer (2014) stated that using a static or dynamic picture in glossing might affect reading comprehension and vocabulary learning differently. Fourth, we did not measure listening comprehension in this study. Measuring listening comprehension can give us a deeper understanding of the relationship between glossing and vocabulary learning. Fifth, we did not measure the frequency of access to each gloss, although this can affect learning of the words. Sixth, we used only a DST as a measure of participants’ vocabulary knowledge in the pre-test. This only measured their productive vocabulary knowledge. Future research can include both DST and MRT in the pre-test to have a more accurate measure of vocabulary knowledge. Finally, as Choi (2016) mentioned, using a longer retention phase could help to better understand the forgetting rate of different types of glossing from listening.

Despite these limitations, the study has demonstrated the potential of MALL in enhancing incidental vocabulary learning, even within the constraints of a formal classroom setting. This underscores the importance of glosses as a complementary educational tool and invites further exploration into its applications across diverse learning environments and proficiency levels. The findings contribute to the growing body of research advocating for the integration of technology in language education, highlighting the need for adaptive learning strategies that cater to the evolving landscape of student engagement.

<A>8. Conclusion<A>

In this article we demonstrated that multimedia glossing, particularly with L1 textual and pictorial elements, significantly enhanced vocabulary learning from listening tasks on mobile devices. The findings confirmed that L1 glosses were more effective than L2 glosses for lower-proficiency learners, and the addition of pictorial information to glosses improved recall of vocabulary. While pictorial glosses had a notable impact when combined with L2 textual glosses, they did not significantly affect vocabulary learning when paired with L1 glosses. This study’s results support the RHM and Mayer’s CTML, suggesting that dual coding and visual aids can be beneficial in language learning. The research advocates for incorporating native language support and multimedia elements in language learning tools for elementary and lower-intermediate learners, aligning with previous studies that favor dual-modality glossing.

**Supplementary material.** To view supplementary material referred to in this article, please visit <TS: Please add the digital object identifier for the supplementary material here.>.

**Acknowledgements.** The authors would like to thank Dr Shona Whyte and three anonymous reviewers for providing us with insightful comments on earlier drafts of the paper.

**Ethical statement and competing interests.** Ethical considerations, such as voluntary participation, anonymity, and confidentiality of the learners, were considered in this study. The authors declare no competing interests. The authors declare no use of generative AI.

<A>References<A>

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. <https://doi.org/10.1007/s11251-007-9016-7>

Ariew, R. & Ercetin, G. (2004) Exploring the potential of hypermedia annotations for second language reading. *Computer Assisted Language Learning*, 17(2): 237–259. <https://doi.org/10.1080/0958822042000334253>

Arpaci, D. (2016) The effects of accessing L1 versus L2 definitional glosses on L2 learners’ reading comprehension and vocabulary learning. *Eurasian Journal of Applied Linguistics*, 2(1): 15–29. <https://doi.org/10.32601/ejal.460988>

Babaie Shalmani, H. & Khalili Sabet, M. (2010) Pictorial, textual, and picto-textual glosses in e-reading: A comparative study. *English Language Teaching*, 3(4): 195–203. <https://doi.org/10.5539/elt.v3n4p195>

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. <https://doi.org/10.1080/09588221.2017.1356335>

Boers, F., Warren, P., He, L. & Deconinck, J. (2017) Does adding pictures to glosses enhance vocabulary uptake from reading? *System*, 66: 113–129. <https://doi.org/10.1016/j.system.2017.03.017>

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. <https://doi.org/10.1017/S0958344017000155>

Çekiç, A. (2024) Incidental L2 vocabulary learning from audiovisual input: The effects of different types of glosses. *Computer Assisted Language Learning*, 37(4): 896–923. <https://doi.org/10.1080/09588221.2022.2062004>

Cheng, Y.-H. & Good, R. L. (2009) LI glosses: Effects on EFL learners’ reading comprehension and vocabulary retention. *Reading in a Foreign Language*, 21(2): 119–142.

Choi, S. (2016) Effects of L1 and L2 glosses on incidental vocabulary acquisition and lexical representations. *Learning and Individual Differences*, 45: 137–143. <https://doi.org/10.1016/j.lindif.2015.11.018>

Elekaei, A., Faramarzi, S. & Koosha, M. (2015) The impact of gloss types on reading comprehension, vocabulary gain and vocabulary retention: A comparative study. *International Journal of Applied Linguistics and English Literature*, 4(5): 97–103. <https://doi.org/10.7575/aiac.ijalel.v.4n.5p.97>

Hsieh, Y. (2020) Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning*, 33(5–6): 567–589. <https://doi.org/10.1080/09588221.2019.1577898>

Hu, S.-M., Vongpumivitch, V., Chang, J. S. & Liou, H.-C. (2014) The effects of L1 and L2 e-glosses on incidental vocabulary learning of junior high-school English students. *ReCALL*, 26(1): 80–99. <https://doi.org/10.1017/S0958344013000244>

Huckin, T. & Coady, J. (1999) Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition*, 21(2): 181–193. <https://doi.org/10.1017/S0272263199002028>

Hulstijn, J. H. (2001) Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In Robinson, P. (ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press, 258–286. <https://doi.org/10.1017/CBO9781139524780.011>

Jacobs, G. M., Dufon, P. & Hong, F. C. (1994) L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading*, 17(1): 19–28. <https://doi.org/10.1111/j.1467-9817.1994.tb00049.x>

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. <https://doi.org/10.1111/1540-4781.00160>

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. <https://doi.org/10.1016/j.system.2017.01.009>

Kim, H. S., Lee, J. H. & Lee, H. (2024) The relative effects of L1 and L2 glosses on L2 learning: A meta-analysis. *Language Teaching Research*, 28(1): 7–28. <https://doi.org/10.1177/1362168820981394>

Ko, M. (2017) The relationship between gloss type and L2 proficiency in incidental vocabulary learning. *Modern English Education*, 18(3): 47–69. <https://doi.org/10.18095/meeso.2017.18.3.03>

Kost, C. R., Foss, P. & Lenzini, J. J., Jr. (1999) Textual and pictorial glosses: Effectiveness on incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals*, 32(1): 89–97. <https://doi.org/10.1111/j.1944-9720.1999.tb02378.x>

Kroll, J. F. & Stewart, E. (1994) Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33(2): 149–174. <https://doi.org/10.1006/jmla.1994.1008>

Kroll, J. F., Van Hell, J. G., Tokowicz, N. & Green, D. W. (2010) The revised hierarchical model: A critical review and assessment. *Bilingualism: Language and Cognition*, 13(3): 373–381. <https://doi.org/10.1017/S136672891000009X>

Laufer, B. (2000) Electronic dictionaries and incidental vocabulary acquisition: Does technology make a difference? In Heid, U., Evert, S., Lehmann, E. & Rohrer, C. (eds.), *Proceedings of the Ninth* *EURALEX International Congress*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 849–854.

Lin, J.-J. & Lin, H. (2019) Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning*, 32(8): 878–919. <https://doi.org/10.1080/09588221.2018.1541359>

Ma, Q. (2017) Technologies for teaching and learning L2 vocabulary. In Chapelle, C. A. & Sauro, S. (eds.), *The handbook of technology and second language teaching and learning*. Hoboken: Wiley Blackwell, 45–61. <https://doi.org/10.1002/9781118914069.ch4>

Mayer, R. E. (2009) *Multimedia learning* (2nd ed.). Cambridge: Cambridge University Press.

Mayer, R. E. (ed.) (2014) *The Cambridge handbook of multimedia learning* (2nd ed.). New York: Cambridge University Press.

McQuiggan, S., Kosturko, L., McQuiggan, J. & Sabourin, J. (2015) *Mobile learning: A handbook for developers, educators, and learners*. Hoboken: John Wiley. <https://doi.org/10.1002/9781118938942>

Mohsen, M. A. & Balakumar, M. (2011) A review of multimedia glosses and their effects on L2 vocabulary acquisition in CALL literature. *ReCALL*, 23(2): 135–159. <https://doi.org/10.1017/S095834401100005X>

Montero Perez, M., Peters, E. & Desmet, P. (2018) Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning*, 31(1–2): 1–26. <https://doi.org/10.1080/09588221.2017.1375960>

Nation, I. S. P. (2017) *The BNC/COCA Level 6 word family lists* (Version 1.0.0) [Data file]. http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Paivio, A. (1990) *Mental representations: A dual coding approach*. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195066661.001.0001>

Ramezanali, N. & Faez, F. (2019) Vocabulary learning and retention through multimedia glossing. *Language Learning & Technology*, 23(2): 105–124.

Ramezanali, N., Uchihara, T. & Faez, F. (2021) Efficacy of multimodal glossing on second language vocabulary learning: A meta‐analysis. *TESOL Quarterly*, 55(1): 105–133. <https://doi.org/10.1002/tesq.579>

Rashtchi, M. & Pourmand, S. (2014) Controlled extensive reading and vocabulary knowledge: Let’s move towards autonomous learning in EFL contexts. *International Journal of English and Education*, 3(2): 132–146.

Tabachnick, B. G. & Fidell, L. S. (1999) *Using multivariate statistics.* New York: Harper Collins.

van Zeeland, H. & Schmitt, N. (2013) Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, 41(3): 609–624. <https://doi.org/10.1016/j.system.2013.07.012>

Varol, B. & Erçetin, G. (2016) Effects of working memory and gloss type on L2 text comprehension and incidental vocabulary learning in computer-based reading. *Procedia – Social and Behavioral Sciences*, 232: 759–768. <https://doi.org/10.1016/j.sbspro.2016.10.103>

Vidal, K. (2011) A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1): 219–258. <https://doi.org/10.1111/j.1467-9922.2010.00593.x>

Waring, R. & Nation, P. (2004) Second language reading and incidental vocabulary learning. In Albrechtsen, D., Haastrup, K. & Henriksen, B. (eds.), *Angles on the English-speaking world: Vol. 4. Writing and vocabulary in foreign language acquisition*. Copenhagen: Museum Tusculanum Press, 97–110.

Warren, P., Boers, F., Grimshaw, G. & Siyanova-Chanturia, A. (2018) The effect of gloss type on learners’ intake of new words during reading: Evidence from eye-tracking. *Studies in Second Language Acquisition*, 40(4): 883–906. <https://doi.org/10.1017/S0272263118000177>

Widdowson, H. G. (1978) *Teaching language as communication*. Oxford: Oxford University Press.

Xu, X. (2010) The effects of glosses on incidental vocabulary acquisition in reading. *Journal of Language Teaching and Research*, 1(2): 117–120. <https://doi.org/10.4304/jltr.1.2.117-120>

Yanagisawa, A., Webb, S. & Uchihara, T. (2020) How do different forms of glossing contribute to L2 vocabulary learning from reading? A meta-regression analysis. *Studies in Second Language Acquisition*, 42(2): 411–438. <https://doi.org/10.1017/S0272263119000688>

Yanguas, I. (2009) Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning & Technology*, 13(2): 48–67. <https://doi.org/10125/44180>

Yeh, Y. & Wang, C.-W. (2004) Effects of multimedia vocabulary annotations and learning styles on vocabulary learning. *CALICO Journal*, 21(1): 131–144. <https://doi.org/10.1558/cj.v21i1.131-144>

Yoshii, M. (2006) L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, 10(3): 85–101.

Yoshii, M. & Flaitz, J. (2003) Second language incidental vocabulary retention: The effect of text and picture annotation types. *CALICO Journal*, 20(1): 33–58. <https://doi.org/10.1558/cj.v20i1.33-58>

Zhang, R. & Zou, D. (2022) A state-of-the-art review of the modes and effectiveness of multimedia input for second and foreign language learning. *Computer Assisted Language Learning*, 35(9): 2790–2816. <https://doi.org/10.1080/09588221.2021.1896555>

**About the authors**

**Arman Shahipanah** is a lecturer with a master’s degree in teaching English as a foreign language (TEFL) from the University of Bojnord. His research interests include mobile-assisted language learning, generative artificial intelligence, e-learning, and digital media.

**Gholam Hassan Khajavy** (PhD) is an associate professor of applied linguistics at the University of Bojnord and postdoctoral researcher at University of Graz. His main research interests are the psychology of language learning and teaching and instructed second-language acquisition. He has published in various international journals such as *TESOL Quarterly* and *Language Learning*.

**Majid Elahi Shirvan** is an associate professor at the University of Bojnord, Iran. His main research interest is the dynamics of psychological aspects of foreign language learning and teaching. He has published in different leading international journals such as *Ecological Psychology*, *Foreign Language Annals*, and *Studies in Second Language Acquisition*.

Author #1 ORCiD. Arman Shahipanah, <https://orcid.org/0000-0001-6473-2639>

Author #2 ORCiD. Gholam Hassan Khajavy, <https://orcid.org/0000-0002-0926-960X>

Author #3 ORCiD. Majid Elahi Shirvan, <https://orcid.org/0000-0003-3363-8273>

**Author queries**

**AQ1:** As there are two references that shorten to Boers *et al*. (2017), to prevent ambiguity, please cite all of the authors to ensure that a distinction can be made between the two.

**AQ2:** As there are two references that shorten to Boers *et al*. (2017), to prevent ambiguity, please cite all of the authors to ensure that a distinction can be made between the two.

**AQ3:** Please double-check phrasing. Is there a word(s) missing here?

**AQ4:** Flesch (1948) is cited in text but does not appear in the reference list. Please supply a reference formatted according to journal style.

**AQ5:** Varol & Erçetin (2019) is cited in text but does not appear in the reference list. Did you intend to cite Varol & Erçetin (2016) instead? If not, please supply a reference for Varol & Erçetin (2019) formatted according to journal style or delete citation from text.