# Effect of Computerized Gloss Presentation Format on Reading Comprehension: A Cognitive Load Perspective

## *Hamideh Marefat, Abbas Ali Rezaee, and Farid Naserieh*
## *University of Tehran, Tehran, Iran*

## marefat@ut.ac.ir; aarezaee@ut.ac.ir; naserieh@gmail.com

## Abstract

In recent years, gloss presentation format or the location where a gloss appears with respect to its related target word has received renewed attention. Research suggested that different gloss presentation formats could have differential effects on reading comprehension and/or vocabulary learning. This study hypothesized that the effectiveness of different computerized gloss presentation formats in reading comprehension could be explained by drawing on the split-attention effect within the framework of cognitive load theory. The effect predicts that when two related sources of information are physically separated (e.g., a target word and the respective gloss), cognitive resources are unnecessarily wasted, and learning is hindered. To test this hypothesis, 39 Persian-speaking L2 learners of English were divided into two experimental conditions, each being exposed to a text enhanced with either in-text or marginal glosses. Two measures of reading comprehension and three measures of cognitive load were employed. The participants' initial differences in terms of grammatical knowledge and vocabulary size were balanced out, and their look-up behavior was also tracked. The results revealed that the participants with access to in-text glosses, compared with those with marginal glosses, experienced lower levels of cognitive load due to the elimination of split-attention and, accordingly, performed better on the reading comprehension measures. Given the participants' L2 proficiency level, the findings suggested that a text enhanced with in-text glosses tends to be instructionally more efficient.

**Keywords***:* gloss presentation format, computerized glosses, reading comprehension, cognitive load, split-attention effect

## Introduction

Enhancing L2 (second language) texts with vocabulary glosses or annotations in the form of, for instance, short definitions or L1 (first language) equivalents of difficult words in the text is a common practice in second or foreign language instructional materials. Previous gloss studies have extensively investigated the effects of different gloss types (e.g., verbal, visual, multimedia, multiple-choice) on promoting reading comprehension and/or enhancing vocabulary learning (Al-Seghayer, 2001; Chun & Plass, 1996; Hulstijn, Hollander, & Greidanus, 1996; Lomicka, 1998; Miyasako, 2002; Yanguas, 2009). Recently, however, gloss presentation format (e.g., in-text, marginal, pop-up) has received renewed attention (e.g.,

AbuSeileek, 2011; Chen, 2016; Chen & Yen, 2013; Cheng & Good, 2009). Such research has shown that the way a gloss is presented on screen or page and its position relative to the respective unknown target word can influence readers' performance on measures of reading comprehension and/or vocabulary learning. Although some scholars (e.g., Chen, 2016) have reported that the shortness of distance between the target word and the related gloss might be a contributing factor in the general usefulness of one gloss presentation format over another, little, if any, empirical evidence can be found in past gloss studies as to how and why this spatial proximity could influence the learning outcomes.

Answers may be found when the effectiveness of different gloss presentation formats is considered within the framework of cognitive load theory (Sweller, 1988). The theory, which takes into account limitations of the human cognitive architecture, delineates an instructional condition called the split-attention effect. According to this effect, when two related sources of information, which are necessary for learning to occur, are separated in space, cognitive resources are wasted, and chances of effective learning are minimized (Ayres & Sweller, 2005). By implication and based on the predictions of cognitive load theory, therefore, when the target word in a text and its related gloss are physically close to one another (as in in-text gloss presentation format), lower levels of working memory or cognitive load are induced, and learning is, accordingly, promoted. In contrast, as the physical distance between the target word and the respective gloss increases (as in marginal gloss presentation format), higher yet unnecessary levels of working memory load are generated that could negatively affect the learning outcomes.

Drawing on the split-attention effect within the cognitive load theory, the current study investigated the effectiveness of two gloss presentation formats (viz., in-text and marginal) in promoting reading comprehension. Set in a CALL (Computer-Assisted Language Learning) environment, the study employed different measures of cognitive load to offer a plausible explanation from a cognitive load perspective as to which gloss presentation format tends to be instructionally more efficient and why.

# Literature Review

Gloss, or annotation, is a short definition, explanation, or synonym providing the meaning of a difficult or important word in a text in order to, primarily, facilitate reading and comprehension processes. Glosses provide easy access to the definition of unknown words without necessarily disrupting the normal flow of reading (Lomicka, 1998; Nation, 2001). One aspect of glossing is gloss presentation format or the location where the gloss appears with respect to the related target word (Roby, 1999). Two common gloss presentation formats include marginal (appearing in the margin of the text) and in-text (appearing within the text close to the target words). Early on, using paper glosses, Watanabe (1997) reported that marginal glosses, compared with glosses embedded within the text in the form of appositives (embedded within the text), tended to be more effective in enhancing vocabulary learning among Japanese students. In Cheng and Good's (2009) study, however, Chinese participants who had access to in-text glosses performed better on vocabulary learning and retention than those with access to marginal glosses. No difference in terms of reading comprehension was reported; the authors argued that the presence of glosses might have distracted the readers from reading for the main idea.

Unlike traditional paper glosses that might clutter the page and distract readers (Roby, 1999), computer glosses are interactive in the sense that they could be initially invisible and appear only when the readers need them (Yun, 2011). With technological advancements, the use of computer glosses has increased in recent years. In particular, gloss presentation format has, accordingly, received special, renewed attention (e.g., AbuSeileek, 2008, 2011; Chen, 2016; Chen & Yen, 2013). For example, AbuSeileek (2011) reported in-text glosses (or what he called "after glossed word" glosses) as the most effective gloss presentation format in enhancing reading comprehen-

sion in comparison with glosses appearing in the margin or at the bottom of the screen. The in-text glosses in his study were initially hidden and were shown only after the reader clicked on the target words. Using such glosses, as the researcher argued, did not distract the participants as their attention was not split between the text and the glosses.

In Chen and Yen's (2013) study among Taiwanese university students, however, in-text glosses were the least beneficial in promoting reading comprehension across all proficiency levels, even less effective than when the text had no gloss at all. The most effective gloss format was what the researchers called pop-up glosses (i.e., glosses that were initially hidden and were shown by the click of the mouse in front of the text). In the study, in-text glosses, like pop-up glosses, were very close to the respective target words, but they, unlike pop-up glosses, were not initially hidden from sight as they were already embedded into the text next to the target words. Chen and Yen argued that such in-text glosses might have interrupted the reading flow, hence impeding comprehension. In stark contrast, Chen (2016) found that a similar type of in-text glosses as used in Chen and Yen's study to be the most beneficial gloss presentation format, with pop-up glosses being the least effective this time. Chen, nevertheless, attributed this to the immediate physical availability of in-text glosses as they were close to the respective target words.

On the whole, regardless of the different names researchers adhered to gloss presentation formats (that may be responsible for part of the inconsistencies), it seems that the crux of gloss presentation format relates to the spatial separation of the gloss content relative to the respective unknown, target word (i.e., the physical distance between the target word in the text and its meaning in the gloss). This could be investigated within the framework of cognitive load theory and one of its effects called the split-attention effect.

Cognitive load theory, developed by John Sweller (1988), offers principled and empirically-supported strategies to help design instructional materials in a way so that the total cognitive load imposed by the materials will not exceed learners' limited working memory capacity. Cognitive load theory is a theory of instructional efficiency; that is, it seeks to explain, given a particular group of learners, which instructional designs can lead to more efficient learning outcomes and which ones may be detrimental to learning (Sweller, 2010; Sweller, Ayres, & Kalyuga, 2011).

Cognitive load can be divided into two types or sources: intrinsic load and extraneous load (Moreno & Park, 2010). Intrinsic load, as the good load, relates to the inherent difficulty or complexity of the material. It is generally regarded as the load relevant and germane to learning. In other words, a certain level of intrinsic load is essential in order for schema acquisition and learning to occur, and "it can be advantageous to increase intrinsic cognitive load" provided that it does not exceed the limited working memory capacity (Sweller et al., 2011, p. 67). The second type of cognitive load, extraneous load, is, in contrast, the bad load. This is related to inefficient instructional design factors that require mental effort to be invested in activities other than schema acquisition (Sweller, 2005). It is generally regarded as the load irrelevant to learning and wasteful of cognitive resources. In other words, extraneous load should always be lowered and kept to a minimum as much as possible because "it is never advantageous to increase extraneous cognitive load" due to its unnecessary consumption of cognitive resources, which may have detrimental effects on learning (Sweller et al., 2011, p. 67).

The total cognitive load is determined as the combination of intrinsic load (which should be managed) and extraneous load (which must be minimized). This is referred to as the additivity hypothesis, according to which if the total cognitive load exceeds the limited working memory capacity, information processing in general and learning in particular will be negatively affected (Sweller, 2010). According to this hypothesis, the only source of cognitive load that can be reduced by instructional design is extraneous cognitive load. In fact, the primary goal of cognitive load theory is to minimize extraneous load so as to set working memory resources free to be used

in other, germane mental activities, thus facilitating schema acquisition during instruction (Moreno & Park, 2010).

One important corollary of the additivity hypothesis is that if the same material, inducing the same level of intrinsic load, is provided to two comparable groups of learners but through two different instructional designs, any difference in the learning outcomes may be attributed to the differential levels of extraneous load; that is, the more efficient instructional design would impose a lower level of extraneous load, resulting in more positive learning outcomes. In cognitive load theory terminology, this design is regarded as instructionally more efficient because it induces a lower level of extraneous load and could, accordingly, result in more desirable learning outcomes (Ayres, 2006; Sweller et al., 2011).

To operationalize how intrinsic load should be managed and how extraneous load should be minimized, scholars have proposed and tested different cognitive load effects, which are experimental demonstrations showing how "an instructional procedure based on cognitive load theory principles facilitates learning or problem solving compared to a more traditional procedure" (Sweller et al., 2011, p. 88). Over the years, cognitive load theory research has evaluated and established several cognitive load effects based on the predictions of the theory in order to guide effective instructional designs. Some of these effects include worked example, imagination, isolated elements, element interactivity, expertise reversal, redundancy, modality, and split-attention effects (Brünken, Seufert, & Paas, 2010; Sweller et al., 2011).

The split-attention effect (as the main focus of the current study) deals with two or more sources of information that are of the same modality (e.g., only written text). It concerns the mental integration of multiple sources of information that are not redundant. Put another way, these sources must be unintelligible in isolation; that is, they are separate but related as they are both required and must be integrated in order to understand the material. The effect occurs when multiple sources of information are separated from one another either in space or in time (Ayres & Sweller, 2005; Mayer & Moreno, 2010; Sweller et al., 2011). Spatial split-attention refers to the physical separation between the sources, whereas temporal split-attention concerns a time interval. In both cases, because learners' attentional resources are split to integrate the sources of information required for the completion of the task, cognitive load will be increased as working memory resources are unnecessarily expended; as a result, learning is likely to be impeded. In contrast, an instructional design strategy without the split-attention effect and based on the physical integration of sources (or synchronization in temporal split-attention) can help reduce cognitive load and improve learning because cognitive resources in working memory are not simply wasted. As one of the early studies exploring the negative consequences of the split-attention effect, Sweller, Chandler, Tierney, and Cooper (1990) found out that when explanatory notes were physically integrated into the diagram at their closest points of reference in science textbooks, the split-attention effect was eliminated, thus reducing total cognitive load with more beneficial learning outcomes.

In relation to the main theme of the current study (i.e., gloss presentation format explored within the framework of cognitive load theory), a previous research attempt is noteworthy. Using subjective ratings of cognitive load, Yeung, Jin, and Sweller (1998; see also Yeung, 1999), in one of their experiments, studied the effectiveness of the use of explanatory notes in reading comprehension among low L2 proficiency learners. They showed that when the notes (i.e., definitions of unknown words) were integrated into the text and in close proximity to the difficult words (viz., integrated format, resembling in-text glosses), the participants' reading comprehension improved. In contrast, when the notes were presented in a separate list of glossary at the end of the text (viz., separated format), reading comprehension was impeded. The authors attributed the results to the elimination of split-attention in the integrated format (as indicated by a lower level of cognitive load reported by the participants) in which more working memory cognitive resources were de-

voted to the primary task of reading comprehension. By comparison, the split-attention effect occurred in the separated format (as indicated by a higher level of cognitive load perceived by the participants) in which a higher extraneous cognitive load was imposed, hence decreasing comprehension.

In sum, according to Yeung et al.'s (1998) findings, the spatial distance between the unknown word and the related gloss may play a significant role (as predicted by the split-attention effect). By implication, for marginal gloss presentation format, as the conventional gloss format, the split-attention effect may occur because there is some physical distance between the target word and its meaning in the gloss. Put simply, "it is obvious that the conventional glosses create the split-attention effect (increase extraneous cognitive load)" (F. Paas, personal communication, September 23, 2014). As a consequence, unnecessary extraneous load may be induced for marginal gloss condition, leading to poor comprehension (due to frequent distractions away from the primary goal of reading for comprehension). The spit-attention is, however, eliminated in the case of in-text gloss presentation format due to the close proximity of unknown word to its gloss; this format is, accordingly, predicted to create an optimal condition conducive to improved text comprehension, the primary goal of glossing.

To the researchers' knowledge, none of the studies exploring different computerized gloss presentation formats used any measure of cognitive load to investigate potentially differential levels of cognitive load (due to the split-attention effect or otherwise) that may result in differential reading comprehension levels. Yeung et al.'s (1998) study, nevertheless, used paper materials and glosses. A problem with their study was that, in the case of integrated glossary, the text page was so cluttered with definitions that the reading texts for the two groups (viz., separated and integrated formats) did not appear comparable in terms of the density of the words on each page, a factor which could, in turn, lead to unaccounted-for extraneous cognitive load (Roby, 1999; Sweller et al., 2011). This confounding factor might have distorted the findings. Additionally, the use of paper glosses made tracking participants' interactions with glosses virtually impossible. As stressed by Laufer and Hill (2000), without tracking look-up behavior, there is no convincing evidence as to whether the learners actually looked up the words or they just ignored them. Put simply, without tracking functionalities, one can never be sure "whether students consult the gloss or not, how long it takes, whether they are associating something else, ..." (S. Rott, personal communication, September 17, 2014).

The purpose of the present study was, thus, to investigate the effect of two L1 gloss presentation formats (viz., in-text and marginal) on L2 reading comprehension from a cognitive load perspective. A text enhanced with glosses was presented on the computer screen, and the participants' interactions with the glosses were tracked. Measures of cognitive load (viz., cognitive load subjective ratings, response time to a secondary task, and time-on-task) were also used to aid in determining the more effective gloss presentation format in terms of facilitating reading comprehension. Based on the purpose of the study, the following research questions were formulated:

1. What is the effect of L1 gloss presentation format (in-text and marginal) on L2 reading comprehension?

2. What is the effect of L1 gloss presentation format (in-text and marginal) on the generated cognitive load level?

3. Which L1 gloss presentation format (in-text or marginal) is instructionally more efficient?

# Method

## *Design*

The design of the study was randomized matched subjects, posttest-only comparison group (Ary, Jacobs, Sorenson, & Razavieh, 2010). The participants were assigned to either of the two experimental treatments (viz., in-text or marginal) based on a mechanical matching of their scores on a grammar test (Fraenkel, Wallen, & Hyun, 2011). In addition, the participants took a L2 vocabulary size test whose scores were used as a covariate in the analyses to statistically adjust for their initial differences in terms of vocabulary knowledge.

## *Participants*

The initial pool of the participants comprised 49 university students coming from two intact classes. Ten students were left out from the final analyses because they could not meet the selection criteria. Two students did not attend the second session of the study; three did not complete at least one of the influential instruments in the study or had a large number of missing values in their data; and five were not attentive enough as they had inordinately short time-on-task and/or they accessed fewer than 75% of the total number of glosses (following the practice of Wallen, Plass, & Brünken, 2005).

The final sample consisted of 39 students (8 males, 31 females) at the University of Applied Science and Technology in western Iran; they were either first- or second-year non-English majors who had already had at least six years of formal English education in school before entering the university. Their age ranged from 19 to 33 years (mean age = 21.08), and they were, in general, at a pre-intermediate (or lower) proficiency level in English as suggested by their scores on the grammar and vocabulary size tests. The participants took part in the study as part of their weekly-scheduled classes of General English course, and 10% of the course grade was awarded to each of them for their participation.

## *Materials*

### Text

The text used in the study was titled "The Future of Water," selected from a chapter in *Reading for the Real World* reading series (Prochaska & Taylor, 2009). The text consisted of 461 words presented in six paragraphs (see the Appendix). The text was selected through two screening phases from among eight texts based on objective and subjective criteria (phase one) as well as the results of a pilot study (phase two). In the first phase of the text selection process, eight texts were chosen from a number of English reading books considering a range of criteria such as topic (Erçetin, 2010), text type (Abraham, 2008; Zumbach & Mohraz, 2008), and text difficulty (Ko, 2005). The texts were also checked and rated by two experienced English teachers for these criteria. To further investigate the textual features of the texts from an objective standpoint, free online tools for text analysis were also used (viz., *Coh-Metrix* developed by McNamara, Louwerse, Cai, & Graesser, 2013; the *Compleat Lexical Tutor* maintained by Cobb, 2015). Considering all the analyses, three texts that were more or less comparable in various aspects were selected.

In the second phase of the text selection process, the three selected texts were piloted using Webb and Nation's (2008) procedures for determining the vocabulary load of texts. To this end, a separate group of students (who did not take part in the study but were selected from the same pool of participants) were asked to complete the bilingual, 50-item version of the vocabulary size test. Then, they were requested to read each of the three texts for comprehension, circle every word

that they did not know, and also rate each text in terms of difficulty, content familiarity, and topic interest.

The mean score for these students' scores on the vocabulary size test was regarded as the approximate average number of word families typical students with similar background knowledge might know (i.e., 28.90 = rounded up to 30, meaning the first 3000 English word families). Then, the three texts were also analyzed by the RANGE program (Nation & Heatley, 2002) to determine their vocabulary load and lexical diversity in terms of the 14,000 English word families. And finally, applying Webb and Nation's procedures and juxtaposing all the above information resulted in the selection of a single text that was most likely appropriate for the final sample of participants considering their vocabulary size.

## Target words

The results of the second phase of the text selection process also aided in selecting the words to be glossed. The output of the RANGE program indicated the words likely to be unknown to the students given their estimated vocabulary size. Two TEFL experts checked these words against a range of other criteria such as the relevance of the words to comprehension of the main ideas (Hulstijn, 1993; Lenders, 2008; Peters, Hulstijn, Sercu, & Lutjeharms, 2009; Rott, 2007), frequency of word occurrence (Al-Seghayer, 2003; Huang & Lin, 2014; Rott, 2005) and parts of speech (Laufer, 1990) and prepared an agreed-upon list of 14 target words to be finally glossed (see the Appendix). The words occurred once in the text and were equally split between verbs and adjectives, which, according to Laufer, are of moderate difficulty level among other parts of speech.

At this stage, the final selected target words were replaced with pseudowords. Using pseudowords is an increasingly common approach to ensure no preexisting differences among participants in terms of knowledge of target words and also to obviate the need for a vocabulary pretest that might threaten the validity of the study (Peters, 2007; Peters et al., 2009; Schmitt, 2010; Webb, 2007). For creating the pseudowords, Wuggy, a multilingual pseudoword generator software (Keuleers & Brysbaert, 2010), was used. Accordingly, based on the word structure of the selected target words, a number of pseudowords were generated using the software. Then, some TEFL experts as well as students were asked to specify those pseudowords looking most like real English words. Based on the results, the most appropriate pseudoword for each target word was identified; the related verb or adjective inflection was also added to the pseudoword before it replaced the respective target word (see the Appendix).

## *Gloss Conditions*

There were two gloss conditions in the study: in-text and marginal. The participants in both conditions were exposed to the same text and gloss content. For both conditions, the glosses were initially invisible, and the target words were highlighted and set in a color different from the other words. The participants could consult the meaning of each target word by clicking on it and holding down the left mouse button to view the respective gloss content in a small pop-up box; as soon as the mouse button was released, the box disappeared again.

The only difference between the two gloss conditions was the location where the gloss appeared. For the in-text gloss condition, the gloss was placed above and spatially close to the target word in the text, whereas for the marginal gloss condition, the gloss was located in the right-hand margin of the page and at a distance from the text and the target word (see Figures 1 & 2 for screenshots of each condition). The gloss content (i.e., what appeared in the pop-up box) for both conditions was the L1 (Persian) equivalents of the target words in the texts; this decision was based on

the finding that lower level learners may benefit more from L1 rather than L2 glosses (Abraham, 2008; Cheng & Good, 2009; Ko, 2005; Miyasako, 2002).
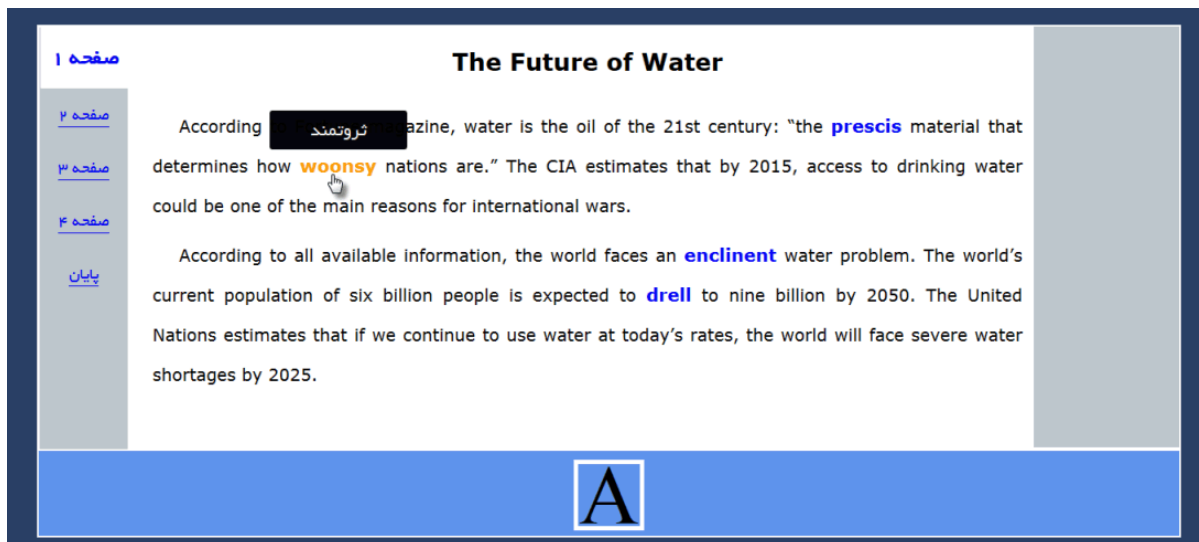


**Figure 1: In-text Gloss Condition**
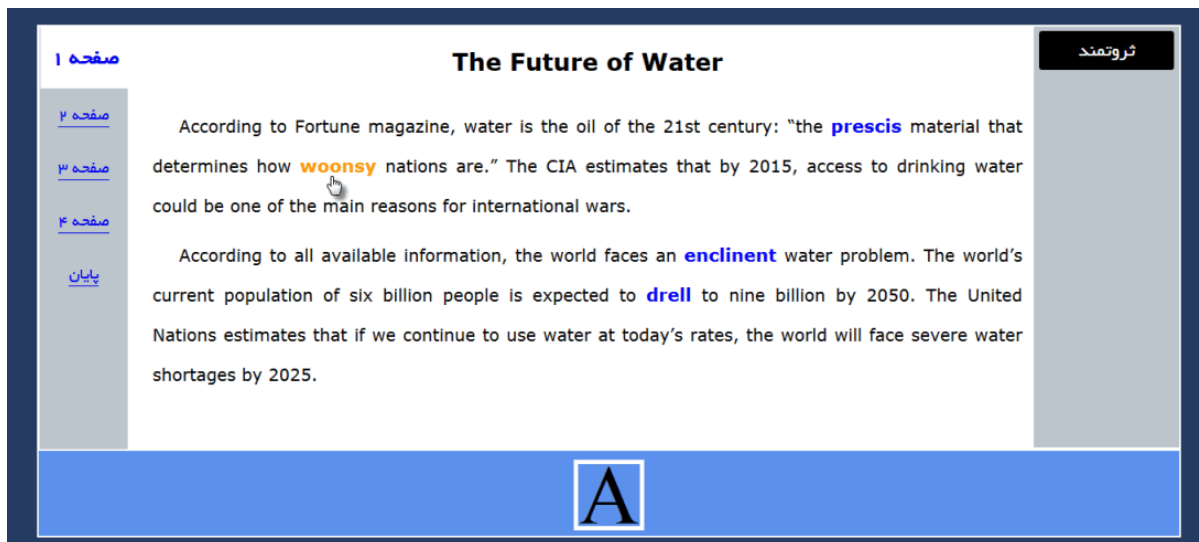


**Figure 2: Marginal Gloss Condition**

It is important to note that the in-text glosses in the current study may resemble Chen's (2016) pop-up glosses. An in-text gloss is usually referred to as a gloss that is embedded right after the target word within the text, and it is already displayed. But, in the present study, in both gloss presentation formats, pop-up technology was employed to set the glosses to be initially hidden; therefore, both gloss presentation formats here could be regarded as pop-up glosses too.

## *Instruments*

The main instruments included a grammar test, a vocabulary size test, cognitive load measures, and reading comprehension measures.

## Grammar and vocabulary size tests

Grammatical knowledge and vocabulary knowledge have been reported to be the strongest correlates of reading comprehension (Jeon & Yamashita, 2014). These two were controlled for in the current study in two different ways. First, the participants' scores on a grammar test were ranked, and, then, each pair of participants, based on their even and uneven rank positions, were assigned randomly to either in-text or marginal group (i.e., through a mechanical matching technique, Fraenkel et al., 2011). The grammar test was a computerized version of one of the Nelson English Language Tests (Fowler & Coe, 1976). Intended for low-level learners, the test used in the study consisted of 50 multiple-choice items measuring different aspects of grammatical knowledge. The Cronbach's alpha for the whole test was .44; the low value might have been due to the restricted distribution of scores on the test ($M = 16.00$, $SD = 4.34$), which was also positively skewed (Bachman, 2004).

Second, the participants' scores on a L2 vocabulary size test were used as a covariate in the analyses to balance out their initial differences in terms of vocabulary knowledge. The Vocabulary Size Test (Nation & Beglar, 2007) used in the study is a measure of written receptive vocabulary size from the first 1,000 to the fourteenth 1,000 word families of English. The test has undergone extensive validation processes with reportedly acceptable reliability indexes (e.g., Beglar, 2010; Schmitt, 2010; Zhang, 2013). A computerized, bilingual (English-Persian) version of the test consisting of only the first five sections (50-item) was employed for the study (Karami, 2012). Using a short, bilingual version (related to the 1000-5000 English word families) of the test is preferable for less advanced learners (Nation, 2012, Webb & Nation, 2008). Each item on the test is in a four-option multiple-choice meaning-recognition format in Persian, with the target word and a non-defining example sentence as the stem. A paper-based version of the test was also used in the second phase of the text selection process. The Cronbach's alpha value for the whole test in the main study was .76.

## Cognitive load measures

Three measures of cognitive load were employed in this study: cognitive load subjective ratings, response time to a secondary task, and time-on-task (for detailed information on each of these, refer to Brünken et al., 2010 and Sweller et al., 2011). Cognitive load subjective rating scale included a computerized two-item self-rating scale originally developed by Paas (1992). The items, which were based on a 9-point Likert-type scale, asked the participants to rate their perceived mental effort invested in comprehending the text and the difficulty level of comprehending the text. The participants' ratings on the two items were averaged for each participant. Both items were presented in Persian. The Cronbach's alpha for the cognitive load ratings was as high as .83. This was related to the significant correlation between the items (about .72), which has been reported in earlier studies using the same scale (e.g., Paas, van Merriënboer, & Adam, 1994).

The second measure included reaction time to a secondary task, with longer times indicating greater cognitive load. The idea is to measure how engaged learners are in the primary task when they are required to perform an additional activity (i.e., the secondary task). This activity is usually dissimilar to the primary task and requires less working memory resources (Sweller et al., 2011). The one used in this study included a visual secondary task that was an adaptation of the spatially non-continuous secondary task applied in Schoor, Bannert, and Brünken (2012). Accordingly, a black English letter wrapped in a box was displayed in the middle lower part of the screen below the main text (see Figures 1 & 2). After the participants started reading the text, the English letter changed color from black to red. The color change probe occurred four times, at specified time intervals. The intervals were set quite arbitrarily, without any discernible patterns. The participants had been instructed to press the space bar on their keyboards as fast as possible when they noticed the color change. After pressing, the color changed back to black. The partici-

pants' reaction times as measured by the time lapse between the beginning of each color change probe and the pressing of the space bar were recorded via the tracking functionality. As practiced by Schoor et al. (2012), the reaction time on the first probe, which was for letting the participants get used to the probes, was not counted, and, only the following probes (#2-4) were calculated. For those participants who did not react within five seconds, the reaction time for the probe was recorded as 5000 (five milliseconds) as an indication of non-reaction, and the letter color was automatically set back to black. The mean reaction time (from the color change probes #2-4) for each participant was calculated in milliseconds.

The third measure of cognitive load was time-on-task or the total amount of time each participant spent on reading the text. The time was recorded in seconds by the tracking functionality embedded in the text web page. Longer time-on-task is usually associated with greater cognitive load (Sweller et al., 2011).

Aside from the distinct measures of cognitive load employed in the study, instructional efficiency was also reported; that is, whether the mental effort invested into the task has been worthwhile given the learning outcome (Sweller et al., 2011). For calculating instructional efficiency for each of the gloss conditions in the current study, first, the participants' ratings of perceived mental effort (i.e., responses to the first item on the cognitive load subjective rating scale) and their reading comprehension composite scores were averaged and transformed into Z-scores, and the resultant values were plugged into the following formula (Paas & van Merriënboer, 1993; Van Gog & Paas, 2008):

$$\text{Instructional Efficiency} = \frac{(\text{ZTest} - \text{ZEffort})}{\sqrt{2}}$$

where ZTest and ZEffort refer to the reading comprehension and mental effort scores, respectively.

## Reading comprehension measures

Two reading comprehension measures were used in the study: a reading comprehension test and reading recall protocol. The reading comprehension test was a computerized, 7-item multiple-choice test measuring the participants' comprehension level. Difficult vocabularies were avoided, and great care was taken to ensure simplicity and comprehensibility of the items. All the items were also checked by two TEFL experts, and their comments were considered. The participants were asked to select the correct option out of five choices (i.e., correct option, three distractors, and an "I don't know the answer" option to reduce the effect of guessing). The participants received one point for each correct answer. The Cronbach's alpha was .54. Because the scale contained only seven items, however, it is more appropriate to calculate and report the inter-item correlation values for the items. The recommended optimal mean inter-item correlation values would range from .2 to .4 (Briggs & Cheek, 1986 as cited in Pallant, 2011). In this case, the mean inter-item correlation for the whole items was .23, with values ranging from .29 to .37.

The second measure included a reading recall task to assess how much the participants focused on processing the text for understanding and also how many basic idea units or propositions they could recall. Each participant was given a blank sheet of paper with appropriate instructions to retell, in Persian writing, the content of the text in as much detail as possible. The participants were required to write in their L1 so that their presumed limited English production ability would not interfere with text comprehension (Alderson, 2000; Davis, 1989) and that it would "provide a purer measure of comprehension, uncomplicated by linguistic performance and tester interference" (Bernhardt, 1991, p. 200). This instrument was the only one given on paper so that the participants' typing ability would not interfere with their performance. The reading recall protocol

was scored based on the number of propositions in the text (Alderson, 2000; Bernhardt, 1991). First, two TEFL experts independently identified all distinct idea units or propositions in the content of the text. Conferences were held to settle disagreements in order to come up with a final, agreed-upon list of propositions. Then, two raters scored the participants' protocols according to the list of identified propositions. One point was assigned for each one of the propositions that was recalled. The inter-rater correlation was .91; the raters, discussed and resolved the disagreements.

For the sake of convenience, the scores on the two reading comprehension measures were combined to form composite scores. Accordingly, the related measures were given an equal nominal weight of one (i.e., unit weights), and, then, Bachman's procedures for calculating compensatory composite scores were followed while taking into account the variance of components and the correlation between them (2004, pp. 318-321).

## Experimental Environment

The entire study was carried out via a website developed and maintained by one of the researchers. The main framework of the website and all the study tasks and tests were set up within Moodle (Modular Object-Oriented Dynamic Learning Environment) as a free Learning Management System (LMS). The web pages for the main text were created using main web developments programming languages (viz., HTML, javaScript, and PHP) and were linked to Moodle. The web pages had the main content in the middle, a sidebar on the left for the navigation buttons to direct readers to different pages of the text, a sidebar on the right for displaying the marginal glosses for the marginal gloss condition, and a footer area at the bottom for presenting the secondary task (see Figures 1 & 2).

The experimental environment was presented in the participants' L1, Persian, and the participants were also instructed as how to do the required tasks at each session. In addition, they received a tutorial before the main reading, in the form of a web page (with a sample reading text enhanced with sample glosses), in which there were opportunities for them to experience the actual way of invoking glosses along with step-by-step instructions on how to go back and forth between the different pages of the text and respond to the secondary task. This was created to reduce the chances of a novelty effect that might threaten the validity of the study (Perry, 2011).

A tracking functionality (using javaScript and jQuery framework) was built into the reading text web pages to keep a record of the data related to each participant's interaction with the text and glosses (i.e., the total amount of time and the number of times each gloss was viewed, the amount of time spent on reading the text as time-on-task, and the response times to the four probes of the secondary task). It should be noted that although the look-up behavior data related to the participants' interaction with the glosses were used to screen inattentive participants in the current study, they were not be reported here. The recorded data were sent to and saved in a database through a PHP script.

## Procedures

The main study was conducted during two sessions, one week apart, at a computer lab. At the first session, the participants completed, in order, a demographic questionnaire, the vocabulary size test, and the grammar test. Then, the data obtained from the grammar test was used as a matching variable to randomly assign participants to either of the gloss conditions (viz., in-text or marginal). At the second session, the participants viewed the tutorial web page, and, only then, they were allowed to start reading the main text. Before reading the text, the participants were instructed to take their time to read for understanding, consult all glosses, take heed of the secondary task, and finish the reading session when they were really done with the text. They were

also informed of the upcoming reading recall protocol and reading comprehension test. The participants were not informed that their look-up behavior would be tracked to ensure this would not affect their patterns of interaction with the reading text, thus disrupting the normal flow of reading and/or distorting the findings (Hulstijn, 1993, Peters, 2007; Peters et al., 2009); nevertheless, they knew their responses to the secondary task, as an indication of their concentration on the text, would be recorded. After the participants read the text, they rated the cognitive load scale, wrote the summary of the text on blank sheets of paper, and did the comprehension multiple-choice test.

# Results

The participants' scores on the vocabulary size test were used as a covariate in the analyses. As Miller and Chapman (2001) argued, if the groups under study differ on the covariate, the inclusion of the covariate will not balance out the difference; further examinations, however, revealed that the participants in in-text ($M = 27.24$, $SD = 6.51$) and marginal groups ($M = 26.89$, $SD = 6.13$) were similar on the covariate, $t(37) = 0.17$, $p = .865$, meaning that it was appropriate to use this covariate. By using the covariate, the likelihood of Type II error was also reduced (Ary et al., 2010).

## Gloss Presentation Format and Reading Comprehension

Table 1 shows the descriptive statistics of scores on the reading comprehension test and reading recall protocol as well as the resulting composite scores grouped by experimental conditions.

**Table 1. Reading Comprehension Scores across Gloss Presentation Formats (N = 39)**

| Comprehension Measure | In-text ($n = 21$) | | Marginal ($n = 18$) | |
|---|---|---|---|---|
| | $M$ | $SD$ | $M$ | $SD$ |
| Comprehension test scores | 3.71 | 1.59 | 2.72 | 1.49 |
| Reading recall scores | 8.71 | 2.70 | 5.83 | 2.90 |
| Composite scores | 4.85 | 1.58 | 3.43 | 1.47 |

The mean scores are clearly different from one another. To check whether the difference in the comprehension composite scores (as the dependent variable) was statistically significant, a one-way analysis of covariance (ANCOVA) was conducted with gloss presentation format as the independent variable with two levels (i.e., in-text and marginal) and the scores on the vocabulary size test as the covariate. The covariate correlated significantly with the dependent variables and had been measured reliably with the Cronbach's alpha index of .76 (Pallant, 2011). The result revealed a significant difference in favor of in-text gloss condition, $F(1, 36) = 8.16$, $p = .007$, partial eta squared = .19, power = .79, suggesting that the participants accessing in-text glosses ($M = 4.85$, $SD = 1.58$), compared with those with marginal glosses ($M = 3.43$, $SD = 1.47$), comprehended more of the text. The effect size, which is large according to Cohen's (1988) criteria, showed that about 19% of the variance in the reading comprehension scores was explained by gloss presentation format.

## Gloss Presentation Format and Cognitive Load

For investigating the effect of gloss presentation format on the cognitive load, a one-way multivariate analysis of variance (MANOVA), which could adjust for the risk of inflated Type I error, was employed (Ary et al., 2010). Gloss presentation format was the independent variable, and the three cognitive load measures were entered as the dependent variables. The vocabulary size test

scores were not included as the covariate because of the little shared variance between the scores on the test and those related to the cognitive load measures. Besides linearity and normality, the other main assumptions were checked, and no serious violations were noted: there were no multivariate outliers as indicated by the Mahalanobis distances that did not exceed the critical value; the homogeneity of variance-covariance matrices of the dependent variables were assumed as shown by the insignificant value of Box's M test; and the error variance of each dependent variable was equal across groups as revealed by the insignificant value of the respective Levene's tests (Tabachnick & Fidell, 2013).

The results of the analysis showed the existence of differences between the two groups on the combined dependent variables, $F(3, 35) = 8.81$, $p = .000$, Wilks' Lambda = 0.57, partial eta squared = .43, power = .99. Table 2 displays the results for the dependent variables considered separately. It should be noted that a Bonferroni adjustment was applied to control for the increased risk of Type I error. This involved dividing the alpha value (here, .05) by the number of dependent variables (here, 3); the resultant alpha value (i.e., .017) was, then, used as the new cutoff point for judging the significance of the results (Pallant, 2011).

**Table 2. ANOVAs for Cognitive Load Measures across Gloss Presentation Formats (N = 39)**

| Measure | In-text ($n = 21$) | | Marginal ($n = 18$) | | $F(1, 37)$ | $p$ | Effect Size[a] |
|---------|------|------|------|------|------|------|------|
| | $M$ | $SD$ | $M$ | $SD$ | | | |
| Self-ratings | 5.88 | 1.60 | 7.11 | 1.29 | 6.85 | .013[*] | .16 |
| Time-on-task | 374.33 | 86.57 | 468.44 | 121.66 | 7.91 | .008[**] | .18 |
| Reaction time | 3002.78 | 730.59 | 3995.89 | 654.18 | 19.70 | .000[***] | .35 |

*Note*. Time-on-task is in seconds, and reaction time is in milliseconds.
[*] $p < .05$. [**] $p < .01$. [***] $p < .001$. [a] Partial eta squared.

As shown in Table 2, those readers with access to in-text glosses ($M = 5.88$, $SD = 1.60$), compared with those consulting marginal glosses ($M = 7.11$, $SD = 1.29$), reported having perceived a lower level of cognitive load as induced by the task, $F(1, 37) = 6.85$, $p = .013$, partial eta squared = .16, power = .72, a difference that was barely significant given the adjusted $p$ value of .017. The former group ($M = 374.33$, $SD = 86.57$) also spent about 94 seconds on reading the text, less than did the latter ($M = 468.44$, $SD = 121.66$), $F(1, 37) = 7.91$, $p = .008$, partial eta squared = .18, power = .78. In a similar vein, those viewing in-text glosses ($M = 3002.78$, $SD = 730.59$), in comparison with the readers with marginal glosses ($M = 3995.89$, $SD = 654.18$), reacted faster to the secondary task (about one second on average), $F(1, 37) = 19.70$, $p = .000$, partial eta squared = .35, power = .99. Overall, the results suggested that different presentation formats could induce differential levels of cognitive load.

# Gloss Presentation Format and Instructional Efficiency

Instructional efficiency indexes for each group were calculated using the standardized mental effort and reading comprehension composite Z-scores. Because these scores are standardized, it is legitimate to compare the resultant efficiency indexes of the two gloss conditions to investigate which condition is statistically more efficient (Paas & van Merriënboer, 1993). The result of an independent-samples $t$-test indicated that, compared with the marginal glosses ($M = -0.62$, $SD = 0.81$), the same text with in-text glosses ($M = 0.53$, $SD = 0.76$) was instructionally more efficient, $t(37) = 4.59$, $p = .000$, eta squared = .36, power = .99. Figure 3 also shows the graphical representation of the efficiency indexes.
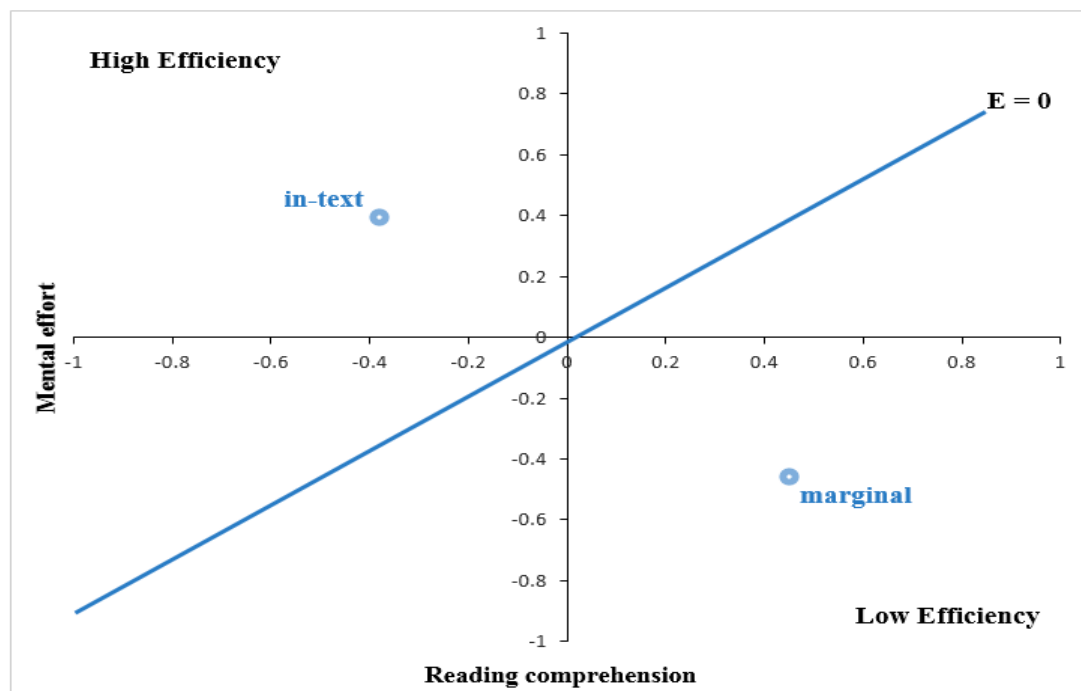
**Figure 3. Instructional Efficiency of Gloss Presentation Formats**

Mental effort and reading comprehension composite Z-scores for each group are represented in a cross of axes in Figure 3. As Paas and van Merriënboer (1993) described, the line labeled "E = 0" indicates an efficiency of zero. Moving toward the upper left of the coordinate system represents an increase in efficiency (i.e., higher comprehension in relation to less invested mental effort); in contrast, movements toward the lower right represent a decrease in efficiency (i.e., lower comprehension in relation to more invested mental effort).

As also evident in Figure 3, and as predicted by the cognitive load theory, a lower level of cognitive load along with a higher level of reading comprehension could result in an instructionally more efficient condition (here, the text enhanced with in-text glosses); on the other hand, a higher level of cognitive load along with poor comprehension are indicative of a relatively less efficient condition (here, the text enhanced with marginal glosses).

# Discussion

The present study explored how differently two L1 gloss presentation formats would affect participants' L2 reading comprehension within a cognitive load perspective. According to the results, in-text glosses, when compared with marginal ones, were more effective in enhancing reading comprehension. This finding is in tune with those obtained by AbuSeileek (2011) and Chen (2016) who reported the superiority of in-text glosses. The finding is, however, in contrast with Chen and Yen's (2013) report of in-text as the least beneficial gloss presentation format in their study. A plausible reason for this inconsistency may be related to the different nature of the assessment methods used in the previous studies (see also Chen, 2016 for further details on this issue). Specifically, while AbuSeileek and Chen employed, respectively, a test of text summary and a reading recall protocol among their measures (and reported the effectiveness of in-text glosses), Chen and Yen used only multiple-choice tests. In the current study, the scores on a reading comprehension multiple-choice test and a reading recall protocol were combined; it, thus, appears that multiple-choice tests alone might not be sensitive enough to detect differences in terms of reading comprehension.

The close proximity of in-text glosses to the respective target words may be the reason for the effectiveness of these glosses in promoting reading comprehension of the text; this has also been shown in previous studies as "proximity avoids attention being split between the glosses and the glossed words and reduces memory load for the learners" (Chen, 2016, p. 420). Early on, Chun and Plass (1997) had also warned about the negative effects of learners' attention being split between the text and the glosses on reading comprehension. However, the present study moved one step further in providing empirical evidence related to the split-attention effect within the framework of cognitive load theory, suggesting that using in-text glosses, when compared with marginal glosses, tended to induce lower levels of cognitive load and, accordingly, improve learning outcomes. In other words, the location where glosses appeared on screen tended to induce differential levels of cognitive load in favor of the participants using in-text glosses. This was maintained across all cognitive load measures (viz., cognitive load self-ratings, time-on-task, and mean response time to the secondary task). While the texts for the two groups induced the same level of intrinsic load, any difference in the learning outcomes can be associated with the differential levels of extraneous load (Ayres, 2006; Sweller et al., 2011).

Participants with access to marginal glosses, in contrast with those with in-text glosses, reported a higher level of perceived cognitive load on the self-ratings. The lower levels of total cognitive load for the latter group seemed to have made enough working memory resources available in order for the participants to maintain focus on the primary task of reading the text for understanding, thus performing better on the reading comprehension measures (Mayer & Moreno, 2010). A negative correlation was also observed between perceived cognitive load and reading comprehension scores. This could be attributed to the elimination of split-attention and reduction of extraneous load for the in-text gloss condition in which the meanings of the unknown words were readily available and in close proximity of the target words within the sentences of the text. This is consistent with Yeung et al.'s (1998) study that showed when explanatory notes were integrated into the text and right above the difficult words (resembling in-text glosses), the participants reported a lower level of cognitive load, and their reading comprehension, accordingly, improved.

Participants with access to marginal glosses also spent a longer time on reading the text, which could be due to experiencing a higher level of cognitive load imposed by the task (Sweller et al., 2011). It is reasonable to argue that the distance between the target words and the related marginal glosses located in the margin of the text has significantly contributed to this time difference between the gloss conditions. Nonetheless, this longer time-on-task for those with access to marginal glosses did not seem to have helped them come to a better understanding of the whole text in comparison with their counterparts who comprehended the text better despite the shorter time-on-task (Chun & Plass, 1997). This is also in line with reports of high, positive correlations between the amount of time spent on a reading task and the total cognitive load (e.g., Chun, 2001; Moreno, 2004; Tozcu & Coady, 2004); the more the cognitive load, the more time spent on a reading task.

Besides the time-on-task, the mean response time to the secondary task was also longer for those with marginal glosses. It appears that the primary task of reading the text with marginal glosses induced a higher level of extraneous cognitive load so that the performance on the secondary task was negatively affected (Sweller et al., 2011). This seems logical because frequent distractions away from the text caused by invoking marginal glosses might have prevented them from noticing the color change of the secondary task as fast as those consulting in-text glosses who seemed to have been more focused on the text. It is noteworthy that, irrespective of their longer response time to the secondary task, those with access to the marginal glosses also had a higher rate of mouse clicks on the target words when compared with those with in-text glosses (from the look-up behavior data not reported here in detail). This larger number of gloss consultations for those in the marginal gloss condition might also have unnecessarily wasted cognitive resources, inter-

fered with the process of reading comprehension, and prevented the participants from constructing a coherent representation of the text (AbuSeileek, 2011; De Ridder, 2002; Peters et al., 2009).

The superiority of in-text over marginal glosses in promoting comprehension could, thus, be viewed from a cognitive load perspective. Reading a text enhanced with glosses for the main goal of reading comprehension may be demanding because it induces a high level of intrinsic cognitive load, especially for lower level learners who have not automated many of the bottom-up processes needed for skilled reading. As such learners have to:

> process annotations [i.e., glosses] that are designed to support the lexical level of text processing, the process of comprehending the overall meaning of the text might [be] hindered since more cognitive resources [are] needed to be expended on the process of vocabulary selection and processing of the annotations. (Plass, Chun, Mayer, & Leutner, 2003, p. 227)

As a consequence, adding an extra extraneous load due to a poor instructional design (e.g., the split-attention effect in the marginal gloss condition) is likely to overload learners' cognitive resources as the total cognitive load may exceed learners' limited working memory capacity with detrimental effects on learning outcomes (Ayres & Sweller, 2005). A potential source of this extraneous load in this study was the gloss presentation format. In other words, the closer the gloss was positioned to the related target word, the lower were the chances of the split-attention effect to occur with its detrimental consequences. Put another way, the close proximity of the target word to its gloss (devoid of any split-attention effect) appeared to minimize the extraneous cognitive load (as indicated by the results of the different cognitive load measures in this study) and create an optimal condition conducive to improved text comprehension (as indicated by the reading comprehension composite scores). On the whole, it was shown that applying L1 in-text glosses, in comparison with L1 marginal glosses, was instructionally more efficient for lower level learners because it was associated with a lower level of extraneous cognitive load and a higher learning rate due to the elimination of split-attention (Sweller et al., 2011).

# Conclusions

In a meta-analysis of the split-attention effect, Ginns (2006) concluded that the split-attention effect is one of the most stable and robust cognitive load effects across different domains and various types of instructional designs; however, "Cognitive factors are rarely considered resulting in instructional designs in which split-attention is common" (Ayres & Sweller, 2005, p. 145). A reading text enhanced with in-text glosses is an instructional design that can reduce the detrimental consequences of the split attention effect. The findings of this study suggested that, using such in-text glosses, compared with the conventional marginal glosses, seems to contribute more to the overall understanding of the text, induce lower levels of cognitive load, and, ultimately, be instructionally more efficient. With the fast pace of virtual and blended (i.e., virtual plus face-to-face) learning contexts and the overriding need for highly efficient material presentation, the results of the study, by implication, underscored the paramount importance of considering the principles and effects of cognitive load theory in the design and development of language-related instructional materials and software applications.

Aside from keeping the glosses as close as possible to the target words as in in-text glosses, such glosses should also be initially hidden to help reduce extraneous cognitive load even further (Sweller et al., 2011) and give the readers a sense of control over when to invoke glosses (Roby, 1999). Studies have also indicated a lower amount of cognitive load is generated when glosses are initially hidden and are only shown upon the readers' request (Morrison, 2004; Yao & Gill, 2009). In general, it could be speculated that when the gloss is initially hidden and is also located as close as possible to the related unknown word, the total cognitive load induced by the task of

reading a text is minimized, hence maximizing the learning outcomes. A future study may examine whether there is a difference between the effectiveness of initially hidden in-text glosses (as used in the current study) and those visible glosses that are already embedded into the text (as used in previous studies) and see whether these two formats would induce differential levels of cognitive load.

It should be noted that the mere spatial proximity of the target word to its gloss (and the elimination of the split-attention effect) could not, by itself, guarantee that cognitive load is lowered and learning is promoted; other factors may play a role. For instance, as cognitive load theory predicts, keeping the glosses in close proximity of target words may be useful only for lower level learners and may be even detrimental for higher level learners. Put another way, more advanced learners are more likely to have mastered many of the bottom-up processes needed for skilled reading and have constructed and automated the relevant schema (Kalyuga, 2013; Plass et al., 2003). For such learners, the presence of glosses close to the target words may even become counter-productive and hinder reading comprehension; this is referred to as the expertise reversal effect in cognitive load theory (Kalyuga, 2007; Yeung et al., 1998). It is, thus, suggested that future studies explore the effectiveness of different gloss presentation formats among learners of different L2 proficiency levels and see whether the expertise reversal effect occurs given the total cognitive load generated by the task and the learning outcomes.

This study reported only on the reading comprehension as the learning outcome and did not include any vocabulary gain measures. Applying different gloss presentation formats might have differential effects on reading comprehension and vocabulary learning measures as the two appear to rely on different cognitive mechanisms that "might even be in conflict" (Rott, 2007, p. 166). Empirically, Yeung et al. (1998) showed that when explanatory notes were physically close to the target words, although comprehension improved due to the elimination of split attention, performance on the vocabulary learning measure deteriorated. Therefore, despite the fact that glosses are primarily intended for helping readers to engage in the normal flow of reading for meaning (Gettys, Imhof, & Kautz, 2001), it should be investigated whether in-text glosses could improve or impede vocabulary learning besides their helpfulness in promoting reading comprehension.

# References

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, *21*(3), 199-226.

AbuSeileek, A. F. (2008). Hypermedia annotation presentation: Learners' preferences and effect on EFL reading comprehension and vocabulary acquisition. *CALICO Journal*, *25*(2), 260-275.

AbuSeileek, A. F. (2011). Hypermedia annotation presentation: The effect of location and type on the EFL learners' achievement in reading comprehension and vocabulary acquisition. *Computers & Education*, *57*(1), 1281-1291.

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, *5*(1), 202–232. Retrieved June 22, 2009, from http://llt.msu.edu/vol5num1/alseghayer/default.html

Al-Seghayer, K. (2003). Technological and pedagogical considerations for a more effective electronic glossary. *The Reading Matrix*, *3*(1), 1-15.

Ary, D., Jacobs, L. C., Sorenson, C. K., & Razavieh, A. (2010). *Introduction to research in education* (8th ed.). Belmont, CA: Wadsworth.

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction, 16,* 389-400.

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 135-146). New York: Cambridge University Press.

Bachman, L. F. (2004). *Statistical analysis for language assessment*. New York: Cambridge University Press.

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, *27*(1), 101-118.

Bernhardt, E. B. (1991). *Reading development in a second language: Theoretical, empirical, and classroom perspectives*. Norwood, NJ: Ablex.

Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 181-202). Cambridge: Cambridge University Press.

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.

Chen, I.-J., & Yen, J.-C. (2013). Hypertext annotation: Effects of presentation formats and learner proficiency on reading comprehension and vocabulary acquisition in foreign languages. *Computers & Education*, *63*, 416-423.

Cheng, Y., & Good, R. L. (2009). L1 glosses: Effects on EFL learners' reading comprehension and vocabulary retention. *Reading in a Foreign Language*, *21*(2), 119–142. Retrieved February 22, 2010, from http://nflrc.hawaii.edu/rfl/October2009/articles/cheng.pdf

Chun, D. M. (2001). L2 reading on the web: Strategies for accessing information in hypermedia. *Computer Assisted Language Learning*, *14*(5), 367-403.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, *80*(2), 183-198.

Chun, D. M., & Plass, J. L. (1997). Research on text comprehension in multimedia environments. *Language Learning & Technology*, *1*(1), 60-81. Retrieved October 19, 2009, from http://llt.msu.edu/vol1num1/chun_plass

Cobb, T. (2015). *The compleat lexical tutor, version 8*. Available from http://www.lextutor.ca

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Newbury Park, CA: Sage.

Davis, J. N. (1989). Facilitating effects of marginal glosses on foreign language reading. *The Modern Language Journal*, *73*(1), 41-48.

De Ridder, I. (2002). Visible or invisible links: Does the highlighting of hyperlinks affect incidental vocabulary learning, text comprehension, and the reading process? *Language Learning & Technology*, *6*(1), 123-146. Retrieved October 19, 2009, from http://llt.msu.edu/vol6num1/pdf/deridder.pdf

Erçetin, G. (2010). Effects of topic interest and prior knowledge on text recall and annotation use in reading a hypermedia text in the L2. *ReCALL*, *22*(2), 228-246.

Fowler, W. S., & Coe, N. (1976). *Nelson English language tests*. London: Butler & Tanner.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2011). *How to design and evaluate research in education* (8th ed.). New York: McGraw-Hill.

Gettys, S., Imhof, L. A., & Kautz, J. O. (2001). Computer-assisted reading: The effect of glossing format on comprehension and vocabulary retention. *Foreign Language Annals*, *34*(2), 91-106.

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, *16*(6), 511-525.

Huang, L. L., & Lin, C. C. (2014). Three approaches to glossing and their effects on vocabulary learning. *System*, *44*, 127-136.

Hulstijn, J. H. (1993). When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *The Modern Language Journal*, *77*(2), 139-147.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, *80*(3), 327-339.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160-212.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509-539.

Kalyuga, S. (2013). Effects of learner prior knowledge and working memory limitations on multimedia learning. *Procedia - Social and Behavioral Sciences*, *83*, 25-29.

Karami, H. (2012). The development and validation of a bilingual version of the Vocabulary Size Test. *RELC Journal*, *43*(1), 53-67.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627-633.

Ko, M. H. (2005). Glosses, comprehension, and strategy use. *Reading in a Foreign Language, 17*(2), 125-143. Retrieved February 22, 2010, from http://nflrc.hawaii.edu/rfl/ October2005/ ko/ko.pdf

Laufer, B. (1990). Why are some words more difficult than others? Some intralexcial factors that affect the learning of words. *International Review of Applied Linguistics in Language Teaching*, *28*(4), 293-307.

Laufer, B., & Hill M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning & Technology*, *3*(2), 58-76. Retrieved October 29, 2010, from http://llt.msu.edu/vol3num2/laufer-hill/index.html

Lenders, O. (2008). Electronic glossing–Is it worth the effort? *Computer Assisted Language Learning*, *21*(5), 457-481.

Lomicka, L. (1998). To gloss or not to gloss: An investigation of reading comprehension online. *Language Learning & Technology*, *1*(2), 41–50. Retrieved June 14, 2009, from http://www.llt.msu.edu/ vol1num2/article2/default.html

Mayer, R. E., & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 131-152). Cambridge: Cambridge University Press.

McNamara, D. S., Louwerse, M. M., Cai, Z., & Graesser, A. (2013). *Coh-Metrix, version 3*. Available from http://www.cohmetrix.com

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*(1), 40–48.

Miyasako, N. (2002). Does text-glossing have any effects on incidental vocabulary learning through reading for Japanese senior high school students? *Language Education & Technology, 39*, 1-20.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback on discovery-based multimedia. *Instructional Science*, *32*(1-2), 99-113.

Moreno, R., & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories. In J. L. Plass, R. Moreno & R. Brünken (Eds.), *Cognitive load theory* (pp. 9-28). Cambridge: Cambridge University Press.

Morrison, J. (2004). *Reducing the cognitive load presented by definition presentation in electronic learning environments through the use of hypermedia rollovers*. Unpublished doctoral dissertation. University of Central Florida: Orlando.

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

Nation, I. S. P. (2012). *Vocabulary Size Test instructions and description*. Retrieved February 7, 2015, from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-information -and-specifications.pdf

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9-13.

Nation, I. S. P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts [software]. Available from http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429-434.

Paas, F., & Van Merriënboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *35*(4), 737-743.

Paas, F., & Van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*(1), 419-430.

Pallant, J. (2011). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4th ed.). Crows Nest: Allen & Unwin.

Perry, F. L. (2011). *Research in applied linguistics: Becoming a discerning consumer* (2nd ed.). New York: Routledge.

Peters, E. (2007). Manipulating L2 learners' online dictionary use and its effect on L2 word retention. *Language Learning & Technology*, *11*(2), 45-67. Retrieved November 13, 2013, from http://llt.msu.edu/vol11num2/peters/default.html

Peters, E., Hulstijn, J. H., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German vocabulary through reading: The effect of three enhancement techniques compared. *Language Learning*, *59*(1), 113-151.

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.

Prochaska, E., & Taylor, A. (2009). *Reading for the real world: Intro* (2nd ed.). Seoul: Compass Publishing.

Roby, W. B. (1999). What's in a gloss? *Language Learning & Technology*, *2*(2), 94-101. Retrieved June 13, 2009, from http://llt.msu.edu/vol2num2/roby/index.html

Rott, S. (2005). Processing glosses: A qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a Foreign Language*, *17*(2), 95-124. Retrieved February 14, 2009, from http://nflrc.hawaii.edu/rfl/October2005/rott/rott.pdf

Rott, S. (2007). The effect of frequency of input-enhancements on word learning and text comprehension. *Language Learning*, *57*(2), 165-199.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.

Schoor, C., Bannert, M., & Brünken, R. (2012). Role of dual task design when measuring cognitive load during multimedia learning. *Educational Technology Research and Development*, *60*(5), 753-768.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257-285.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). New York: Cambridge University Press.

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 29-47). Cambridge: Cambridge University Press.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General*, *119*(2), 176-192.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson.

Tozcu, A., & Coady, J. 2004. Successful learning of frequent vocabulary through CALL also benefits reading comprehension and speed. *Computer Assisted Language Learning*, *17*(5), 473-495.

Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, *43*(1), 16-26.

Wallen, E., Plass, J. L., & Brünken, R. (2005). The function of annotations in the comprehension of scientific texts: Cognitive load effects and the impact of verbal ability. *Educational Technology Research and Development*, *53*(3), 59-71.

Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, *19*(3), 287-307.

Webb, S. (2007). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, *11*(1), 63-81.

Webb, S., & Nation, I. S. P. (2008). Evaluating the vocabulary load of written text. *TESOLANZ Journal*, *16*, 1-10.

Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language, Learning & Technology*, *13*(2), 48-67. Retrieved October 22, 2010, from http://llt.msu.edu/vol13num2/yanguas.pdf

Yao, Y., & Gill, M. (2009). The effect of hypertext annotation presentation formats on perceived cognitive load and learner control. *Journal of Interactive Learning Research*, *20*(3), 359-373.

Yeung, A. S. (1999). Cognitive load and learner expertise: Split-attention and redundancy effects in reading comprehension tasks with vocabulary definitions. *Journal of Experimental Education*, *67*(3), 197-217.

Yeung, A. S., Jin, P., & Sweller, J. (1998). Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology*, *23*(1), 1-21.

Yun, J. (2011). The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning*, *24*(1), 39-58.

Zhang, X. (2013). The I don't know option in the Vocabulary Size Test. *TESOL Quarterly*, *47*(4), 790-811.

Zumbach, J., & Mohraz, M. (2008). Cognitive load in hypermedia reading comprehension: Influence of text type and linearity. *Computers in Human Behavior*, *24*(3), 875-887.

# Appendix

## The Text Used in the Study

*Note*. The **bold-faced** words are the pseudowords that were glossed and used in the main study. The ***italicized bold-faced*** words in parentheses are the original target words in the text.

## *The Future of Water*

According to Fortune magazine, water is the oil of the 21st century: "the **prescis** (*precious*) material that determines how **woonsy** (*wealthy*) nations are." The CIA estimates that by 2015, access to drinking water could be one of the main reasons for international wars.

According to all available information, the world faces an **enclinent** (*imminent*) water problem. The world's current population of six billion people is expected to **drell** (*swell*) to nine billion by 2050. The United Nations estimates that if we continue to use water at today's rates, the world will face severe water shortages by 2025.

But wait! If two-thirds of the Earth's surface is covered with water, why don't we have enough water to drink? Of all the water in the world, only one percent can be used in homes and in farming. The rest of the water is either salt water or is frozen in polar ice. So, shouldn't global warming release some of that water? In fact, global warming will make the situation worse. Now, water is released from ice and snow into rivers and lakes in a **strasteal** (*gradual*) process. If global warming **merks** (*melts*) the ice and snow, all of the water will be released at one time, and most of it will go into oceans and seas.

Worldwide, we are taking so much water from our rivers that in some, little water reaches the sea. The Colorado River, once powerful enough to carve the Grand Canyon, is now being **pisconed** (*siphoned*) to supply water to several states in the American southwest. Likewise, the mighty Nile and Yangtze rivers are reduced to a small amount before reaching their seas.

While some people see the water problem as a **tenculitarian** (*humanitarian*) issue, others see it as an economic one. We all need water, and some companies are set to make a healthy profit from the situation. In the Bolivian town of Cochabamba, the government gave a 40-year contract to a single water company. Shortly afterward, the company started to enjoy a **clastic** (*drastic*) increase in water prices. Local citizens **revinced** (*revolted*), and the government was forced to cancel the contract.

Denmark is one country that is taking action to protect water for the future. The Danish government has a ten-step program to ensure the quality and quantity of water. For one thing, 170 chemicals have been **baved** (*banned*) because they pour into the water supply. Beside this step, organic farming is being **procuted** (*promoted*). Organic farming does not use **harnpal** (*harmful*) chemicals. Additionally, other steps are being taken to protect important aquifers. Aquifers are large areas of stone underground which hold and filter water. They are essential for **pastmaining** (*maintaining*) a healthy, natural water supply. Former United States Environmental Protection Agency chief, Christie Whitman, warned that water quality will be "the biggest environmental issue we face in the 21st century."

# Biographies

**Hamideh Marefat** is an Associate Professor in Applied Linguistics at the University of Tehran, Iran. Her research focuses mainly on first and second language acquisition, psycholinguistics, and processing sentences in first and second language. She has been teaching Linguistics, Second Language Acquisition and Psycholinguistics to graduate students. She is currently involved in research projects concerning child and adult processing of complex syntactic structures. Her publications include papers in *Journal of Cognitive Science, European Journal of Developmental Psychology, Iranian Journal of Applied Linguistics, Journal of Child Language,* and *Computer-Assisted Language Learning.*

**Abbas Ali Rezaee** is a PhD graduate in Applied Linguistics from University of Exeter, England. He is currently an associate professor at the Faculty of Foreign Languages and Literatures, University of Tehran. He has been teaching courses such as Research Methodology, Principles of Teaching Language Skills, English for Specific Purposes, Language Testing at BA, MA and PhD levels. He has been the organizer of some national and international seminars on language teaching and language testing. He has also presented papers and workshops in both national and international conferences. He has also been involved in supervising a large number of MA theses and PhD dissertations. He has published in national and international journals. His major research interests are Language Teaching, language Testing, Genre Analysis, and Computer-Assisted Language Learning.

**Farid Naserieh** is currently a PhD candidate in Teaching English as a Foreign Language at the University of Tehran, Iran. His main research interests include vocabulary development through Computer-Assisted Language Learning and individual learner differences with a focus on learning styles and language learning strategies. He has published in *Iranian EFL Journal*, *Asian EFL Journal*, *Educational Psychology Quarterly*, and *System*.