6-1-2017

Does Adding Pictures to Glosses Enhance Vocabulary Uptake from Reading?

Frank Boers
fboers@uwo.ca

Paul Warren
Victoria University of Wellington

Lin He
Xi'An International Studies University, China

Julie Deconinck
Vrije Universiteit Brussel, Belgium

Does Adding Pictures to Glosses Enhance Vocabulary Uptake from Reading?

Frank Boers, Paul Warren, Lin He and Julie Deconinck

Abstract

This article reports three trials of a pen-and-paper experiment where adult L2 learners'
recollection of glossed words was tested after they had read a text with or without
pictures included in the glosses. Unlike previous studies in which a superiority of
multimodal glosses over text-only glosses was claimed, the experiment furnished no
evidence that the addition of pictures helped the learners to retain the glossed words'
form-meaning association any better than providing glosses containing only verbal
explanations. When learners were prompted to recall of the written form of the words,
the gloss condition without pictures in fact led to the better performance. The results
suggest that the provision of pictures alongside textual information to elucidate the
meaning of novel words may reduce the amount of attention that L2 readers give to the
form of these words.

Keywords: vocabulary, glosses, multimodality, recall, attention.

1. Introduction

Glosses or annotations that accompany text to clarify the meaning of unfamiliar
words not only facilitate text comprehension but can also promote learners' acquisition
of the glossed words (Hulstijn, Hollander & Greidanus, 1996). Glosses draw attention to
words that might otherwise be overlooked by the learner, they reiterate the word of
interest within the gloss, and they ensure adequate interpretation — provided the gloss

is sufficiently informative and the information is presented in a way that is accessible to the learner. Many studies in this area have compared the benefits of different kinds of glosses. For example, one line of research compares the benefits of annotations in L1 (translations) and in L2 (definitions) (Jacobs, Dufon & Hong, 1994; Ko, 2012; Yoshii, 2006). Another line evaluates ways of stimulating cognitive engagement with the glosses, for example by incorporating an interpretation challenge in the gloss itself (Boers, 2000; Nagata, 1999; Watanabe, 1997). A third line of research concerns the potential benefits of multimodal glossing, defined here as the combination of textual clarification and pictorial elucidation of word meaning. That is the line of research to which the present article aims to contribute.

Several studies (see below) have reported findings that appear to support the thesis that multimodal information helps L2 readers retain the meaning of glossed words better than textual clarifications alone. The present article first evaluates those findings and the conclusions drawn from them, and points to a need for approximate replication studies to re-assess the benefits of multimodal glosses. This is followed by a report of such a study, the results of which cast doubt on the proclaimed superiority of multimodal glosses over text-only glosses for L2 vocabulary uptake.

2. Literature review

2.1 Studies on the benefits of providing more than one gloss for a word

When pictures are used in printed materials for second language learning such as text books, they are presented to learners together with the textual input, typically on the same page, in a single gloss. It is that co-presentation of words and pictures in printed materials that is the object of the experiment we report further below. However, a fair

amount of the evidence that has been interpreted in favour of adding pictures to verbal clarifications of word meaning comes from research conducted in the context of computer-assisted reading where participants consulted separate textual and pictorial annotations consecutively, by mouse-clicking on highlighted words. Chun and Plass (1996) is an early study of this kind. In their experiment, half of the targeted words only had a textual annotation while the other half had both a pictorial and a textual annotation. In post-tests where the students were asked to match the L2 words with their meaning, the words for which both pictorial and textual annotations were available generated the best scores. While this attests to the advantage of providing a textual annotation in addition to a textual one, the authors did not explore whether this might be due to multiple look-ups (in the case of two available annotations) rather than the multimodality *per se* of the information that was made available for half of the words. In a similar study, Plass, Chun, Mayer and Leutner (1998) showed that participants were indeed more likely to retain the meaning of target words if they consulted two annotations about a word than if they consulted only one.

Jones and Plass (2002) investigated the benefits of word annotations in the context of computer-assisted L2 listening practice. Learners listened (in a self-paced manner) to a text with its transcript appearing on the computer screen. Again, some of the words in the transcript were highlighted as having annotations, accessible by mouse-clicking on them. For some of the learners only one annotation (either text or picture) was available for each of these words, while for other learners both a textual and a pictorial annotation was made available. The students who inspected two annotations outperformed those who were provided with only one in a post-test about the meaning of the L2 words.

Akbulut (2007) is another study where L2 readers could mouse-click on highlighted words to access annotations. Again, the participants' retention of the meaning of the annotated words was found to be better after reading a text in a condition where pictorial annotations were available in addition to textual ones than in a condition where only a textual annotation was available per target word. Although the author explains that the computer software recorded how often participants accessed given annotations, this data is unfortunately not included in the article, and so it is again impossible to tell whether the better post-test performance under the multimodal condition is to be attributed to the multimodality *per se* or rather to the greater number of look-ups prompted by the availability of more than one annotation.

In sum, what this handful of studies suggests is that making more than one annotation available stimulates look-ups. A word whose meaning is looked up twice also receives attention twice, and the amount of attention given to a word is known to be one of the predictors of word learning (e.g., Godfroid, Boers, & Housen, 2013; Godfroid & Schmidtke, 2013; see Schmitt, 2008, for a review that emphasizes the role of attention or "engagement" in vocabulary learning). What the above studies do *not* show, strictly speaking, is that it matters whether the available annotations which stimulate multiple look-ups and thus multiple episodes of engagement with the same word also include pictorial annotations.

2.2 Studies comparing the benefits of multimodal and text-only glosses for L2 vocabulary uptake

Let us now turn to studies where pictures and textual information were co-presented together in a single gloss, which is the reading condition that the present

article aims to re-evaluate. In Kost, Fost and Lenzini (1999), L2 learners read a short narrative text with marginal glosses for 14 unfamiliar words. Three groups of learners encountered a different version of the glosses: a version with L1 translations, a version with pictures, and a version with both an L1 translation and a picture in each gloss. In post-tests, the students were presented with the L2 words and asked (a) to provide the L1 translation, (b) to match the word with its corresponding picture, and (c) to match the word with its corresponding L1 translation. No significant differences between the three groups' scores were found in the first test (i.e., where the students were required to supply the meaning of the L2 words). On the picture recognition test the students who had received glosses containing a picture outperformed the students who had received only translations. This is not very surprising as the latter students had not seen those pictures before.[1] On the third test, where students were asked to match the L2 words with their L1 translation, it was the group that had received picture-only glosses that performed most poorly. Again, this is not so surprising, as these students had not seen the translations before. Interestingly, however, the combination of translation and picture in glosses yielded the best scores on this third test, which suggests that multimodal clarifications of word meaning were helpful for the learners in Koss et al.'s experiment.

A few other studies where visuals and verbal clarifications were combined in single glosses were conducted in computer-aided reading contexts again. In Al-Seghayer (2001) seven words in a text came with a definition only, for another seven words this definition was accompanied by a picture, and for yet another seven words the definition was accompanied by a video clip. In the post-test, which gauged students' recognition and recall of the meaning of the target words, the set of words that had been annotated

only by means of a definition received the lowest mean score. It needs to be conceded, though, that it is hard to create sets of target words that are perfectly matched with regard to factors that influence their learnability (Ellis & Beaton, 1993, pp. 560–569; Laufer, 2013). This is a difficulty inherent to a within-participants design on vocabulary learning, to be acknowledged also in connection with some of the aforementioned studies, such as Chun and Plass (1996). Yeh and Wang (2003) used a between-participants design, with three annotation conditions: text only, text plus picture, and text plus picture and an audio-file of the pronunciation of the word. Learning was gauged through a mixture of word association, multiple-choice and cloze tests (but the article provides no details or examples, thus making the study hard to replicate). The authors conclude from the overall post-test scores that the text-plus-picture annotations were the most effective, but this is slightly misleading because there was in fact no significant difference with the scores obtained under the text-only condition. Interestingly, the annotations enhanced with audio-files yielded the poorest post-test scores, which suggests that more is not always better.

Yoshii and Flaitz (2002) also used a between-participants design. Students were assigned to one of three annotation conditions: L2 text only, picture only, or L2 text plus picture. The post-test on 14 words consisted of three parts: (a) matching the L2 word with the corresponding picture; (b) matching the L2 word with the corresponding L2 definition; and (c) explaining the meaning of the given L2 word. As could be expected given considerations of input format – test format congruency, the text-only condition led to the poorest performance on the picture-recognition test, while the picture-only condition led to the poorest performance on the verbal tests. Interestingly, the text-plus-picture condition led to the best scores on the provide-the-meaning test.

Yoshii (2006) used roughly the same materials in a partial replication study, and found again that learners who had been given access to annotations in which pictures were added to verbal information did best on a post-test that prompted them to explain the meaning of the words. A question that needs to be asked, though, is whether those better post-test scores under the multimodal glossing condition in Yoshii and Flaitz (2002) and Yoshii (2006) were entirely due to the multimodal nature of the information. For instance, the verb *dash* was defined in the text-only gloss as 'to move very quickly', in the picture-only gloss it was illustrated by a drawing of a running figure, and in the multimodal gloss both the definition and the picture were presented (Yoshii & Flaitz, 2002, p. 37). The verbal definition in this example leaves the manner of motion underspecified. It is the picture of the running figure that helps to add precision to the verbal explanation by signalling that *dash* does not denote just any fast movement, but typically involves *running*. Conversely, the picture of a running figure alone may be taken by the learner as intended to depict the action of running, but perhaps not necessarily as depicting specifically *fast* running. Crucially, the participants in the experiments were required in the post-test to supply an explanation of the meaning of the word that contains both the notions 'fast' and 'run' in order to obtain full points. This obviously puts participants who were provided with an underspecified gloss — which happens to be the case with both the single-mode glosses — at a disadvantage. It is certainly true that different modes of input can usefully complement one another in case the information provided by each separately is not optimally helpful. Where words fail, visuals may indeed come to the rescue, and vice versa. In this particular case, however, it would have been feasible to define *dash* with more precision in the text-only gloss (e.g., 'to run somewhere very quickly because you are in a hurry', *Macmillan*

English Dictionary for Advanced Learners [2007]). The participants provided with this more precise verbal definition would then have stood a fairer chance in the post-test where they needed to explain the meaning of the word.

Despite such methodological concerns, the evidence reviewed so far is at least corroborative of the proposition that adding pictures to glosses promotes uptake of word meaning. Not all studies have produced such supporting evidence, however. A study that casts doubt on the benefits of combining pictures and verbal explanations in a gloss is an experiment by Acha (2009), where young L2 learners were randomly assigned to one of three gloss conditions: L1 translations only, pictures only, and a combination of L1 translations and pictures. In an immediate and a two-week delayed post-test where the students were prompted to explain the meaning of the glossed words, the translation gloss condition yielded the best scores while the multimodal gloss condition yielded the poorest. These results stand in stark contrast with those reviewed above, and indicate that the question of whether adding visuals to glosses truly enhances word learning is not yet settled.

It is also worth noting that all aforementioned studies measured only receptive word knowledge (e.g., matching the given L2 word with its L1 translation or with a picture, or providing the meaning of the given L2 word prompt).[2] Post-reading tests that focus on receptive knowledge are perfectly justified, of course. After all, glosses are meant first and foremost to aid comprehension. On the other hand, acquiring word knowledge comprises many things, including knowledge of the form of the word. The question this raises is how the addition of pictures to glosses influences learners' uptake of the (written) form of target words. There are actually grounds for expecting a *negative* effect. Barcroft (2015) has demonstrated repeatedly that activities which orient

learners toward aspects of the meaning of words in the initial stage of word learning aid retention of word meaning, but often do so at the cost of retention of word form. Pictorial glosses obviously orient learners toward word meaning — pictures depict referents or concepts; they do not as such direct learners' attention to the (orthographic) form of the word. It is therefore not unlikely that, in the process of dividing their attention between different stimuli (e.g., Craik, Govoni, Naveh-Bejamin, & Anderson, 1996; Mulligan, 1998; Yeung, Jin, & Sweller, 1997), learners will allocate attention to a picture which they might otherwise give to the actual word which the picture is intended to elucidate.

Summing up, the aforementioned studies which claimed positive effects for multimodal glosses are in need of approximate replication for at least two reasons: (a) the benefits of adding pictures to glosses may have been overestimated as a result of methodological choices made by the authors, and (b) the effect of adding pictures on learners' uptake of the *form* of novel words has yet to be addressed.

3. Research questions

The experiment reported here compares the effects of two gloss types: text-only and text-plus-picture. The questions addressed are whether the addition of pictures to textual glosses affects L2 readers' post-reading (a) recognition of the *meaning* of the glossed words, and (b) their recall of the *form* of the glossed words.

4. Method

Like Kost et al. (1999), ours was a pen-and-paper experiment and it adopted a between-participants design. In recognition of the need for multiple replications in order

to avoid premature conclusions (Porte, 2012), we conducted the experiment three times, with different populations of learners.

4.1 Materials and procedure

A narrative text of about 800 words was created for the purpose of this study (see Appendix 1). Six pseudowords (all used as nouns, and referring to concrete referents) were incorporated in the text, each occurring once. The meanings of these pseudowords were explained in marginal glosses (see Appendix 1). The pseudowords were borrowed from Godfroid et al. (2013), where the phonological and orthographic plausibility of the proposed word forms was checked by four native speakers. In addition to the six pseudowords, 12 real words were glossed in order to reduce the risk of participants' becoming suspicious about the pseudowords. Only the pseudowords served as real targets in the experiment, however.

The marginal glosses for the six pseudowords differed according to treatment condition. One group of participants received glosses that only contained textual information (*text-only* glosses), while for the other group the same glosses were complemented with pictures to elucidate the word's meaning (*multimodal* glosses; see Appendix 1). In trials 1 and 2, the textual information in the glosses consisted of a brief English definition of the word. In trial 3, it consisted of a translation equivalent in the participants' first language. The latter adaptation helped to blend the materials in with the participants' regular course work — the students in trial 3 were enrolled in a programme preparing them to become translators.

In all three trials, the reading activity and the tests were integrated in a regular English language lesson.[3] The participants were given 15 minutes to read the text in

silence and were told to expect content questions afterwards. Telling them this was intended to provide an incentive for the participants to read for understanding. They were not told that their recall of vocabulary from the text would also be tested — which is in keeping with the objective to examine incidental rather than deliberate learning (Hulstijn, 2001). A question presented after the post-tests confirmed that it is unlikely that the participants tackled the reading task as an intentional vocabulary learning task. This was a yes / no question asking the participants whether they had expected, while reading the story, that their recollection of words from the text was going to be tested. None of the participants in any of the three trials replied they had expected this.

When the reading was finished, the participants' copies of the text were collected and they were given the first of two tests (Appendix 2). Six items in this test were fill-in-the-blank items meant to elicit the written form of the pseudowords. The participants were encouraged to write what they could remember, even if it was just part of the word or even if they were not sure about it. This form recall test contained three prompts: (a) the sentence from the text where the word was used but with a blank where the target word occurred, (b) the definition that was given in the marginal gloss, and (c) the picture from the multimodal condition. Participants in either gloss condition could thus rely on the prompts they had been exposed to during the reading activity. Although the pictorial prompts in the test were new for the students in the text-only gloss condition, these students did not need to rely on these pictures in the test because the other prompts (which *were* congruent with their reading condition) were available.

Apart from the gap-fill items targeting the six pseudowords, the first test contained six multiple-choice questions, each with three answer options to choose from, about the content of the story. Answering these questions did not require comprehension of any of

the six target words. Whether or not multimodal glosses aid recall of text content more than text-only glosses was not initially set up as a research question for this study [4], but the participants' performance on the content questions was nevertheless considered in our statistical models (see below).

After this first test (on word form and text content) was collected from the participants, they were given a second test (Appendix 3). On the left-hand side of the sheet the textual explanations and the pictures from the marginal glosses were re-presented and on the right-hand side the target words were listed in jumbled order. In addition to the six pseudowords two existing words that were glossed in the reading text were included in this matching test, but responses to these were excluded from the data analysis. The task for the participants was to match each explanation (plus picture) with its corresponding word by connecting them with a line. We shall refer to this second test as a meaning-recognition test for short, although successful responses obviously require recognition of word form as well. The meaning-recognition test was administered after the form-recall test, because otherwise the participants' performance on the form-recall test would have been aided by their seeing the target-word prompts in the meaning-recognition test.

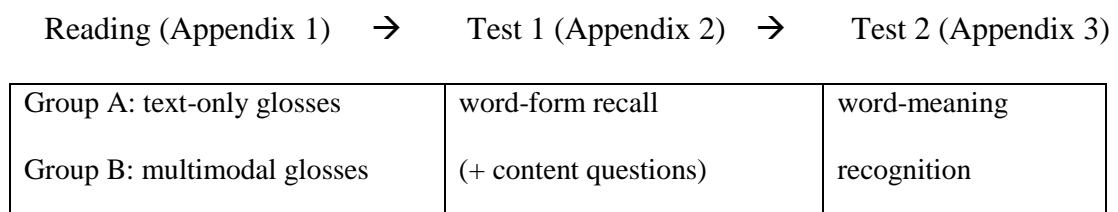Figure 1 sums up the data collection procedure.

Reading (Appendix 1)   →   Test 1 (Appendix 2)   →   Test 2 (Appendix 3)

| Group A: text-only glosses | word-form recall | word-meaning |
| Group B: multimodal glosses | (+ content questions) | recognition |

Figure 1: Data collection procedure

4.2 Scoring and analysis

The meaning-recognition test was scored in a binary fashion for each of the six pseudowords (0=incorrect, 1=correct form-meaning match). The same applies to each of the six multiple-choice questions concerning text content.

Scoring of the form-recall test required a more intricate procedure. Given that each pseudoword occurred just once in the text and once more in the marginal gloss, and given that the students were not told a word-recall test would follow, it is not surprising that these words tended to be poorly recalled in the form-recall test. Participants' attempts typically resulted in only a partial reproduction of the pseudoword (e.g., *banlion* instead of *bandilon*; *redat* for *redaster*; *mat* for *canimat*) and/or misspelled forms (e.g. *bandollion* instead of *bandilon*; *radester* instead of *redaster*; *pinlin* instead of *panipline*). In each of the three experimental trials, two research assistants who were blind to the reading condition to which participants had been assigned gave scores to the participants' responses. We originally envisaged using Barcroft's (2002) scoring protocol for partial knowledge, according to which partially correct responses are awarded a score of 0.25, 0.50 or 0.75. However, a score of .25 corresponds to accurate reproduction of one fourth of the target word, and in many cases the participants in the present study reproduced just the first letter (e.g., *p* for *panipline*). While this is minimal recall, it is nonetheless evidence of a memory trace and thus worth distinguishing from responses that were left blank altogether. We therefore asked the assessors to give scores to responses on a scale from 0 to 1. Partially correct responses were thus given scores between 0.1 (recall of the first letter) and 0.9 points (complete recall apart from one missing letter, a substituted letter or inversion of two letters). Even though the scoring instructions for partial word recall in the range between 0.1 and 0.9 points were

kept vague, the Pearson's correlation coefficient between the scores given by the two assessors was reassuringly high: $r = .98$.[4] For example, the response *red* (for *redaster*) was awarded a score of 0.3 by both assessors, and the response *banlions* (for *bandilons*) was awarded a score of 0.8 by both assessors. When the two assessors' scores did diverge, this never exceeded a difference of more than 0.1 point. Given the high inter-rater reliability, the means of the two assessors' scores were used in the statistical analyses.

We analysed the students' test performance by means of a mixed effects regression model, using the *glmmADMB* package in R (Skaug, Fournier, Nielsen, Magnusson & Bolker, 2015).[5] The predictor of primary interest in the analysis was, of course, the type of gloss, i.e., whether the gloss was multimodal or text-only. As mentioned, we also included the students' performance on the multiple-choice questions concerning text content in the analyses as a potential predictor of vocabulary uptake. In none of the regression models was this found to be a significant predictor of either form recall or meaning recognition. However, another factor, which we had not anticipated, emerged: Pseudowords encountered earlier in the text tended to be recalled better than those encountered later. Therefore, the sequential position of the six pseudowords in the story was included as a predictor in the analyses, and so was the interaction between gloss type and the words' sequential position.

4.3 Participants

Participants in the first trial were 48 English majors, aged 19–20. They were in their second year of study at a university in China. They all followed the same English language programme, had taken the same university entrance exam and the same exams

to move from the first to the second year of university study. The student cohort was divided into same-sized classes for their English proficiency courses. The experiment was conducted with two such intact classes (n 24 and n 24), which were assigned to treatment condition by the flip of a coin.

Participants in the second trial were 48 Malaysian university students, aged 21–24 and enrolled in a four-year programme to become teachers of English. The second and third years of their programme were offered at a university in New Zealand. The students were halfway through the third year of the programme and had thus been living and studying in an English-language environment for a year and a half. The student cohort was randomly divided into two groups. Due to logistic circumstances, the group sizes were not the same (n 31 for the multimodal glosses and n 17 for the text-only glosses).

Participants in the third trial were 29 English majors (aged 19–20) at a university college in Flanders, Belgium, where they were enrolled in a programme preparing them to become translators. Their mother tongue was Dutch and they were halfway through the first semester of their second year of the programme. They had all taken the same English exams to move from the first to the second year. The students were randomly assigned to one of the two reading conditions (n 14 for the multimodal glosses and n 15 for the text-only glosses).

5. Results

The descriptive statistics regarding the three experimental trials are summed up in Table 1. The results of the inferential statistics are incorporated in the sections below.

Table 1: Descriptive statistics

5.1 Trial 1

5.1.1 Meaning recognition

Although the meaning-recognition test was administered second, its results will be presented first, as this test served to answer the first research question — about the effect of multimodal glosses on meaning retention. The mean per-item score in the text-only gloss condition was 0.35 and in the multimodal condition it was 0.31, a non-significant difference. Both groups' scores showed a decrease for pseudowords encountered later in the text, and this effect was significant ($z=-2.19$, $p<.05$). This trend occurred irrespective of whether the glosses contained a picture.

5.1.2 Form recall

Turning now to the form-recall test, the mixed effects model returned significant simple effects for both gloss type (z: -3.00, $p<.005$) and position of the target word in the story (z: -3.20, $p<.005$), as well as a significant interaction of these two factors (z: -3.06, $p<.005$). The effect for gloss type was such that recall was more accurate in the text-only gloss condition: Per-item means were 0.082 and 0.041 for the text-only and multimodal conditions, respectively. The effect for position in the story confirmed the observation noted above, i.e., that participants recalled the pseudowords that appeared earlier in the text better than those encountered later. The interaction effect between gloss type and position arose because the decrease in accuracy for words encountered later in the story was larger in the multimodal gloss condition.

5.1.3 Content questions

The type of gloss did not appear to have an impact on text comprehension — the average per-question score was 0.63 in the text-only and 0.61 in the multimodal condition.

5.2 Results of Trial 2

5.2.1 Meaning recognition

The gloss condition with pictures resulted in poorer meaning recognition, with per-item means of 0.67 and 0.59 for text-only and multimodal glosses, respectively. The difference fell short of significance, however. There was a marginal effect for the position of the target words in the story (z: -1.86, p=.06), but no significant interaction between gloss type and position.

5.2.2 Form recall

The analysis of the form-recall data reiterates the finding in Trial 1 of a *dis*advantage in form recall for the multimodal gloss condition. The per-item means were 0.158 and 0.100 for the text-only and multimodal conditions, respectively.

Significant effects were again found for gloss type (z: -3.97, p<.001) and for the position of the target word in the text (z: -2.23, p<.05). Again, a significant interaction of these two factors emerged (z: -3.20, p<.005), meaning that the decrease in test performance on target words encountered later in the story was most dramatic under the multimodal gloss condition.

The overall test performance was markedly better in Trial 2 than in Trial 1, suggesting that the participants in Trial 2 were more advanced learners (which is not surprising given their profile as pre-service ESOL teachers and their long immersion in an English-language environment – see section *4.3 Participants*). A further statistical analysis of the form-recall scores was run which combined the data from Trials 1 and 2, and included group of participants (Chinese students in Trial 1 vs. Malaysian students in Trial 2) as an additional factor. This analysis confirmed the effects of gloss type (z: -3.71, p<.001), position of the words in the text (z: -4.00, p<.001), and the interaction of these two predictors (z: -3.54, p<.001). The Malaysian group were found significantly more accurate in their recall (z: 4.46, p<.001), and there was also an interaction between participant group and position of the words (z: 4.02, p<.001), with the (less advanced) Chinese group showing a more dramatic decrease in their recall of words encountered later in the story.

5.2.3 Content questions

No significant difference was found in content recall between the text-only and multimodal conditions. Mean per-question scores were 0.76 and 0.67, respectively.

5.3 Trial 3

5.3.1 Meaning recognition

For this third trial, the two gloss conditions produced very similar scores on the meaning test, with per-item means of 0.44 and 0.45 for text-only and multimodal glosses, respectively. The analysis returned a significant effect for the position of the

words in the text again (z=-3.30, p<.001), with poorer scores for items that appeared later in the text. This pattern emerged irrespective of the gloss type.

5.3.2 Form recall

As in the other two trials, recall of word form was worse in the gloss condition with pictures. Per-item means were 0.129 and 0.068 for the text-only and the multimodal gloss condition, respectively. This difference fell short of significance, however. Regardless of gloss type, a marginally significant effect was found for the position of the words in the story (z: -1.81, p=.07), with poorer recall of words encountered later.

5.3.3 Content questions

The scores on the content questions were similar under the two conditions, with per-item means of 0.81 and 0.85 for the text-only and the multimodal gloss condition, respectively.

Judging by the vocabulary test results, the participants in the third trial were more advanced than those in the first trial, but not yet as advanced as those in the second. Their performance on the content questions was better than that observed in both earlier trials, however. This might be due to the fact that the subject of the text (Vikings) was more familiar to these Flemish students than to the Chinese and Malaysian students who took part in trials 1 and 2.

6. Discussion

None of the three experimental trials reported here furnish evidence that adding pictures to marginal glosses aids L2 readers' retention of the glossed words. Contrary to

what was found in some earlier studies (e.g., Kost et al., 1999; Yoshii & Flaitz, 2002), the post-reading test we used to gauge retention of word meaning showed better results in two of the trials in the gloss condition without pictures — which is in agreement with Acha's (2009) findings —, although it is important to remember that in the present study the differences were not pronounced enough to reach statistical significance. This at least illustrates the usefulness of replication studies to assess the generalizability of research findings, and, in this case, to caution against over-optimistic expectations about the effectiveness of a proposed pedagogic intervention based on a relatively small number of publications.

Concerning the effect of multimodal glosses on L2 readers' recall of the *form* of glossed words — a facet of word knowledge left unexplored by previous studies on the subject —, the results of all three trials suggest a *negative* effect (relative to that of text-only glosses). In Trials 1 and 2, form recall was significantly poorer when pictures accompanied the verbal information in the glosses. The result in Trial 3 was in the same direction, but it fell short of statistical significance. The sample in this third trial was quite small, however, and so less likely to reveal effects at p<.05, despite the marked difference in mean scores. Cohen's *d* effect size calculation (based on group means, standard deviation and sample size) actually suggests there were medium-sized effects in favour of the text-only glosses which were virtually identical across the three trials: Trial 1: *d*=0.516; Trial 2: *d*=0.555; Trial 3: *d*=0.515.[6] Overall, then, the findings suggests that far from helping participants learn the forms of novel words, the addition of pictures may have hindered this learning. As explained at the end of the literature review, this may be accounted for by the phenomenon of split attention, whereby — in the case of multimodal glosses – pictures usurp attention that learners might otherwise

give to the words proper. Evidence for this should be noticeable especially in tests that require accurate recall of word form, and that is precisely what was observed in the experiment.

While our study was not designed with a view to examining the effect of the position of target words in a story, this factor emerged as a significant predictor, with words encountered later in the text generating poorer scores than those encountered early on. Regarding performance on the meaning-recognition test, this effect was significant in Trials 1 and 3, and marginally significant in Trial 2. In the case of the form-recall test, it was significant for Trials 1 and 2, and marginally significant for Trial 3. Of particular interest for the purpose of the present study, this position effect tended to be greater for multimodal glosses than for text-only glosses, as shown by the significant interaction between gloss type and word position in Trials 1 and 2. Form recall was thus negatively affected by the presence of a picture in the gloss, but this effect was worse for items encountered later in the text. Why this happened is something which can only be speculated about at this stage.

One such speculation is that the participants became gradually more engrossed in the story as it unfolded itself, and therefore took less time to inspect the glosses. If so, readers may have been inclined to quickly take in the pictures in the multimodal gloss condition while overlooking the textual parts of the glosses. The finding is also compatible with Cognitive Load Theory (Sweller, 1994). Reading a text accompanied by multimodal glosses entails more input overall than reading the text with single-mode glosses. While a learner may find the diverse stimuli manageable at the start of an activity, this may become increasingly challenging as the information to be processed accumulates. If so, it is again conceivable that, as readers progressed through the text,

their attention to the textual component of the glosses decreased when also pictures were available.

However, as mentioned, these are of necessity speculations about reading behaviour, because the experiment reported here was of the pen-and-paper kind. Follow-up research will need to use eye-tracking technology to investigate which aspects of glosses and passages of the text itself attract attention from L2 readers as they read under varying gloss conditions.

While the experiment reported here focused on the use of visuals in *L2* reading by adults, it is worth mentioning that their proclaimed benefits have also been the subject of debate in the context of *L1* literacy development, in particular children's development of a sight vocabulary for reading (see Samuels [1970] and Sadoski [2005: 224–228] for reviews from opposing standpoints). Research on children with developmental disabilities, for example, has generated evidence that the co-presentation of a written word and a picture of the word's referent hinders learning of the former, because the picture 'overshadows' (or even blocks attention to) the to-be-learned word form (Dittlinger & Lerman, 2011). The present study can therefore be considered as bridging that line of research to the realm of adult L2 vocabulary learning.

7. Pedagogic Implications and Conclusions

If further research were to confirm the impression from the present study that pictures carry the risk of slowing down learners' initial uptake of formal facets of words, then this would by no means imply, of course, that visuals should be banned from materials for second language learning. There is no doubt that judiciously chosen visuals serve important purposes, such as making learning materials appealing and

elucidating unfamiliar concepts, unfamiliar referents, or unfamiliar procedures, where words alone fail (e.g., Mayer, 2009). The implication would rather be that care should be taken to provide learners with complementary learning opportunities that foster acquisition of those aspects of word knowledge (such as accurate phonological / orthographic representations and usage patterns) that pictures *per se* do not lend themselves well to. For example, output activities that incorporate the target words which the learner encountered in an input text have been shown to be beneficial (e.g., Hulstijn & Laufer, 2001; Laufer & Roitblat-Rozovski, 2015). This is consistent with Laufer and Hulstijn's (2001) *Involvement Load Hypothesis*, which holds that the likelihood of word retention is influenced by how learners engage with new words in terms of their 'need' (i.e., whether they experience a need to understand and/or use the word to perform a given task), 'search' (i.e., the effort they invest in determining the meaning or other properties of the word), and 'evaluation' (the effort they invest in assessing the accuracy of their interpretation and/or use of the word). Output tasks that require use of glossed words normally include 'need' and 'evaluation', and can thus, according to the *Involvement Load Hypothesis* be expected to be relatively conducive to learning, including learning of the word's form. Output tasks figure as predictors of word learning also in Nation and Webb's (2011) *Technique Feature Analysis*, where it is emphasized that it is tasks where learners make an effort at retrieving newly learned words from memory and where they try to use the new words in new contexts (e.g., Joe, 1998) that are particularly effective. [For a comparison of the *Involvement Load Hypothesis* and *Technique Feature Analysis*, see Hu and Nassaji, 2016.]

Still, the fact remains that not all reading is followed directly by an output activity, let alone an output activity that requires recycling newly encountered words. This, then,

raises the question of how the conditions for vocabulary uptake from glossed text might be improved through further manipulations of the input materials. A combination of the following suggestions could serve this purpose. One suggestion is to include several instances of the same target word in the reading passage. It is well known that frequency of encounters matters for incidental vocabulary acquisition (e.g., Waring & Takaki, 2003). Another suggestion is to increase the visual salience of the target words by means of typographic enhancement (e.g., Peters, Hulstijn, Sercu & Lutjeharms, 2009). A third suggestion is to provide two separate glosses for the same word, for example a pictorial elucidation to accompany an early occurrence of the word, and a textual clarification to accompany a later occurrence. If it is true that combining pictorial and textual information in a single gloss creates competition for attention, then separating the two modes of information is a way of avoiding this competition. This in effect corresponds to what was done in some of the studies that used hyperlinks to separate annotations in the context of CALL which we reviewed at the start of this article (e.g., Chun & Plass, 1996). In the case of glosses in the margin of a text, a design challenge may be to present the glosses such that a pictorial gloss for one target word does not interfere with the reader's processing of the textual gloss for another, which might happen if the latter is presented on the same page. This implementation of glossing may therefore be suitable only if the number of target words to be glossed is small. That need not be a drawback, however, because targeting only a small number of words in a text may actually be pedagogically judicious, and consistent with the observation in our experiment that recall rates for words fell rapidly as the number of encounters with unfamiliar words in the text increased.

At this stage, these suggestions for text manipulation with a view to enhancing vocabulary uptake from glossed text are of necessity just that — suggestions. Further empirical research (including, as mentioned, eye-tracking studies; e.g., Pellicer-Sanchez, 2015) would be welcome to evaluate them. For now, what we hope to have demonstrated is that the benefits of including pictures in marginal glosses should not be overestimated.

Notes

1. See Jones (2004) for evidence of input-test congruency effects in the context of word learning from single-mode and multimodal annotations.

2. The tests where participants were asked to provide the meaning of the given L2 words are referred to as "productive" tests in some of the articles reviewed here, which may cause some ambiguity. In only one of the studies (Yeh & Wang, 2002) were the participants required in one section of the post-test to produce the L2 words, but the participants' performance on this section is not reported separately from their performance on the receptive knowledge sections. Neither is it mentioned what proportion of the total test score was based on this productive measure. It is also unfortunate that no information is given about the participants' pre-test scores so as to ascertain that the treatment groups were on par regarding prior knowledge of the target words.

3. In compliance with the authors' institutional human ethics guidelines, all the participants in the three trials had been informed the activity was part of a study on reading in a second/foreign language, and they had given consent to their data being used for this research. After the test papers had been collected at the end of the

experiment, the participants were briefed about the precise purpose of the experiment (i.e., the effect of gloss type) and about the pseudowords.

4. The assessors' mean scores for form recall also correlated strongly (r=.93) with a more objective measure, that of orthographic similarity proposed by Van Orden (1987).

5. In the case of the form-recall test, the dependent variable was a linear transformation of the assessors' mean score. This transformation simply involved multiplying the score by 20 in order to obtain the integer values required by the analysis. These values thus ranged from 0 (no recall whatsoever) to 20 (fully accurate recall). The statistical model included a zero-inflation parameter to account for the fact that a large proportion of responses contained no recognisable trace of the pseudoword.

6. Because the language of textual glossing in Trial 3 differed from Trials 1 and 2 (L1 translations instead of L2 definitions), we decided not to combine the data of all three trials in a single statistical analysis. Whether L1 or L2 glosses are more effective was not a research question here (see Jacobs et al., 1994, and Ko, 2012, for results suggesting they work equally well). According to the results of a study by Yoshii (2006), the choice of language (L1 or L2) is no significant moderator in comparing the effectiveness of text-only and multimodal glosses. Still, we do not want to rule out the possibility that L1 equivalents require less effortful processing, and may consequently be less subject to a split-attention effect where the textual part of the gloss is overlooked in favour of the pictorial part.

References

Acha, J. (2009). The effectiveness of multimedia programmes in children's vocabulary learning. *British Journal of Educational Technology, 40*, 23–31.

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*, 499–517.

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology, 5*, 202–232.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning, 52*, 323–363.

Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. Amsterdam, Netherlands: John Benjamins.

Boers, F. (2000). Enhancing metaphoric awareness in specialised reading. *English for Specific Purposes, 19*, 137–147.

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

Craik, F. I. M., Govoni, R., Naveh-Bejamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General, 125*, 159–180.

Dittlinger, L. H., & Lerman, D. C. (2011). Further analysis of picture interference when teaching word recognition to children with autism. *Journal of Applied Behavior Analysis, 44*, 341-349.

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning, 43*, 559–617.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition, 35*, 483–517.

Godfroid, A., & Schmidtke, J. (2013). What do eye-movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honour of Richard Schmidt* (pp. 183–205). Honolulu, HI: University of Hawai'i.

Hu, H-C. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement Load Hypothesis versus Technique Feature Analysis. *System, 56*, 28–39.

Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automacity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge: Cambridge University Press.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*, 539–558.

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*, 327−339.

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*, 19−28.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics, 19*, 357–377.

Jones, L. C. (2004). Testing recognition and recall using pictorial and written test items. *Language Learning & Technology, 8*, 122−143.

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

Ko, M. H. 2012. Glossing and second language vocabulary learning. *TESOL Quarterly, 46*, 56−79.

Kost, C. R., Fost, P., & Lenzini, J. J. (1999). Textual and pictorial glosses: Effectiveness of incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals, 32*, 89−113.

Laufer, B. (2013). Second language word difficulty. *The Encyclopaedia of Applied Linguistics*. DOI: 10.1002/9781405198431.wbeal1328.

Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a foreign language: The construct of task-induced involvement. *Applied Linguistics, 22*, 1–26.

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research, 19*, 687–711.

Macmillan English Dictionary for Advanced Learners (2nd ed.). 2007. Oxford, UK: Macmillan.

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

Mulligan, N. W. (1998). The role of attention during encoding in implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 27−47.

Nagata, N. (1999). The effectiveness of computer-assisted interactive glosses. *Foreign Language Annals, 32*, 469–479.

Nation, P., & Webb, S. (2011). *Researching and analysing vocabulary*. Boston: Heinle.

Pellicer-Sanchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition, 38*, 97–130.

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*, 113–151.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 9*, 25–36.

Porte, G. (Ed.) (2012). *Replication research in applied linguistics*. Cambridge: Cambridge University Press.

Sadoski, M. (2005). A dual coding view of vocabulary learning. *Reading & Writing Quarterly, 21*, 221−238.

Samuels, S. J. (1970). Effects of pictures on learning to read, comprehension and attitudes. *Review of Educational Research, 40*, 397−407.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*, 329–363.

Skaug, H., Fournier, D., Nielsen, A., Magnusson, A. & Bolker, B. 2015. General Linear Mixed Models using 'AD Model Builder'. Version 0.8.3.2. Accessed from http://glmmadmb.r-forge.r-project.org.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, *15*, 181−198.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*, 130–163.

Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition, 1*, 287–307.

Yeh, Y., & Wang, C-W. (2003). Effects of multimedia vocabulary annotations and learning styles on vocabulary learning. *CALICO Journal, 21*, 131–144.

Yeung, A., Jin, P., & Sweller, J. (1997). Cognitive load and learner expertise: Split attention and redundancy effects in reading and explanatory notes. *Contemporary Educational Psychology*, 23, 1–21.

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

Yoshii, M. & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of picture and annotation types. *CALICO Journal, 20*, 33–58.

Appendices

<div align="center">

Appendix 1

</div>

The sound of the horn. Again... 'That was the fifth time,' Anne whispered. The sound came from the other side of the hill, from the valley where her village was and the castle beyond it. 'What's going on, Sam?' Anne said to her dog, 'Let's find out.'

Anne and Sam had been looking after the sheep on this hillside since early in the morning. She was waiting for Jack, her brother, to come and join her after he finished helping their uncle mend the redaster that was damaged in yesterday's storm.

The uphill walk was slow, not because it was a steep climb, but because the ground was so muddy. At times, Anne's feet sank deep into the mud and she had to be careful not to lose her bandilons. Sam didn't seem to be bothered by the mud and ran ahead of her. The horn was blown once again. At long last, they arrived at the top of the hill so they could see the village and, beyond it, the castle. That was where the sound of the horn came from. Anne could see movement on the battlements.

Villagers were hurrying towards the castle. Some were carrying hastily gathered possessions. There was a lot of shouting, and although Anne was too far away to recognize the words, people were unmistakably in a panic.

redaster = chicken house

bandilon = simple, light shoe

battlements

panic = fear

The first villagers were now crossing the prantenon of the castle.

'Where are mum and dad?' Anne whispered. Then she saw them, among a group of people who must have been the last to leave the village. Mum kept turning around to point to the hillside where Anne was standing. 'She doesn't want to leave me behind,' Anne realized. Then, she saw how dad threw her mum over his shoulder and staggered on, trying to keep up with the other villagers.

It didn't take long before Anne understood why everyone was seeking refuge in the castle. As her eyes travelled along the river that passed the village, she saw what was causing the panic – there, at the canimat normally occupied only by small fishermen's boats, was a ship of a kind that she'd never seen before but that she'd heard about in horror stories. The dragon-shaped front left no doubt about it. 'Vikings!' she gasped.

She'd heard terrifying things about these warriors from the north. How they would not only attack towns along the coast but sail up rivers in their paniplines and attack villages inland as well.

prantenon = bridge that can be pulled up

stagger = walk with difficulty

refuge = safety

canimat = platform for boats

panipline = long boat

A group of about thirty heavily armed Vikings had disembarked and were approaching the village. They looked terrifying. Their shields displayed monsters and the sunlight that momentarily broke through the clouds was reflected by the sharp blades of their swords and battle axes.

The Vikings did not seem to be in a hurry. They must have realised they were too late to catch up with the fleeing villagers or to launch a surprise attack on the castle. The last group of villagers, including Anne's mum and dad, had meanwhile made it to the safety of the castle walls. There was no movement on the battlements, as if the soldiers were holding their breath. What were the intentions of the Vikings?

The Vikings searched all the houses and sheds in the village one by one and brought out everything that might be of value.

disembark = get off the ship

battle axe

flee = run away

Two men were giving directions as to what should be loaded onto a large stavener, to be taken to the ship.

Through her tears, Anne suddenly noticed a figure leaving the shelter of the haystacks between the village and where she was standing, and run up the hillside towards her. It was her brother, Jack, who must have been on his way to her when the Vikings arrived. The Vikings were busy looting the village and wouldn't have noticed the boy, but unfortunately, that's when Sam started barking and ran down to meet Jack. 'Shhh, Sam!' Anne hissed, but it was too late. She heard shouts and saw several Vikings point in their direction. She had been so transfixed by the events she'd been witnessing that it hadn't even occurred to her that she should stay out of sight. The Vikings had now seen Jack, too. 'Run, Jack!' Anne shouted. But it was too late. Some of the Vikings had taken their crossbows and took aim at her brother. The first two arrows missed him, but the third one went into his shoulder. He took four more steps and then went down with a yelp. 'No!' cried Anne. But she realized there was nothing she could do. Some of the evil men were coming up the hill towards her. She turned and ran as fast as she could, desperately hoping that her pursuers' heavy armour would slow them down.

stavener = a cart with two wheels
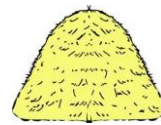
haystack

loot = steal things and do damage

transfixed: so shocked that you stop moving

crossbow

armour = body protection

Appendix 2

1. Why was Anne in the hills when she heard the sound of the horn? Underline the right answer:

 A. to walk her dog   B. to herd the sheep   C. to hide from her brother

2. Anne was wearing <u>simple, light shoes</u> on her feet. Can you remember the word that is used in the text for that? Here are the sentence from the text and a picture (perhaps they help you remember):

Anne's feet sank deep into the mud and she had to be careful not to lose her
_____ (Write as much of the word as you can remember)

3. Anne saw the villagers going somewhere in a hurry. Where were they going? Underline the right answer:

A. the river   B. the hills   C. the castle

4. Anne's brother, Jack, wasn't with her because he had to help their uncle mend their <u>chicken house</u>. Can you remember the word used in the text for that? Here are the sentence from the text and a picture (perhaps they help you remember):

She was waiting for Jack, her brother, to come and join her after he finished helping their uncle mend the _____ that was damaged in yesterday's storm. (Write as much of the word as you can remember)

5. Who was blowing the horn? Underline the right answer:

A. someone in the castle   B. someone in the village   C. someone on the Viking ship

6. Why did Anne's father carry her mother? Underline the right answer:

A. she didn't want to follow   B. she was hurt   C. he wanted exercise

7. What is the word used in the text for the type of <u>long boat</u> used by the Vikings?
Here are the sentence from the text and a picture (perhaps they help you remember):

…they would not only attack towns along the coast but sail up rivers in their
_____ and attack villages inland as well. (Write as much of
the word as you can remember)

8. Anne knew it was a Viking ship because of the shape of its prow (its front part).
What shape was it? Underline the right answer:

A. a snake   B. a mermaid   C. a dragon

9. To enter the castle, the villagers crossed a <u>bridge that can be pulled up</u> to prevent
others from entering. What is the word for that used in the text? Here are the
sentence from the text and a picture (perhaps they help you remember):

The first villagers were now crossing the _____ of the castle.

(Write as much of the word as you can remember)

10. What drew the Viking's attention to Anne and Jack? Underline the right answer:

A. the dog started barking   B. Jack started sneezing   C. Anne started crying

11. In the river there was a sort of <u>platform for boats</u> to be tied. Can you remember the word used in the text for that?

…at the _____ normally occupied only by small fishermen's boats was a ship of a kind that she'd never seen before but that she'd heard about.

12. The Vikings used a <u>two-wheel cart</u> to take what they'd stolen from the village to their ship. Can you remember the word used in the text for that?

Two men were giving directions as to what should be loaded onto a large

_____ ,to be taken to the ship.

13. <u>While</u> you were reading the text, did you expect that that you'd be asked to try to recall the words that were explained in the margin? Underline your answer:

        A. Yes, I expected this   B. No, I did not expect this.

Appendix 3

Connect the definitions/illustrations with the right words. One correct match has already been made for you as an example

2-wheel cart

armour

long boat

battle axe

bridge that
can be pulled up

bandilon

stavener

chicken house

canimat

body protection

prantenon

simple light shoe

panipline

platform for boats

redaster

Appendix 4

Per-group means (max = 6) (and standard deviations)

		Trial 1 (N 48)	Trial 2 (N 48)	Trial 3 (N 29)
Form recall	Multimodal	0.24 (0.36)	0.60 (0.55)	0.41 (0.75)
	Text only	0.49 (0.60)	0.95 (0.69)	0.77 (0.70)
Meaning recognition	Multimodal	1.83 (1.66)	3.53 (1.62)	2.71 (1.73)
	Text only	2.08 (1.64)	4.03 (1.40)	2.67 (0.82)
Content questions	Multimodal	3.67 (1.37)	4.00 (1.37)	5.14 (0.77)
	Text only	3.79 (1.35)	4.55 (1.12)	4.87 (0.99)