**Title:** **CALL versus Paper:**

**In Which Context Are L1 Glosses**

**More Effective?**

**XXXX. XXXX**

**Brigham Young University-Idaho**

**Title:** **CALL versus Paper: In Which Context Are L1 Glosses More Effective?**

**Abstract:** CALL glossing in the first language (L1) or second language (L2) texts has been shown by previous studies to be more effective than traditional, paper-and-pen L1 glossing. Using a pool of studies with much more statistical power and more accurate results, this meta-analysis demonstrates more precisely the degree to which CALL L1 glossing can be more effective than traditional L1 glossing. Results indicate, as previous research has shown, that CALL L1 glossing is significantly more effective in L2 reading comprehension than traditional L1 glossing. That is, the mean effect size is significantly higher (*p* < .001) for studies that use CALL L1 glosses (*g* = 1.44) when compared to studies that use paper-based L1 glosses (*g =* .50). This article explains how and under what circumstances CALL L1 glosses may be more or less effective than traditional L1 glosses.

**CALL versus Paper: In Which Context Are L1 Glosses More Effective?**

**Introduction**

Reading is widely considered an important skill and an excellent source of second language (L2) input (e.g., Day & Bamford, 1998; Gettys et al., 2001). Because we live in the information age, online reading has become extremely important (e.g, Bowles, 2004; Gettys, Imhof, & Kautz, 2001). With the amount of online texts, there is a greater opportunity to take advantage of this resource in order to expose the L2 learner to more L2 input. Since we find ourselves in an online world in its various forms of computers such as desktops, I-Phones, I-Pads and other devices, it is paramount to determine the effectiveness of CALL contexts in L2 reading comprehension. If such contexts have good possibilities for positive L2 reading comprehension, how can this best occur? In the study of computer-assisted language learning (CALL), the degree to which native-language (L1) glossing has an effect on L2 reading comprehension has been a subject of research and discussion (e,g., Bowles, 2004; Guidi, 2009; Yanguas, 2009). The present study, a quantitative meta-analysis, attempts to combine all CALL L1 glossing experimental research in order to better understand the effects of CALL L1 glosses on L2 reading comprehension.

One of the advantages of CALL reading with L1 glosses is the enhanced ability to allocate attentional resources to the interaction of the text with the L2 reader that already occurs with traditional L1 glossing (Davis, 1989). When a reader is processing a text, putting form on meaning, this can be impeded by the necessity of having to find the meaning of a key lexical item. Studies clearly demonstrate that the effects of L1 glosses are not always significant and, in certain studies, L1 glosses can even be detrimental to L2 reading comprehension (Baumann, 1994; Joyce, 1997). Indeed, some researchers would agree that, to some extent, there is or at least has been a debate on the question of whether L1 glosses are generally effective (e.g., Bowles, 2004; Jacobs, 1994).

L1 glosses are generally preferred by the L2 learner (Bell & Leblanc, 2000; Cheng and Good, 2009) and with CALL glossing, in which there is instant access to the items, learners can choose which items they would like to access instead of the teacher or author of a textbook determining the glossed items. This empowers students to become more independent in their L2 learning because the glossed items can be more tailored to their needs. However, at some point, L1 glosses may become more of a hindrance for the L2 learner, and, consequently, L2 glosses, strategy training, and dictionary skill may become more useful.

It should be mentioned that one of the disadvantages of L1 glossing, whether CALL or traditional, is that it is in the L1. If the learner leans too heavily on the L1 while reading, or during any other task, it is possible that it can become a crutch on which the learner could possibly lean too heavily. There is obviously a role for the L1 in L2 learning (Cook, 2001; Turnbull, 2001). However, L1 glossing can become a short-term fix that may prove to be detrimental to the development of long-term strategies that will be needed when the L2 learner arrives in the target culture. Learners may not always have access to L1 glosses while reading the news on a computer or studying in a classroom in a foreign country, in which case reading strategy training, especially at more intermediate and advanced levels (XXXX, Stevens, & Asher, 2006), can be more beneficial.

There are several CALL glossing studies that have attempted to provide different kinds of glosses, besides L1 glosses, to the L2 reader in order to promote deeper processing. Studies have provided L2 glossing, audio, and pictorial glossing (e. g., Stoehr, 1999; Hayden, 1997; Yanguas, 2009). Another factor to consider with CALL L1 glosses is accessibility. In the case in which everything is glossed or at least has the possibility of being consulted, assuming that most if not all of the text is linked (e. g., Lomicka, 1998; Stoehr, 1999), CALL glossing can be an effective way of assisting L2 text comprehension. However, linking all of the text may not always be to the reader’s advantage, since the readers’ effort to learn the vocabulary may not be reinforced by putting forth effort to guess at the words. Aweiss (1994), in one of the first CALL studies, suggested that glossing may not significantly affect L2 text comprehension longitudinally. This concern relates to the idea that much of the L1 glossing studies have used the effective immediate recall protocol as recommended by Bernhardt (1991). Using immediate recall protocols may be a good idea, since some research has suggested that it may be the most effective means of measuring L2 reading comprehension (e.g., Bernhardt, 1983, 1991). L2 recall protocols are problematic, however because by default they need to be immediate, otherwise they are difficult to conduct. A meta-analysis (that included CALL studies) XXXX (2002b) found that out of 19 L1 glossing studies, 15 (79%) used the immediate L1 recall. Further, one of the characteristics of recent L1 glossing studies is that the measurement of the effects of glossing on reading comprehension is secondary, in that the researcher includes a multiple choice reading comprehension test along with more extensive vocabulary tests that are both immediate and delayed (e.g., Bowles, 2004; Ko, 1995; Yanguas, 2009). This is not necessarily a major weakness in these studies; it simply means that the longitudinal effects of L1 glossing (which includes CALL studies) on L2 reading comprehension have not been sufficiently measured. Thus, it is possible that long-term retention of items glossed in the L1 may be weaker or stronger than other types of glossing. On the other hand, since the use of L1 glosses in the minimally instrusive CALL environment, is controlled by the L2 reader, he or she may be more attentionally ready for the intake of the items into the developing system (XXXX, 2006a; 2009).

**Literature Review**

**CALL Glossing Studies**

Past research has generally found that CALL glosses, especially L1 glosses, are more effective than traditional glosses in L2 reading comprehension. One such study was conducted by Aweiss (1994), who investigated whether there was a causal relationship between L1 CALL reading supports and L2 reading comprehension in English-speaking learners of Arabic. A repeated-measures, within-groups design was used with a treatment of four successive texts (three with a different type of gloss and a control), with an approximate word total of 500-600 words of which 14 were glossed (3% of text). The participants for Aweiss’ experiment were twenty-four beginning-level learners randomly assigned to different orders of four successive treatments. The texts chosen were considered authentic and seemed to contain material not unfamiliar to the participants. No time limit was reportedly given to the participants who were tested individually. Aweiss’ results revealed that those with L1 glosses recalled significantly more pausal units than those without glosses (*p* < .05;  = 43.00; = 30.00). Pausal units are idea units at which a native speaker would likely pause in a natural reading of the text (Johnson, 1970).

Additional research on CALL glosses was conducted by Goyette (1995). Goyette conducted a small-scale CALL experiment using English-speaking learners of French (n = 24). Authentic French texts were adapted for upper intermediate level learners. Participants were asked to immediately recall aloud in the L1 what they understood after each paragraph was read and were given the option of using either a bilingual (French-English) or monolingual (French only) dictionary. 95% of the participants used the bilingual dictionary exclusively. Results indicated that learners with L1 glosses performed significantly higher than those without glosses (*p* > .01; = 43.00; = 25.2). Goyette’s study had several flaws: 1) each subject was tested individually, which might have contributed to confounding variables affecting the treatment, and 2) the subjects were given L1 English both printed and CALL texts either before or after L2 French texts, further confounding the effect of L1 glossing (instant lookup on the computer); and (3) a within-groups design was implemented. Goyette’s results are pertinent, however, because the L1 glossing treatment (called ‘dictionary’ in the study) was probably not extensively used for the English text, whereas the L1 glosses would be used for the French text.

In a ground-breaking CALL study using a linked text, Lomicka (1998) found no significant difference for L1 and L2 glossing in a CALL experiment. Participants were native English-speaking learners of French at the beginning level (second semester). Participants (n = 12) read a text (a Francophone poem) containing 95 words with 92 of them glossed (97% of text) which means that the learner had much flexibility in which glossed items to consult. The text and glosses were equipped with instant look up capability rather than the traditional L1 or L2 glosses. Participants were also told to read at their own pace*.* The dependent variable was the number of oral, recorded explanations that were correct with regard to the text. Comparison groups included those with no glosses (control group), those with L1 glossing (experimental group 1) and those with pictures, questions and pronunciation (experimental group 2). The results ( 1 = 15.4%; = 21.7%) suggested that glossing had no significant (*p* > .05) effect on reading comprehension. Given the very small number of participants (n = 12), these results should be interpreted with caution. By way of comparison, in a large (n = 62), randomized CALL study of English learners of German, Stoehr (1999) found that participants with L1 glosses recalled a significantly higher amount of L2 text than those without glosses (*p* < .0001; = 15.33;  = 8.03). The text used had 621 words and was deemed authentic and meaningful by the researcher. Stoehr used a recall task immediately after reading the text in which the participants in the L1 were asked to recall factual content of the text. A factor to be considered in explaining the results of Stoehr’s experiment is that most words were instantly available to the L1 glossing group (100%), which can be considered a different condition from that of ‘traditional’ glosses accompanying the text as in the case of Davis (1989), Jacobs (1994), and Luo (1993). However, Stoehr’s findings were significant. Learners with access to L1 glosses simply did better than learners without access to glosses. Stoehr also found that L2 paraphrases had a significant effect on L2 text comprehension.

Guidi (2009) conducted a large-scale CALL experiment in which first-year English speaking learners of Spanish (n = 65) were randomly assigned to treatment and control groups. The text for Guidi’s experiment was a non–literary article written by the researcher. Among other tasks, participants were asked to take a multiple-choice reading comprehension test (called a “questionnaire”) in the L1. Results indicated that learners with L1 glosses performed significantly higher on the reading comprehension measure than those without glosses (*p* > .05;  = 16.21; = 10.56). The authors conclude that glossing works in tandem with other variables to increase L2 reading comprehension.

Bowles (2004) studied the effects of what she termed “L2 glossing,” which actually meant L1 glossing in her study. One interpretation of “glossing” is the assumption that the glosses are in the L1 which is, of course, not always true. There can be L1 and L2 glossing, sometimes at the same time (Lomicka, 1998) or in the same experiment (e.g., Hayden, 1997; Stoehr, 1999). In a study with 93 participants of which 43 were eliminated for either prior L2 lexical knowledge or for not fully completing the experiment, Bowles found that both L1 glosses in a CALL or traditional format significantly improved L2 text comprehension. In fact, Bowles found that the effect sizes were higher for “paper-and pen” (i.e., traditional) glosses versus CALL glossing. Traditional glosses had an effect size of 2.96 and CALL L1 glosses had an effect size of 1.75. Bowles eliminated 43 participants based on whether they had completed all assessments (there were two vocabulary posttests), or looked at the reading text after finishing the text, or had prior knowledge of the target vocabulary that was being glossed--perhaps the main question of her study. Bowles also had the participants think aloud while reading the L2 text, a technique that may be unnatural and intrusive. Additionally, it is important to note that Bowles’ study was ambiguous about random assignment of participants to control or experimental groups.

Salem (2006) conducted a large (n= 93) CALL study with second-year students of Spanish. Salem created his own text of which 4% was glossed. Salem randomly divided students into four test groups: a text-only, a textual plus audio glossing, a textual, audio and picture glossing, and a textual, audio, picture and writing (writing down the glossed word) glossing group. Results indicated that participants with L1 (textual) glosses performed significantly higher (*p* < .01; = 18.89;  = 9.08) than the control group on a 25-item multiple-choice reading comprehension test. Very interestingly, Salem also found that L1 textual glosses (not textual glosses with audio and/or pictures or writing) provided the highest reading comprehension scores. The extremely high mean score for the glossing group compared to the control group was a salient feature of the study. One possible explanation is that the researchers constructed the text insomuch that the glosses were not only unknown words but also completely essential for a basic understanding of the L2 text.

Yanguas (2009) studied the effects of L1 glosses on L2 reading comprehension of fourth-semester native English speakers learning Spanish (n = 94) by dividing the test subjects into four groups: 1) control, 2) textual, 3) pictorial, or a 4) combination (pictorial + textual gloss). These intermediate-level participants performed significantly higher on an immediate multiple-choice test task than the control group without glossing (*p* < .01; = 6.25;  = 4.13). The authentic text on the reading comprehension test included 21 out of 534 words glossed, or 4% of the text. Yanguas made use of hyperlinked L1 glosses which the learner was able to click on to have access to the text. Yanguas concluded that, although there was no significant difference among the experimental groups (excluding the control group), it was possible that other types of glossing, (i.e., pictorial) may be beneficial because readers may use different mental capacities to process different types of glosses. Indeed, textual and pictorial glosses together had a higher overall mean score than only textual glosses. We should keep in mind, however that Yanguas’ study had intermediate learners and that glosses generally should generally ‘fit’ with the learner level if they are to be effective (Joyce, 1997).

**A Traditional Glossing Study**

The CALL studies mentioned above are compared in the present study to traditional glossing studies. Joyce’s (1997) study on traditional glossing is a salient example of the effects of learner level on L1 glossing studies. Joyce (1997) conducted a fairly small study (n = 35) to determine the effect of L1 glosses on the L2 reading comprehension of native English-speaking participants of different levels (first, second and third semester learners). Joyce observed there was a significant effect for the level of the participants, and that participants did not perform significantly higher on L1 recall tasks than those without glosses (*p* > .05), which was the case for all three group levels. At the beginning-level, first-semester group with L1 glossing did not perform significantly better than those without glosses (*p* > .05; = 9.50;  = 8.64). The beginning-level (102) second semester intermediate group had the highest results in favor of L1 glosses of the three groups, even though these were, again, not significant (*p* > .05; = 14.47;  = 10.89). Surprisingly, participants with L1 glosses had lower recall scores than those without L1 glosses for the intermediate level group (the most advanced group in the study (*p* > .05; = 13.08;  = 15.06)). The results of two other groups (first year) suggest that L1 glosses might have more of an effect on L2 reading comprehension for learners of lower proficiency, although it appears that the intermediate level was the best ‘fit’ for the text and L1 glossing. Given the impact of learner level on text recall, one must wonder if an optimal text level exists for a given L2 reader for which L1 glosses would be effective (Cheng & Good, 2009; Joyce, 1997). The second (intermediate) semester students in Joyce’s study were at a level conducive to the text and to L1 glossing conditions. The text chosen by Joyce (1997) had a “large number of cognates and high-frequency words in the passage [to make] it easier for beginning French readers to understand” (p. 59). The choice of text and how it related to learner level may have affected the results of Joyce’s study. Other L1 glossing studies have been conducted showing that L1 glossing can be effective (Davis, 1989; Jacobs, 1994; Luo, 1993; Palmer, 2003).

**How CALL Glosses Can Be Effective**

**Previous Meta-Analytic Research**

XXXX (2002b) found that learners with at least one year of instruction made the best use of L1 glosses, whether in computer or traditional format. He also observed effect sizes of .37 for first-year, .73 for second year and .14 (of no practical importance) for third year or higher L2 learners. Thus, these results suggest that there may be a linguistic threshold at which glosses may have a ‘peak effectiveness’, or the L2 developmental point at which the L2 learner can best make use of the L1 glosses. That is, L1 glosses for L2 reading are not always effective (e.g., Baumann, 1994; Jacobs, Dufon, & Hong, 1994). XXXX (2006), in one of the first CALL meta-analyses, located 18 study reports or experiments (sometimes a study will have several experiments within it), to include in his meta-analysis. XXXX extracted the effect sizes, weighted them and compared the CALL studies to the traditional glossing studies. The effect sizes were 1.09 for CALL studies, a large effect size, and .39 for traditional glossing studies, a smaller effect size. Differences between groups were highly significant with a fixed effect analysis (*p* = .0001). XXXX suggested that, even though there was a significant difference between CALL and traditional glosses and it likely enhances L2 reading comprehension, there were good arguments for and against the use of CALL L1 glosses. Arguments for including L1 glosses include faster L2 access, more motivation to read in the L2 and more efficient allocation of attentional resources. Arguments against using L1 glosses include glossing requires too little effort so retention may be minimized, learners may think that there is always one translation that is perfect for an glossed item and the L1 should generally be avoided while learning the L2. XXXX suggested that more studies should be included before making large claims as to the effectiveness of L1 glosses. XXXX further suggested that L1 glossing research is important because, when placed side by side, L1 glosses are much more widely used than other ancillary aids such as pictures or audio glosses.

In a larger meta-analytic study in 2009, XXXX studied the effects of L1 or L2 glosses on L2 reading comprehension and whether CALL glossing, either in the L1 or L2, had a significant comparative effect on L2 reading comprehension. XXXX’s meta-analysis was comprised of 32 study reports out of which he identified 7 CALL studies and 25 traditional glossing studies. The overall effect sizes were .92 for CALL studies and .43 for traditional glossing studies. This means that, for learners who were given L1 glosses in CALL contexts, 81% of them performed higher than those without glosses. He further claimed that traditional glossing was effective as well, but not as effective as CALL glossing. On measures of L2 reading comprehension, he found that 64% of learners who were provided with traditional glosses performed better than learners without traditional glosses. XXXX conducted a test of homogeneity between the two groups and found that there was a significant difference (*p = .*0001*)* between these groups. In other words, both types of glossing were effective; however, CALL glossing was significantly so when combining L1 and L2 glosses. XXXX concluded that even though the results were significant, there was more potential for CALL glossing than was evident in his study:

CALL glosses are more flexible and much faster than traditional glossing for providing the meaning of lexical items. To make the comparison truly fair, one can imagine a meta-analysis examining the difference between a CALL hypertext dictionary (simply clicking on a word) and a paper-based dictionary. Assuming the inclusion of well designed experimental research in such a meta-analysis, it is possible that the difference in L2 reading comprehension would be even greater than that presented here (XXXX, 2009; 156).

XXXX included a few studies that, while not technically glossing studies, made use of “electronic dictionary access,” which fit into his definition of glossing, which was "immediate or near-immediate access to the textual definition of L2 words” (p. 147). His findings suggested that, if all of the studies included in his meta-analysis were true glossing studies rather than electronic dictionary studies (e.g., Goyette, 1995; Knight, 1994), the results would have been much larger. The analysis of this finding is that he was probably correct (see Appendix A). If he had not included the Goyette (1995) or Knight (1994) studies, which used an electronic dictionary, the overall effect size would have been an effect size *(g) of* 1.20, a larger effect size similar to the one in XXXX’s previous meta-analysis (2006) with a significant difference between groups (*p =* .0002). Interestingly, the p-value goes down because the *n* size is reduced when we do not include the extra studies. Of course, whether or not XXXX had included the Goyette (1995) or Knight (1994) studies would not have significantly changed the overall results. Further, electronic dictionary access is perhaps more similar than different from L1 glossing; thus pedagogically, the inclusion of such studies did not seem to corrupt XXXX’s (2009) results.

In the present study, which is on the effects of L1 glosses (not L2 glosses as in XXXX, 2009) we are attempting to update XXXX’s 2006 meta-analysis which included only 18 studies. Our study includes 28 study reports, which is a larger meta-analysis, and, therefore, has more statistical power. XXXX’s original meta-analysis also had a much smaller sample size (*n =* 875) compared to the sample size in the present meta-analysis (*n* = 1458).

**Method**

A meta-analysis is only as valid as its studies so we included only true or well-designed quasi-experiments. A true experiment is where participants are assigned randomly to treatment or control groups. There are many study designs that fit under the definition of quasi-experiments (for a discussion of study quality, see the excellent book by Campbell & Stanley, 1963). However, we included only quasi-experiments that had two intact groups, and that provided a pretest and a posttest to each group.

In order for the meta-analysis to be of a sufficiently high quality, the included studies had to meet the following criteria: (a) the study had to be either an experiment or a quasi-experiment, (b) the analysis included all studies up to and including the year 2011, (c) at least one of the dependent variables of the study was reading comprehension and (d) the effect of immediate access to textual glosses (in the L1) versus no access to glosses was tested. A total of 27 study reports met the four criteria for our meta-analysis.[[1]](#footnote-1) In this study, we attempted to include all methodologically sound research, regardless of whether it has been published, mainly because meta-analysis has shown that research oftentimes is published or not because of significant results. This has been shown to be the case at least in the social sciences (Glass, McGaw & Smith, 1981). As a result, if we only included published studies, there is a chance that our overall effect size would be larger than it should be. After conducting an analysis (see Appendix B) to make sure that published studies (i.e., in refereed journals) did not significantly differ from non-published (i.e., dissertation, ERIC documents, etc.) studies in our meta-analysis, we found that the effect sizes were not even close to being significantly different (p = .94) although, interestingly, published studies had a slightly higher average effect size (g = .74) than non-published studies (g = .72).

Several studies, even though important and informative to the L1 glossing field; did not meet our criteria. For example, Lomicka’s (1998) seminal study on glossing included both L1 and L2 glosses so it was not included. Bowles (2004) was ambiguous about random assignment of participants so we were not sure whether the study was a true experiment. Some studies were correlational or reports of student perceptions of glossing (e.g., Bell & LeBlanc, 2000) and thus were not included in our meta-analysis.

**Effect Sizes**

Effect sizes, the standardized difference between two means, are calculated to determine how strong relationships or differences between groups actually are. Effect sizes can be useful statistical tools because they can be combined, since they are on the same standardized scale. Thus, we can combine different studies and make comparisons based on variables of interest. In this case, the variable of interest is CALL L1 versus traditional L1 glosses. The effect size was entered into the software Comprehensive Meta-Analysis Version 2.2 (Biostat, 2010), from which all other statistics were derived (see Table 1). It is interesting to note that Table 1 includes the Baumann (1994) and Joyce (1997) studies, which had several outcomes, indicating that there were several experiments within their “study”, each of which compared the effects of glosses versus no glosses on L2 reading. In Table 1, a *d* statistic indicates the initial effect size. In our study we also use a Hedges *g* as well, which is a slightly more conservative estimate of effect (see Table 1).

Each outcome has its own effect size, some of which are positive (experimental group did better than control group) and a few of which are actually negative (experimental group did worse than control group). It should be noted that this meta-analysis compares 8 CALL studies to 21 traditional glossing studies. Hedges and Olkin (1985) described effect sizes as small for .20-.50, medium for .50-.80, and large for anything over .80. Cohen (1988) used similar terms for describing effect sizes, arguing that an effect size of .80 or above can be considered to have a “large” effect. Glass et al. (1981) suggested that the categorization of effect sizes as “large” and “small” is not useful since “depending what benefits can be achieved at what cost, an effect size of 2.0 might be ‘poor’ and one of .1 might be ‘good’” (p. 104). For example, if the overall effect size of studies on the effectiveness of L1 glosses on L2 reading comprehension is found to be .50, yet textbooks including L1 glosses are slightly more expensive than those without L1 glosses, an informed decision can be made as to whether the product is worth the cost.

**Results**

Table 2 summarizes statistical information for each study, such as the effect sizes and the P-values for each individual study report. As reported at the bottom of Table 1, the overall effect size across studies was .73 (a medium effect size), which means that L2 readers who have L1 glosses should perform on average over seven-tenths of a standard deviation higher than those who do not have access to such glosses. Such results suggest that, in general, learners provided with glosses perform consistently better than those without such aids. Transferring this statistic to a percentage scale, one can suggest that approximately 74% of learners provided with glosses perform higher than those without such glosses.

The main research question of the present meta-analysis is whether a significant difference exists between studies that have CALL L1 glosses versus those that have traditional L1 glosses. In Table 2, we can see a large effect size was observed for CALL outcomes (*g =* 1.44) and a small (almost medium) effect size was obtained for the traditional glossing group (*g =* .47), with a significant difference (*p =* .000) between the CALL and traditional groups (See Table 3). It should be mentioned that Table 3 is the statistical test comparing the traditional and CALL L1 glossing mean group effect sizes. It is evident that CALL L1 glosses generally have a stronger effect on L2 reading comprehension than traditional L1 glosses.

These effect sizes suggest that L2 readers provided with CALL L1 glosses comprehend L2 text over one standard deviation above those who are not provided with CALL L1 glosses. L2 readers with traditional L1 glosses comprehend L2 texts almost half a standard deviation above those without traditional L1 glosses. Converting these statistics to a percentage scale, approximately 90% of learners provided with CALL glosses should perform higher than those without such glosses. L2 readers provided with traditional glosses should perform, on average, nine-tenths of a deviation above those without such aids. As for traditional glosses, 66% of learners with traditional, paper-based glosses should perform higher than those without such glosses.

We were concerned that the Salem (2006) study with its enormous effect size of 3.50 was skewing our results so we ran the analysis without the Salem study and there was still a significant difference between groups (*p =* .005). We even ran the analysis including the Bowles study, which had two groups, CALL and traditional glossing, and which had a larger effect on traditional glossing. The Salem study was then excluded and found that the difference between CALL and traditional glossing studies was still significant, although to a lesser degree (*p =* .01). This again demonstrates the power of meta-analysis; when we have enough studies in our data pool, a missing study (or two) will not often significantly affect the overall conclusions.

It should be mentioned that if we are missing more than a study or two, then the overall results of the meta-analysis can be called more into question, especially assuming that more than several of the missing studies might contradict the overall results. However, this seems improbable based on the pattern of results we see in the studies we have collected thus far.

Table 1.

*Descriptive Statistics of the Study Reports*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Study | NE | NC | XE | XC | SDE | SDC | D |
| Aweiss, 1994 | 24 | 24 | 43.50 | 30.00 | 21.47 | 19.42 | 0.66 |
| Baumann, 1994 (beginning level; Bicycle text) | 6 | 6 | 37.17 | 33.50 | 24.45 | 24.45 | 0.14 |
| Baumann, 1994 (beginning level; Breakfast text) | 8 | 7 | 25.75 | 18.29 | 30.93 | 30.93 | 0.23 |
| Baumann, 1994 (intermediate level; Bicycle text) | 8 | 7 | 61.29 | 51.86 | 24.45 | 24.45 | 0.43 |
| Baumann, 1994 (intermediate level; Breakfast text) | 7 | 7 | 57.71 | 78.00 | 30.93 | 30.93 | -0.67 |
| Cheng, 2009 Level 1 | 9 | 7 | 2.67 | 1.43 | 1.32 | .78 | 1.10 |
| Cheng, 2009 Level 2 | 12 | 8 | 3.25 | 2.63 | 1.29 | 1.51 | .45 |
| Cheng, 2009 Level 3 | 12 | 11 | 3.58 | 3.36 | 1.17 | 1.21 | .19 |
| Cheng, 2009 Level 4 | 5 | 7 | 2.80 | 2.57 | 1.10 | 1.30 | .19 |
| Davis 1989 | 23 | 26 | 28.22 | 11.10 | 8.69 | 9.00 | 1.93 |
| Goyette, 1995 | 12 | 12 | 43.00 | 25.60 | 11.60 | 12.10 | 1.33 |
| Guidi, 2009 | 33 | 32 | 16.21 | 10.56 | 4.01 | 2.57 | 1.67 |
| Huang, 2003 | 46 | 46 | 4.43 | 3.30 | .83 | 1.19 | 1.10 |
| Jacobs, 1994 | 58 | 58 | 18.16 | 13.70 | 5.68 | 5.44 | 0.80 |
| Jacobs, Dufon & Hong, 1994 | 33 | 27 | 17.30 | 16.40 | 9.00 | 7.30 | 0.11 |
| Joyce, 1997 (beginning level) | 12 | 11 | 9.50 | 8.60 | 7.76 | 7.76 | 0.11 |
| Joyce, 1997 (beginning level) | 17 | 18 | 14.47 | 10.89 | 7.76 | 7.76 | 0.47 |
| Joyce, 1997 (intermediate level) | 13 | 18 | 13.10 | 15.10 | 7.76 | 7.76 | -0.26 |
| Ko, 1995 | 64 | 63 | 13.05 | 12.86 | 3.95 | 3.21 | 0.05 |
| Ko, 2005 | 30 | 31 | 20.90 | 19.58 | 2.14 | 3.52 | .45 |
| Knight, 1994 | 54 | 51 | 74.01 | 56.65 | 27.29 | 23.35 | .68 |
| Kwong-Hung 1995 | 55 | 60 | 6.15 | 5.97 | 1.99 | 1.77 | 0.23 |
| Lou 1993 | 16 | 17 | 14.63 | 6.24 | 8.85 | 5.14 | 1.17 |
| Martínez-Fernández, 2010 | 28 | 14 | 6.71 | 6.57 | 2.34 | 3.01 | .05 |
| Palmer, 2003 | 31 | 28 | 20.37 | 11.52 | 6.35 | 5.96 | 1.44 |
| Salem, 2006 | 19 | 18 | 18.89 | 9.06 | 2.51 | 3.10 | 3.50 |
| Stoehr, 2000 | 33 | 29 | 15.33 | 8.03 | 5.31 | 4.26 | 1.51 |
| Yanguas, 2009 | 20 | 23 | 6.25 | 4.13 | 1.37 | 1.21 | 1.65 |
| **Overall Effect Size, Hedges g, Random Effects** |  |  |  |  |  |  | **.73** |

**NE** = Number of participants in experimental group

**NC**= Number of participants in comparison group

**XE**= Mean of the experimental group

**XC**= Mean of the control group

SDE=Standard Deviation of the experimental group

SDC=Standard Deviation of the control group

D= Raw Effect Size, or standardized difference between experimental and control means

Table 2.

*Overall Meta-Analysis, CALL versus Traditional Glosses*



Table 3.

*Test of Homogeneity of CALL versus Traditional L1 Glossing Variable*

Q-value Degrees of Freedom P-value

12.41 1 .000

**Discussion**

So what does all this mean? In reality, L1 glosses are generally effective in L2 reading comprehension whether in a CALL or traditional context. We would logically expect that L1 glosses would not impede L2 text comprehension. However, in looking at the studies without meta-analytic techniques, we cannot be sure. A closer look at Table 1 reveals the general benefit of meta-analysis; if we simply look at P-values in Table 2, we can observe that 15 out of 28 of our study reports (e.g., Baumann, 1994; Cheng & Good, 2009; Jacobs et al., 1994; Joyce, 1997; Ko, 1995; Kwong Hung, 1995; Stoehr, 1999) actually *did not obtain significant results* (meaning that their P-value is higher than .05)–which would be more than half of the studies. If we were to make an overall decision based on the usefulness of glossing, we could logically conclude that L1 glosses are not effective and should not be used. However, most researchers and practitioners would intuitively not agree with such a conclusion. The overall effect size of .73 helps us more clearly understand the general effect of L1 glosses on reading comprehension, without being misled by the comparing of significant versus non-significant results.

It can be argued with all confidence that there is a significant difference between CALL L1 glossing and traditional L1 glossing studies. In fact, the initial difference in mean effect sizes between the mean effect sizes of these two groups was so large that our software could not provide an exact significance level within three decimal places (*p=* .000*)* indicating that the significant difference was at least *p = <* .001. It is interesting to note that many of the above studies controlled the amount of glossing in their texts (e.g., Yanguas 2009). One of the larger effect sizes was from the Stoehr (1999) study, which essentially glossed all the text in a “dictionary” condition. Stoehr (1999) and Goyette (1995) both obtained large effect sizes by glossing the whole text and simply letting the learner decide which items to look up. Their studies provided not only the choice of CALL L1 glossing, but also there was sufficient bottom-up, lexical help that can be considered essential for effective L2 reading comprehension. L1 glosses can be considered effective because they offer the bottom-up help that facilitates top-down processes. Interactive models of the reading process take into account both the importance of bottom-up and top down models of the L2 reading process. On this point, Jacobs (1994) suggested:

This idea of the proper use of glosses is consistent with Eskey’s (1988) concept of the need to “hold in the bottom” of the reading process. Eskey supports an interactive view of reading and cautions against too much faith in exclusively top-down routes of comprehension. He believes that vocabulary knowledge provides an essential base for top-down processing to occur. Thus, information from the printed page (the bottom) interacts with readers’ previous knowledge and reading strategies (the top) to enable readers to derive their own meanings from texts (p. 129).

Of course, CALL L1 glossing is effective only to the extent that it meets the learner’s needs. Because traditional L1 glossing is generally separate from the text (in the margin or below the text), the learner's attention is drawn to the particular item and away from the text. This can be problematic for the L2 learner, especially for those with a lack of lexical comprehension, because the text is important for understanding content and meaning. Even if the learner does not need the glossed item, attention can still be drawn away by a traditional gloss and prove distracting to the learner’s concentration. (XXXX, 2006b), although one could argue that if the learner clicks too much, the gloss may still be distracting in a CALL format.

CALL L1 glossing is arguably consulted because of a lack of L2 text comprehension. However, because the learner desires to know the meaning of the lexical items consulted, distraction is limited. CALL L1 glosses may assist when the L2 learner notices a mismatch between what is expected at the global level and what is linguistically understood or vice-versa. The ‘mismatch’ is an important point in this discussion, evoking the ‘noticing the gap’ principle which is part of Swain’s theory of L2 acquisition (1995, 1998). To a certain degree, attention involves noticing, which may be a key processing component in L2 learning (Schmidt, 1990, 1994). Similarly, noticing the gap also involves comparing and integrating the attended input (Swain, 1995; Ellis, 1997)

As discussed in the literature review of Joyce’s (1997) study, glossing also seems to be more effective for L2 learners based on the level of the text and their own proficiency level (Chun, 2006), which may be based, on the lexical threshold of L2 learners. (Bernhardt & Kamil, 1995; Laufer, 1996). CALL glossing may help address the problem of lack of vocabulary for beginning-level L2 learners by motivating them to read more. Perhaps the most essential component of learner level is the lexicon of the learner (e.g., Bernhardt & Kamil, 1995; Hu & Nation, 2000; Laufer, 1996), and glossing can address the need for lexical assistance while reading. Thus, if a learner does indeed require a vocabulary of at least 3,000 words (preferable 5,000) for reading most L2 texts, but has not yet acquired it, glossing may be able to assist at least in the short term. Of course, as mentioned previously, glossing is not always equally effective for every learner level (e.g., Cheng and Good, 2009; Joyce, 1997) precisely because of varying lexical knowledge.

The question then arises: besides assisting text comprehension and learner motivation, can glossing (including electronic dictionary access) help the L2 learner in real time to *acquire* lexical items, if we consider glossing a pedagogical aid? Knight (1994) and Salem (2006) both found that the more glossing is consulted in a CALL context, the more vocabulary is learned. Besides helping the L2 learner to become more motivated, glosses may help the L2 learner actually learn lexical items through the allocation of attentional resources, since attention to lexical items may be as essential as attention to language structure (Cook, 2001). L2 learners seem to process lexical items before or instead of grammatical items (Lee & VanPatten, 1995). When trying to ascertain what is in a text, L2 learners may not concern themselves as much about the structure of the language as they do about the lexical items. Examples of this lesser concern for structural comprehension can be seen in CALL experiments where learners access grammar explanations much less extensively than L1 glosses or a dictionary (Hayden, 1997). VanPatten and Lee (1995) explain:

The most efficient way for learners to get meaning is to process the lexical items and “skip over” the grammatical items . . . They can do so because lexical items have a rather high informational value, or what VanPatten calls communicative value . . . defined to be the relative value a form contributes to overall sentence meaning. (p. 97)

According to Vanpatten’s Model of language processing, L1 glossing may facilitate lexical acquisition at least at the level of intake. L1 glossing enables the L2 learner to have the option of attending to the input, making the text comprehensible. In CALL reading contexts, lexical items are accessed and attended to by the L2 learner when the choice (among grammatical explanations and lexical definitions) is provided (Hayden, 1997). Because the learner controls the amount of attention allocated to the input, L1 glossing can be more amenable to different learning styles. This amenability to learning styles is even more enhanced in a CALL environment.

**Conclusion**

Essentially, a CALL lesson with instant-lookup capability may be the most effective way of L2 text reading. Not only is the word glossed in the L1, the *right word for the learner* is glossed because the learner can simply click on the item he or she needs for comprehension. CALL hyperlinked texts simply provide the most options for L2 learners. Another advantage of CALL L1 glosses is the hidden nature of the glosses. The L2 reader can attempt to guess from context without have the glossed item on the same page. This may be a tremendous advantage and provides learners with greater control over their own language acquisition processes. Perhaps most importantly, learners provided with CALL L1 glosses are, in a way, highly encouraged to pay attention to the glossed item and identify whether they know the item sufficiently for L2 comprehension. Of course, there is always the possibility of too much of a good thing: thus the disadvantage of CALL L1 glossing may be that if too much of the text is glossed, the ability of the L2 learner to gain the skill of guessing vocabulary and content meaning from context would arguably be diminished.

The present meta-analysis supports the conclusion that CALL L1 glossing works. The overall mean effect size for CALL studies was found to be large--significantly larger than that of traditional studies. This result is in contrast to Bowles’ (2004) finding that CALL L1 glosses are not significantly more effective in L2 reading comprehension than traditional L1 glosses. We would have more confidence in our finding, however, since we have a larger data pool from which to draw our conclusions. Bowles simply found that both traditional (or pen-and-paper) glosses and CALL L1 glosses (albeit with a higher mean score for traditional glossing) work for L2 learners. The apparent discrepancy can be resolved when analyzed with the proper statistical analysis. We found that while both traditional and CALL L1 glosses work, it is in fact CALL L1 glosses that are shown to be significantly more effective when studies are pooled and effect sizes are combined. That is, most of the CALL studies had large effect sizes. Even if a text is only 4% glossed, as in the Salem (2006) study, we can assume that the quantity of glossing, while important (Jacobs, 1994), is not necessarily the most important variable in glossing studies. It may be that it is the hypertext environment that Chun discussed (2001) which is attractive and effective for L2 readers.

**Pedagogical Implications**

It seems important to inquire as to why this manner of glossing is not more common in L2 glossing studies. Why not put the onus of learning on the L2 reader and let him or her decide which items to pay attention to and select? If the learner could control which items were glossed, perhaps such a condition would provide even more bottom-up support for L2 reading comprehension. More studies should be conducted that demonstrate the power of learner choice in reading a text with complete CALL L1 glossing.

We should emphasize that we are not suggesting that it is time to do away with books in favor of electronic texts. Glossing is effective in either format; it is just especially so in a CALL learning environment for the reasons expressed above. The advantages of CALL glossing can be summarized as the following: (a) the learners’ attention is fully brought to the glossed item with CALL glossing, (b) it is faster and in some instances such as in the Stoehr (1999) study, more flexible, and (c) it gives the learner more control over the reading process. The above findings further suggest that L2 instructors, who would like their students to effectively understand L2 texts in the L2 classroom, should make use of online textbooks that include CALL L1 glosses with L2 authentic texts. CALL L1 glosses, especially in the short-term, can be effective for providing reading comprehension assistance to the L2 learner. In a classroom environment, the teacher should be able to use L1 glosses for the purpose of providing differentiated instruction and thus more effectively addressing the invariable differences among the linguistic levels of the students. For example, L2 learners classified as “faux-débutants” (those who may have had high school L2 instruction) may not need as much of the L1 glosses accompanying a text as true beginners. Of course, the optional nature of L1 glossing is effective in addressing this concern.

Instructors, however, should keep mind that L1 glosses should be judiciously used by the instructor, since L1 use can be helpful in general but should not be overly encouraged in the L2 classroom (e.g., Turnbull, 2001). Also, L2 learners could rely too heavily on CALL glossing, leading to decreased skill-development in deriving word or content meaning from context. The competency to acquire lexical and syntactic items may be augmented by the occasional use of texts with little or no glossing. On the other hand, L2 learners invariably face issues of frustration level and motivation. If the L2 learner must look up every difficult word of a passage whose lexical level is too high for their ability, the reading experience can be frustrating. Thus, CALL use likely should depend to a large extent on the goal(s) of the L2 instructor. If the goal is to achieve fairly unified comprehension of a passage for preparing the learner for a speaking or writing activity, motivational and comprehension needs can be addressed with a CALL L1 glossed text. If the classroom goal is to show how certain lexical items are used in context, CALL L1 glossing may be effective as well. If the L2 teacher’s goal is to provide the learner with the skill of deriving meaning from context, perhaps providing less CALL L1 glosses may be a more logical option, especially at more advanced levels.

**Limitations and Future Studies**

One important aspect of strong science is replicability, and meta-analysis is no exception. All of the studies we have included in this meta-analysis are cited and are easily accessible. The reader of this study may not agree with the results that we have found. Perhaps we have included or excluded an experiment or two. These can be added to our study for future meta-analyses by other researchers who can make use of the effect sizes we have already extracted and add to them. Other study characteristics can be coded and compared. With the statistical power in our meta-analysis, however, we feel that adding or discarding a study from our data pool will not greatly influence the overall effect sizes. This, of course, does not mean that the question of CALL glossing has been completely resolved, since our meta-analysis was heavily dependent on the primary studies included. Primary studies should continue to be innovative and find new directions and gaps in the current research which will in turn enable meta-analysts to synthesize these new, salient features of their respective studies.

Other future meta-analyses could study the moderating variable of test task that is used to measure reading comprehension in our respective studies. More research is still needed to study the long-term effects of L1 glosses on L2 reading comprehension. Most experiments have simply measured reading compression as a kind of afterthought, focusing more on the long-term effects of vocabulary instruction, although Stoehr’s study (1999) is an exception to this. It was also found that, since the XXXX (2002b) meta-analysis, in which 15 of 19 studies used the immediate recall protocol as recommended by Bernhardt, more studies have used multiple choice tests (Bowles, 2004, Cheng & Good, 2009; Salem, 2006; Yanguas, 2009). This is a potentially important moderating variable in discovering the general effect of L1 glossing on L2 reading comprehension.

A limitation of our study is the inclusion of instant lookup with and electronic dictionary instead of with electronic glosses. Although electronic dictionaries and electronic glosses seem similar, future meta-analyses could study the difference between actually typing the word as opposed to simply clicking on it. It is possible that having to type may be either more or less helpful in L2 reading comprehension and vocabulary learning. Other limitations of the present meta-analysis center around the test task used to measure reading comprehension. As Bernhardt (1983, 1991) suggests, the effects of L1 glosses on L2 reading comprehension may be best measured via an L1 recall protocol task. However, in our meta-analysis, no L1 recall protocol test tasks (or any others, for that matter) have been used for measuring the long-term effects of L1 glosses on L2 reading comprehension. As suggested above, classroom considerations should be weighed before testing (or glossing) is done in L1. Although L1 glossing studies have not used an L2 recall protocol task, its use should also be investigated in future experimental research. Productive testing in the L2 may be possible for second or third-year learners and may be more congruent with L2 instructional goals.

List of References[[2]](#footnote-2)

\*Aweiss, S. (1994). Situating learning in technology: The case of computer-mediated reading supports. *Journal of Educational Technology Systems, 23,* 63-74.

\*Baumann, C. C. (1994). *The effect of previews and glosses on the reading comprehension of beginning and intermediate students of German*. Unpublished Doctoral Dissertation, University of Minnesota.

Bell F. L., & LeBlanc, L. B. (2000). The language of glosses in L2 reading on computer: Learners’ preferences. *Hispania, 83,* 274-285.

Bernhardt, E. B. (1983). Testing foreign language reading comprehension: The

immediate recall protocol. *Die Unterrichtspraxis, 16,* 27-33.

Bernhardt, E. B. (1991). *Reading development in a second language: Theoretical,*

empirical and classroom perspectives. Norwood, NJ: Ablex.

Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting relationships between L1 and L2 reading:

Consolidating the linguistic threshold and the linguistic interdependence hypotheses.

*Applied Linguistics, 16,* 15-34.

\*Bowles, M. A. (2004). L2 glossing: To CALL or not to CALL. *Hispania, 87,* 541-552.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental

*designs for research.* Chicago: Rand McNally.

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

Chun, D. M. (2006). CALL technologies for L2 reading. In L. Ducate & N.

Arnold (Eds.), *Calling on CALL: From theory and research to new directions*

*in foreign language teaching* (pp. 69-98). San Marcos, TX: CALICO.

Cohen, J. (1988). Statistical power analysis for the behavior sciences. (2nd ed.). Hillsdale, NJ: Erlbaum.

Comprehensive Meta-Analysis version 2.2 [Computer software]. (2010). Englewood, NJ. Biostat.

Cook, V. (2001). Using the first language in the classroom. *The Canadian Modern Language Review, 57,* 402-422.

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

Davis, J. N., & Lyman-Hagar, M., (1997). Computers and L2 reading: Student performance, student attitudes. *Foreign Language Annals*, *30*, 58-72.

Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge, England: Cambridge University Press.

Ellis, R. (1997). SLA research and language teaching. Hong Kong: Oxford.

Eskey, D. E. (1988). Holding in the bottom. In P. L. Carrell, J. Devine, & D. E.

Eskey (Eds.) *Interactive approaches to second language reading* (pp. 93-100). Cambridge, UK: Cambridge University Press.

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,* 91-106.

\*Goyette, E. S. (1995). *The effects of dictionary usage on text comprehension.* Doctoral Dissertation, University of McGill.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*.

Beverly Hills: Sage.

\*Guidi, C. (2009). *Glossing for meaning and glossing for form. A computerized study of the effects of glossing and type of linguistic item on reading comprehension, noticing, and L2 learning.* Unpublished Doctoral Dissertation, Georgetown University.

Hachette Dictionnaire Oxford v.2 version iFinger. [Computer software] (1997). London : Oxford University Press.

Hayden, S. (1997). *An investigation into the effect and patterns of usage of a computer-mediated text in reading comprehension in French.* Unpublished Doctoral Dissertation, University of Pennsylvania.

Holley, F. M., King, J. K. (1971). Vocabulary glosses in foreign language materials. *Language Learning, 21,* 213-219.

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language 13*, 403-430.

\*Jacobs, G. M. (1994). What lurks in the margin: Use of vocabulary glosses as a strategy in second language reading. *Issues in Applied Linguistics, 5,* 115-137.

\*Jacobs, G. M., Dufon, P., & Hong, F. G. (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.

Johnson, R. E. (1970). Recall of prose as a function of the structural importance of the

linguistic units. *Journal of Verbal Learning and Verbal Behavior, 9,* 12-20.

\*Joyce, E. E. (1997). Which words should be glossed in L2 reading materials? A study of first, second and third semester French students’ recall (report number FL 024 770). *Pennsylvania Language Forum*, (ERIC Document Reproduction Service No. ED 427 508).

## \*Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal, 78*, 285-299.

\*Ko, M. H. (1995). *Glossing in incidental and intentional learning of foreign language vocabulary and reading comprehension.* Unpublished Thesis (M.A.), U of Hawai’i, Manoa.

\*Ko, M. H. (2005).Glosses, comprehension, and strategy use. *Reading in a Foreign Language,* 17, Retrieved January, 2, 2006 from http://nflrc.hawaii.edu/rfl/October2005/ko/ko.html.

\*Kwong-Hung, L. (1995). Bilingual Texts: *A study of the effects of providing L1 Chinese terms in L2 English texts on text comprehension and on English vocabulary acquisition.* Unpublished Masters Thesis, University of Hong Kong.

Laufer, B. (1996). The lexical threshold of second language reading comprehension: What it is and how it relates to L1 reading ability. *Jyvaskyla Cross-Language Studies, 17*, 55-62.

Laufer, B., & Hill, M. (2000). What lexical information do learners select in a

CALL dictionary and how does it affect word retention? *Language Learning and*

*Technology, 3,* 58-76.

Lee, J, F., & VanPatten, B. (1995). *Making communicative language teaching happen.*

New York: McGraw-Hill.

\*Lomicka, L. L. (1998). To gloss or not to gloss: An investigation of reading comprehension online. *Language Learning and Technology, 1,* 41-50.

\* Luo, J. (1993). *A study of the effects of marginal glosses on the reading*

*comprehension of intermediate college students of French*. Unpublished Doctoral

Dissertation, University of Pennsylvania.

\*Martínez-Fernández, A. M. (2010). *Experiences of remembering and knowing in SLA, L2 development and text comprehension: A study of levels of awareness type of glossing and type of linguistic item.* Unpublished Doctoral Dissertation, Georgetown University.

Norris, J. M., & Ortega, L. (Eds.). (2006). *Synthesizing research on language learning and*

*teaching*. Amsterdam: John Benjamins.

\*Palmer, R. C. (2003). *A comparison of the effect of glossed self-instruction reading materials and traditional teacher fronted instruction*. Unpublished Doctoral Dissertation, Indiana University of Pennsylvania.

\*Salem, E. B. (2006). *The influence of electronic glosses on word retention and reading comprehension with Spanish Language Learners*. Unpublished Doctoral Dissertation, the University of Kansas.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 1,* 129-158.

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and

SLA. In N. Ellis (Ed*.), Implicit and explicit learning of languages* (pp. 165-209).

London: Academic Press.

\*Stoehr, L. E. (1999). *The effects of built-in comprehension aids in a CALL program on student-readers’ understanding of a foreign*. Unpublished Doctoral Dissertation, University of Texas at Austin.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics* (pp. 125-144). Oxford: Oxford University Press.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.). *Focus on form in classroom second language acquisition* (pp. 125-144). Oxford, UK: Oxford University Press.

XXXX, XXXX (2002). L’alternance de codes chez deux enseignants de français langue seconde,

niveau débutant, dans l’enseignement assisté par ordinateur. *The Canadian Modern*

*Language Review*, 54, 623-652.

XXXX, XXXX (2002b). *A meta-analysis on the effects of L1 glosses on L2 reading comprehension.* Unpublished Doctoral Dissertation, Purdue University.

XXXX, XXXX (2006a). The effects of CALL versus traditional L1 glosses on L2

reading comprehension. *CALICO Journal, 23*, 1-10.

XXXX, XXXX (2006b). Factors associated with glossing: Comments on Ko (2005). *Reading in a Foreign Language, 18(1).* [*http://nflrc.hawaii.edu/rfl/April2006/discussion/XXXX.html*](http://nflrc.hawaii.edu/rfl/April2006/discussion/taylor.html)*.*

XXXX,XXXX, Stevens, & Asher (2006). The effects of Explicit Reading Strategy Training on L2 reading comprehension: A meta-analysis. In J. Norris, & L. Ortega, (Eds.), *Synthesizing research on language learning and teaching* (pp 231-344). Philadelphia, PA: John Benjamins.

Turnbull, M. (2001). There is a role for the L1 in second and foreign language teaching,

but . . . *The Canadian Modern Language Review*, 57, 531-540.

Appendix A

Table A.

Schema comparing CALL studies to Traditional Glossing without Goyette (1995) or Knight (1994) in XXXX (2009).



Table B.

*Comparison of CALL studies to traditional glossing studies without Goyette (1995) or Knight (1994) from XXXX (2009).*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | | |
|  |  |  | |  | |  | |  |  | |
|  |  |  | |  | |  | |  |  | |
|  | **Effect size and 95% confidence interval** | | | | | | |  | | |
|  |  | |  | |  | |  |  | |  |
| **Group** | **Effect Size** | | **Standard error** | | **Lower limit** | | **Upper limit** |  | |  |
|  |  | |  | |  | |  |  | |  |
| CALL | 1.0190 | | .1513 | | .7209 | | 1.3172 |  | |  |
| Traditional | .4279 | | .0534 | | .3231 | | .5327 |  | |  |
| Combined | .4935 | | .0504 | | .3947 | | .5923 |  | |  |
|  |  | |  | |  | |  |  | |  |
|  | **Summary of fit statistics** | | | | | | |  | |
|  |  |  | |  | | |  |  | |
| **Source** |  |  | | **Q-Value** | | | **df** | **P-Value** | |
|  |  |  | |  | | |  |  | |
| Between classes | | | | 13.5780 | | | 1.0000 | .0002 | |
| Total | | | | 156.3740 | | | 29.0000 | .0000 | |

Appendix B

Table B.

*Published versus non published studies*



1. Bowles (2004) was ambiguous about random assignment of participants. We are not sure about whether Huang( 2003) randomly assigned participants to respective groups. He reported that the “three classes were randomly chosen” which is not really random assignment. In another part of his study, he claims that the participants were “randomly assigned,” so we are including his study in our pool. Lomicka (1998 ) used L1 and L2 glosses at the same time. [↑](#footnote-ref-1)
2. The studies included in our meta-analysis are marked with an asterisk. [↑](#footnote-ref-2)