Glossing Frequency and L2 Reading Comprehension: The Influence of CALL Glossing

ALAN M. TAYLOR

Brigham Young University – Idaho

ABSTRACT

This study challenges the assertion that an increase in L1 glossing results in more L2 reading comprehension. The results of this study, a quantitative meta-analysis, indicate that there is a significant difference (p = .04) in L2 reading comprehension between groups based on how much L1 glossing is provided. It was found that the group with the highest average effect size–which had all computer-assisted language learning (CALL) studies–included studies with 50% or more L1 glossing. However, the second largest mean effect size came from a group that contained 5% or less L1 glossing. In looking across groups in this meta-analysis, it was found that of the six studies with the largest effect sizes, five of them were CALL studies. In light of this finding, this paper will discuss the interaction of the variables of CALL glossing and percentage of text glossed in order to determine their possible influence on L2 reading comprehension.

INTRODUCTION

In the field of second language (L2) learning, the use of glosses is abundant. Glosses, or first language (L1) translations in the margin of L2 texts, are commonplace in language teaching materials. However, glosses have not always been shown to be useful, nor is the use of L1 glosses in L2 reading comprehension very standardized. Indeed, the variability in amount and type of L1 glossing that occurs in L2 studies can be problematic and has not been studied in great detail. In fact, many L1 glossing studies have shown few significant effects of L1 glosses on L2 reading comprehension (Baumann, 1994; Cheng & Good, 2009; Jacobs, Dufon, & Hong, 1994; Joyce, 1997; Ko, 1995; Kwong-Hung, 1995; Stoehr, 1999). Indeed, some experimental studies show that the mean of the L1 glossing group did worse than that of the control (non-glossing) group (e.g., Baumann, 1994; Joyce, 1997). Such conflicting results merit further examination. Thus, this study, a quantitative meta-analysis, is designed to shed more light on the question of whether the percentage of L1 glossing significantly influences L2 reading comprehension. The reason why we are focusing on L1 glosses is because they are generally preferred to L2 glosses by learners (Bell & LeBlanc, 2000) or other glosses when given the choice (Hayden, 1997).

THEORETICAL FRAMEWORK

Glossing and Reading Comprehension

In examining experimental studies on L1 glossing and L2 reading comprehension, Taylor (2013) found that 56% of the studies did not obtain significant results. Thus, while it may be true that L1 glossing can be effective in some studies, we still do not know many of the variables that may be confounding the results of L1 glossing studies. The present study attempts to elucidate the reason for why so many studies have not obtained significant effects for L1 glosses on L2 reading comprehension by explaining how the amount, or percentage, of glossing may influence the results of these studies.

Past research supports the general effectiveness of L1 glosses in L2 reading comprehension (e.g., Taylor, 2002, 2006, 2009, 2013). Indeed, even though some studies

*CALICO Journal, 31(*3), p-p 374-389. doi: 10.11139/cj.31.3.374-389

in the past have not resulted in significant results for L1 glossing (e.g., Baumann, 1994; Joyce, 1997), L1 glossing has generally been found to be effective in meta-analyses which combine studies with significant and non-significant results and produce effect sizes with much more statistical power. CALL L1 glossing has especially been found to be effective (e.g., Stoehr, 1999; Yanguas, 2009). However, there are multiple reasons explaining why L1 glossing may not always be effective. For instance, some studies on the effects of glossing on L2 reading comprehension may not accurately test the amount of reading comprehension that has actually taken place in the respective study (Bernhardt, 1983; 1991). Or, perhaps other studies may have included texts that are not difficult enough for the L2 learner so there may not always be a need to consult the glosses.

We must also consider the degree to which glossing may actually be a distraction in L2 learning. Since, as commented above, some experiments have shown that L1 glosses are not always effective (Cheng & Good, 2009), we must examine the usefulness of L1 glossing more closely. Are the glosses distracting? Do they hinder the L2 reader in some way? Do they actually facilitate L2 reading comprehension? The present study addresses whether there may be a threshold at which too much L1 glossing is provided in reading an L2 text. Clearly, the entire L2 text cannot be glossed; otherwise, there would be no purpose to the reading activity. Yet assuming that a certain amount of L1 glossing accompanying an L2 text is facilitative for comprehension, it seems helpful to determine whether or how the percentage of text glossed may be influential.

One aspect of L1 glossing studies that merits consideration is the idea of how much need there is for the glosses. One way to look at such a need is through the lens of the lexical threshold of the reader. The threshold of vocabulary knowledge and how it relates to reading comprehension has been studied fairly extensively (e.g., Hu & Nation, 2000; Laufer, 1996; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006). The basic premise of the lexical threshold theory is that 95% lexical coverage is a threshold for basic understanding of most texts (Laufer, 1996). Optimal comprehension is about 98% coverage (see Laufer & Ravenhorst-Kalovski, 2010). Thus, if the glossing is making up the difference in the small percentage of the text that is unknown, more comprehension can occur. The amount of text that is glossed has much to do with the L2 learner's linguistic competency level.

Prichard and Matsumoto (2011) found that the 90-95% percent lexical coverage level was inadequate for L2 reading comprehension. In their study, participants could access any word in a reading test with an online electronic dictionary. Prichard and Matsumoto found that there was no significant difference between groups based on electronic dictionary use. However, if the pretest is considered and calculated with the posttest, the difference becomes highly significant. By our own calculation, taking into account the pretest scores (which were actually standardized tests of overall linguistic competency, not reading comprehension scores), it was found that the overall effect size of the pretest was -.73 and the posttest effect size was .27 which, added together (.27-(-.73)), yielded an effect size of 1.01 with a significant *p* value ($p = .0001$). Thus, even though Prichard and Matsumoto did not consider the pretest with intact groups, there actually may have been a significant effect for electronic dictionary groups. Further, there was a significant difference between groups according to how much time was spent on reading: 16.1 minutes versus 10.6 minutes ($p < .0001$). Prichard and Matsumoto claim that there was no difference between groups because the dictionary did not contain enough lexical coverage (only up to 94%) to make a difference. However, with our above analysis, this is not necessarily true. It is possible that approaching the 95% threshold can still produce a significant result even in a well-designed quasi-experiment (according to Campbell & Stanley, 1963). Very likely, other aspects of reading come into play once there is sufficient coverage, background knowledge, the ability to guess, forming hypotheses, etc. It may be that linguistic support such as lexical knowledge, rather than more top-down strategies (Eskey, 1988; Taylor, 2009), is more fundamentally influential in L2 reading.

Noticing and Comparing

From time to time, learners may not be developmentally ready for a particular linguistic item presented via L1 glosses. What if the learners are not developmentally ready for the L1 glossing? Also one might ask: will the items be integrated into the interlanguage (IL) system of the L2 learner? That is, when an item is perhaps too abstract to be understood in its context in a passage or too difficult to relate to prior knowledge, more processing difficulty may occur for the L2 learner. This assumption is primarily based on the L2 learner's linguistic threshold and L2 competency and perhaps secondarily on the choice of the L2 text content.

According to Ellis (1997), the direct integration of analyzed L2 features into the learner's IL depends partially on the developmental readiness of the learner. Thus, an L1 glossed item (i.e. any analyzed input, such as a grammar rule, word, or idiomatic expression), which is far above a learner's general language level, may be partially learned but not fully integrated into the IL. This can differ from learning formulaic speech, however, since the complete item can include several lexical items. If the learner is developmentally ready for an item, the inclusion of a glossed item may have the effect of providing a form-meaning connection for the learner of an item or chunk of items.

Glossing can be considered a technique used for the manipulation of input in the processing system of the L2 learner. Processing models can provide insight into how input becomes intake in the context of L1 glossing and reading comprehension. L1 glossing is a type of input manipulation. Attention to input can be crucial for the acquisition of a particular lexical item. In L2 learning, attention to lexical items is arguably as important as attention to language structure (Cook, 2001). Lexical items are central to this discussion because L2 learners, especially those reading an L2 text with L1 glosses, may process lexical items before or instead of grammatical items (Lee & VanPatten, 1995). Lexical items are what are accessed and attended to by the L2 learner when the choice among grammatical explanations and lexical definitions is provided (Hayden, 1997). When trying to comprehend a text, L2 learners do not seem to care as much about the structure of the language as they do about the content of the words. Evidence of this may be found in CALL experiments in which grammar explanations are accessed much less liberally than L1 glosses or a dictionary (e.g., Hayden, 1997). Lee and VanPatten (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).

Correlatively, L1 glossing may facilitate lexical acquisition at the level of intake. L1 glossing enables the L2 learner to have the option of attending to the input, making it comprehensible. Because the learner controls the amount of attention allocated to the input, L1 glossing can be more amenable to different learning styles.

L1 glossing is effective only to the extent that it meets the individual learner's needs. Because 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, especially in the occurrence of a lack of lexical comprehension, where the L2 learner can rely on attempting to understand content and meaning. L1 glossing is arguably consulted because of a lack of L2 text comprehension. There may be 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, showing the close overlap occurring between different aspects of processing models and the 'noticing the gap' principle that is part of Swain's theory of L2 acquisition (1995, 1998). To a certain degree, attention involves noticing. This is perhaps the most key processing component in L2 learning (Schmidt, 1990, 1994).

Comparing and Contrasting L1 and L2

Comparing the L2 to the L1 is essential for L2 acquisition. Comparing occurs when one consults L1 glosses. The L2 learner may notice a lack of understanding of the text or a word (i. e., 'notice the gap' see Schmidt & Frota, 1986) and make a comparison between the L1 and the L2 (for an investigation of the cognitive benefits of using the L1 in L2 learning, see Kern, 1994). As suggested above, it seems that comparing occurs either during or after one notices the gap. Noticing the gap, and subsequently, making the comparison, may occur when there is a comprehension problem when the learner uses L1 glosses, which may contribute to explicit knowledge of an item. Gass and Selinker (1994) argue that comparing occurs in intake, when "information is matched up against prior knowledge and where, in general, processing takes place against the backdrop of the existing internalized grammatical rules" (p. 303). In a L1 glossing context, comparing may take place when the learner uses the glossed items, providing a bottom-up environment for the integration of the item (Kern, 1994). Certain studies, especially CALL studies, may provide a way for the participant to notice the gap by allowing consultation when a comprehension breakdown (or the perception of one) is occurring. Thus, the L1 becomes a reference point from which the learner can read texts above his or her linguistic threshold level (Prichard & Matsumoto, 2011). Explicit knowledge of lexical semantic content can occur with bottom-up support, which is more indicative of L2 reading comprehension than L1 skills or world knowledge (Bernhardt, & Kamil, 1995). Ellis (1997) commented, "New items and rules only become part of the developing interlanguage system if learners can establish how they differ from their existing interlanguage representation" (p. 121). To a certain extent, the learner and the text are brought together through the enhanced input of the L1 glosses.

L1 glosses can enhance textual input, depending on (a) if the items glossed are essential to an understanding of the text and (b) if the learner attends to the glossed item. Of course, the learner can consult a glossed item that is relatively unimportant to the overall storyline of a text, misunderstand it, and then come to perhaps a wrong conclusion as to the general thesis of the text, perhaps as a result of other confounding variables that may exist in L1 glossing studies.

Moderating Variables

Percentage of text glossed

There is considerable variability across L1 glossing studies. Differences occur in the findings among studies on a particular question because most experiments do not perfectly replicate each other. Thus, there are usually variables in human subject research that confound the results. I argue that one such moderating variable in our pool of studies may be the percentage of text that is glossed. Jacobs (1994) claims that the percentage of the text glossed can influence the results of studies. One may assume that as more of the text is glossed, more of the text will be understood. However, most studies contain limited, targeted glossing—whether paper or computer-based (e.g., Davis, 1989; Yanguas, 2009)—and only a very few studies (e.g., Knight, 1994; Stoehr, 1999) have included unlimited access to L1 glossing. Studies have included unlimited glossing, theorizing that the higher the percentage of text L1 glossed, the more options the L2 reader has, and, one would logically assume, the higher the reading comprehension. If there is a significant difference among effect sizes of studies with a larger percentage of text glossed than studies with less of the text glossed in the L1, researchers and L2 classroom teachers can address this issue by choosing their texts glossed in the L1 accordingly. One can also infer that the question of percentage of text glossed needs further primary study and that it is possible that L1 glosses have an effect on L2 reading comprehension proportionate to the size of the overall text in which they are featured.

The results of Taylor's (2002) study support the assertion by Jacobs (1994) that a higher percentage of text glossed may result in higher L2 reading comprehension gains. Obviously, this assertion has important pedagogical implications. If the L2 teacher, in

choosing an L2 text, would like his or her L2 learners to understand the text (and perhaps to be more motivated about the activity as a result), L1 glosses should be used liberally. On the other hand, if the instructor would like the L2 learners to learn to get meaning from context, L1 glossing use should be diminished or not used at all. To our knowledge, there have been no meta-analytic studies conducted on how the number of words glossed in the L1 can affect the amount of L2 text comprehended.

Context: CALL vs. non-CALL

Another moderating variable may be the context in which a study is conducted, whether in a CALL environment or otherwise. Past research of CALL glossing has shown a positive, and often large, effect for L1 and L2 glossing (e.g., Stoehr, 1999; Yanguas, 2009). Of course, CALL glossing is not always effective. Hayden (1997), for example, did not find a significant difference between CALL glossing and traditional glossing when she combined glossing types. When learners were provided CALL glosses in multiple formats (L1, L2, grammar, sentential and cultural glossing), Hayden found that learners consulted L1 glosses most frequently and generally ignored the other formats, especially at lower levels of competency (1997). In his meta-analysis, Taylor (2006) claimed that CALL L1 glossing not only was a more effective, bottom-up means of assisting the L2 learner, allowing more focusing on top-down aspects of L2 reading, but that glossing was also, as a result, motivating to the L2 learner. He suggested that L1 glossing could result in more "lookup" behavior with students. Taylor's 2009 CALL study similarly concluded that CALL glosses are more effective, regardless of whether they were in the L1 or L2.

METHODS

Statistical Methods

The effect size, or *g*, is the standardized difference between the mean of the control groups versus the mean of the experimental group. The effect size is also sometimes called the "point estimate" which indicates the estimate of effect of the independent variable on the experimental groups. A positive effect size means that the effect of the independent variable (in the present meta-analysis, L1 glosses) on the dependent variable (L2 reading comprehension) is stronger than the effect of no treatment. Assuming that other confounding variables are controlled, a negative effect size indicates that the independent variable has a negative effect on the dependent variable. In other words, a negative effect size means that the mean of the control group is higher than that of the experimental group. These effect sizes and the sample sizes of the control and experimental groups were entered into the program Comprehensive Meta-Analysis (2010). Effect sizes are generally characterized as large (*g* = .80 or above), medium (*g* = .50-.80) small (*g* = .20 - .50) or of no practical importance (less than .20). An effect size of .20 means that on average, the learners provided with the experimental treatment (independent variable) will perform two-tenths of a standard deviation above those participants who did not receive the experimental treatment. The following research questions motivated the present study:

1. What is the overall effect size of studies conducted on the effects of L1 glossing on L2 reading comprehension?
2. What are the overall effect sizes of groups based on differing percentages of L1 glossing?
3. Does the percentage of L1 glossing significantly affect studies on L2 reading comprehension?

Procedure: Search and Inclusion of Studies

A total of 20 studies with 28 study reports met the four criteria for our meta-analysis.[1] Along with the consultation of each bibliography of each study, a variety of electronic search methods were used to find relevant studies for this meta-analysis. The most important of these were Dissertation Abstracts International (DAI), Languages and Literatures Behavior Abstracts (LLBA), the Educational Resources Information Center (ERIC), Psychology Information (Psych INFO) and Google searches. These were also used to search for theses and dissertations.

In our meta-analysis, we attempted to include all methodologically sound research, regardless of whether it had been published, because past studies have shown that research oftentimes is published 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 (or smaller) than it should be. We included non-published studies because past meta-analytic research has demonstrated their validity (Glass et al, 1981; Taylor, 2002). Non-published studies are studies that meet the above criteria for experimental quality but are not included in refereed journals such as a PhD dissertation, or an ERIC article. 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 needed to be either an experiment or a quasi-experiment (quasi meaning that there was a control and an experimental group with a pre- and post-test), (b) the meta-analysis must include all studies written in English from the beginning of experimental research up to and including the year 2012, (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 in the study.

Data Analysis

After conducting an analysis to verify that published studies did not significantly differ from non-published studies in our meta-analysis, we found that the effect sizes were not close to being significantly different ($p = .95$) although, interestingly, published studies had a slightly higher average effect size ($g = .69$) than non-published studies ($g = .67$) (see Appendix A). These results are similar to Taylor's (2002) results where it was found that, in accumulated studies on the effects of L1 glossing on L2 reading comprehension, the effect sizes were almost statistically identical when comparing published and non-published groups of studies.

It should be noted that studies with 50%+ glossing were instant dictionary studies, which means that the student had to either type the lexical item in a computer dictionary (Goyette, 1995) or click on the item in the text, of which all or most was glossed (Stoehr, 1999). All other categories were either clicking on a glossed item or seeing it in the margin in a traditional, paper-based format.

RESULTS

The overall effect size, Hedges, g, of L1 aides on L2 reading comprehension was .68 with considerable heterogeneity among the studies ($Q = 126.422$ $p < .001$). These statistics indicate that generally on posttests, learners with L1 glosses comprehend L2 texts almost three-fourths of a standard deviation above those without L1 glosses. The overall results for the present meta-analysis are found in Tables 1 and 2. Table 1 alphabetically displays the descriptive statistics for each study, indicating the overall average effect size near the bottom of the table. Table 2 reorganizes the table according to common percentages of glossing found in the literature. The positive effect sizes indicate that learners with L1 glosses did better on measures of L2 reading comprehension than those without L1 glosses. A negative effect size means that learners without L1 glosses did better than those with L1 glosses on L2 reading comprehension measures.

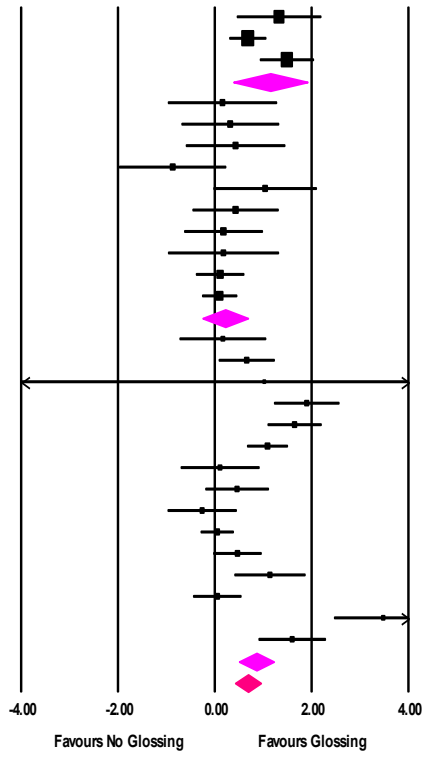
Table 1
*Descriptive Statistics of the Study Reports**

Study	NE	NC	XE	XC	SDE	SDC	D
Al-Jabri, 2009	30	30	8.20	7.90	1.85	1.73	0.17
Aweiss, 1994	24	24	43.50	30.00	21.47	19.42	0.66
Azari, 2012	19	19	25.07	13.03	15.37	5.35	1.05
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 & Good, 2009 Level 1	9	7	2.67	1.43	1.32	.78	1.10
Cheng & Good, 2009 Level 2	12	8	3.25	2.63	1.29	1.51	0.45
Cheng & Good, 2009 Level 3	12	11	3.58	3.36	1.17	1.21	0.19
Cheng & Good, 2009 Level 4	5	7	2.80	2.57	1.10	1.30	0.19
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, 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	0.45
Knight, 1994	54	51	74.01	56.65	27.29	23.35	0.68
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	0.05
Salem, 2006	19	18	18.89	9.06	2.51	3.10	3.50
Stoehr, 1999	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							**.68**

*Note: The effect sizes are not yet corrected or wrighted so they may slightly differ, though not significantly from those effect sizes in later charts.
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
Percentage of Text Glossed

Group by Percent of text glossed	Study name	Point estimate	Lower limit	Upper limit	p-Value
50% or more	Goyette, 1995	1.327	0.445	2.209	0.003
50% or more	Knight, 1994	0.682	0.290	1.074	0.001
50% or more	Stoehr, 1999	1.490	0.922	2.058	0.000
50% or more		1.136	0.369	1.903	0.004
5-10%	Baumann, 1994 Beginning level; Bicycle text	0.160	-0.977	1.297	0.783
5-10%	Baumann, 1994 Beginning level; Breakfast text	0.320	-0.699	1.339	0.538
5-10%	Baumann, 1994 Intermediate level; Bicycle text	0.430	-0.609	1.469	0.417
5-10%	Baumann, 1994 Intermediate level; Breakfast text	-0.870	-1.987	0.247	0.127
5-10%	Cheng, 2009 Level 1	1.040	-0.038	2.118	0.059
5-10%	Cheng, 2009 Level 2	0.430	-0.472	1.332	0.350
5-10%	Cheng, 2009 Level 3	0.180	-0.643	1.003	0.668
5-10%	Cheng, 2009 Level 4	0.180	-0.976	1.336	0.760
5-10%	Jacobs, et al., 1994	0.110	-0.400	0.620	0.672
5-10%	Kwong-Hung, 1995	0.100	-0.272	0.472	0.599
5-10%		0.204	-0.265	0.673	0.394
less than 5%	Al Jabri, 2009	0.166	-0.740	1.072	0.719
less than 5%	Aweiss, 1994	0.660	0.072	1.248	0.028
less than 5%	Azari, 2012	1.024	-6.295	8.343	0.784
less than 5%	Davis, 1989	1.900	1.214	2.586	0.000
less than 5%	Guidi, 2009	1.650	1.082	2.218	0.000
less than 5%	Huang, 2003	1.090	0.659	1.521	0.000
less than 5%	Joyce, 1997 101 level	0.110	-0.713	0.933	0.793
less than 5%	Joyce, 1997 102 level	0.460	-0.206	1.126	0.176
less than 5%	Joyce, 1997 201 level	-0.260	-0.985	0.465	0.482
less than 5%	Ko, 1995	0.050	-0.303	0.403	0.781
less than 5%	Ko, 2005	0.470	-0.040	0.980	0.071
less than 5%	Lou, 1993	1.140	0.395	1.885	0.003
less than 5%	Martinez- Fernandez, 2010	0.052	-0.458	0.561	0.843
less than 5%	Salem , 2006	3.480	2.455	4.505	0.000
less than 5%	Yanguas, 2009	1.600	0.894	2.306	0.000
less than 5%		0.850	0.490	1.210	0.000
Overall		0.675	0.407	0.942	0.000



In Table 2, we can see that the percentage of L1 glosses included in a study may have influenced L2 reading comprehension. Average effect sizes were observed of 1.13 for studies with more than 50% of the text glossed, .85 for studies that had less than 5% of the text glossed and .20 for studies that had 5-10% of the text glossed. There was a significant difference between groups (Q = 6.16; p = .04), which means that the difference between the groups was attributable to something other than random chance. In Table 3, we can see that for each group, the effect sizes have been organized in descending order, from highest to lowest. From Table 3, we can observe that the four highest effect sizes come from the group with the least amount of glossing.

Post hoc analysis further demonstrates that the CALL learning environment may also have played a role in the results. Table 4 displays the studies with the six largest effect sizes in our meta-analysis. Table 4 further shows the type of experiment these top studies performed, whether CALL or traditional, paper-based glossing. Davis (1989) had the only traditional paper-based L1 glossing study in Table 4.

Table 3
Percentage of Glossing Highest Effect Sizes to Lowest in Each Group

Group by Percent of text glossed	Study name	Point estimate	Lower limit	Upper limit	p-Value
50% or more	Stoehr, 1999	1.490	0.922	2.058	0.000
50% or more	Goyette, 1995	1.327	0.445	2.209	0.003
50% or more	Knight, 1994	0.682	0.290	1.074	0.001
50% or more		1.136	0.369	1.903	0.004
5-10%	Cheng, 2009 Level 1	1.040	-0.038	2.118	0.059
5-10%	Baumann, 1994 Intermediate level; Bicycle text	0.430	-0.609	1.469	0.417
5-10%	Cheng, 2009 Level 2	0.430	-0.472	1.332	0.350
5-10%	Baumann, 1994 Beginning level; Breakfast text	0.320	-0.699	1.339	0.538
5-10%	Cheng, 2009 Level 3	0.180	-0.643	1.003	0.668
5-10%	Cheng, 2009 Level 4	0.180	-0.976	1.336	0.760
5-10%	Baumann, 1994 Beginning level; Bicycle text	0.160	-0.977	1.297	0.783
5-10%	Jacobs, et al., 1994	0.110	-0.400	0.620	0.672
5-10%	Kwong-Hung, 1995	0.100	-0.272	0.472	0.599
5-10%	Baumann, 1994 Intermediate level; Breakfast text	-0.870	-1.987	0.247	0.127
5-10%		0.204	-0.265	0.673	0.394
less than 5%	Salem , 2006	3.480	2.455	4.505	0.000
less than 5%	Davis, 1989	1.900	1.214	2.586	0.000
less than 5%	Guidi, 2009	1.650	1.082	2.218	0.000
less than 5%	Yanguas, 2009	1.600	0.894	2.306	0.000
less than 5%	Lou, 1993	1.140	0.395	1.885	0.003
less than 5%	Huang, 2003	1.090	0.659	1.521	0.000
less than 5%	Azari, 2012	1.024	-6.295	8.343	0.784
less than 5%	Aweiss, 1994	0.660	0.072	1.248	0.028
less than 5%	Ko, 2005	0.470	-0.040	0.980	0.071
less than 5%	Joyce, 1997 102 level	0.460	-0.206	1.126	0.176
less than 5%	Al Jabri, 2009	0.166	-0.740	1.072	0.719
less than 5%	Joyce, 1997 101 level	0.110	-0.713	0.933	0.793
less than 5%	Martinez- Fernandez, 2010	0.052	-0.458	0.561	0.843
less than 5%	Ko, 1995	0.050	-0.303	0.403	0.781
less than 5%	Joyce, 1997 201 level	-0.260	-0.985	0.465	0.482
less than 5%		0.850	0.490	1.210	0.000
Overall		0.675	0.407	0.942	0.000

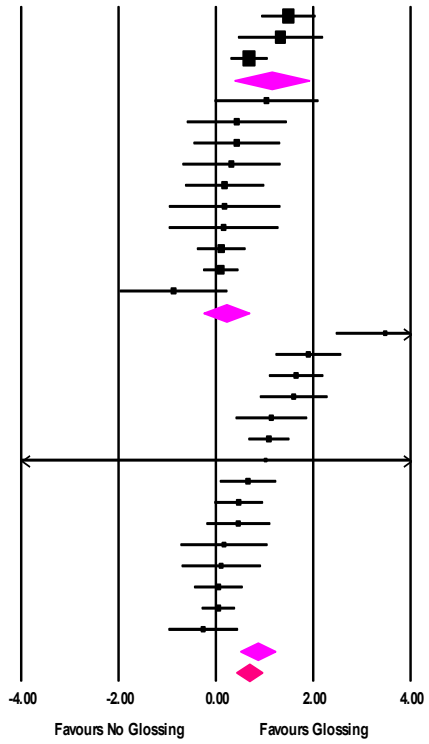


Table 4
Top Six Effect Sizes across Categories, Descending Order

Study	Effect Size	Category	Type of Study
Salem, 2006	3.48	less than 5%	CALL
Davis, 1989	1.90	less than 5%	Traditional Gloss
Guidi, 2009	1.65	less than 5%	CALL
Yanguas, 2009	1.60	less than 5%	CALL
Stoehr, 1999	1.49	more than 50%	CALL
Goyette, 1995	1.33	more than 50%	CALL

DISCUSSION

It is surprising to note that the highest group effect size was found for texts that included the least amount of glossing. However, what was surprising was the finding that the second highest group (and the groups that had the most accurate effect size, based on its much larger size) was the group of the studies with more than 50% of the study glossed. In fact, the top four largest effect sizes in the whole meta-analysis were all from the group with the least amount of glossing. One possible conclusion would be that there may not yet be

enough data from the 50%+ glossing group. It does seem that one might be able to comprehend L2 texts more easily by simply using judicious amount of glossing in textbooks, rather than simply including unlimited glossing. This, of course, does not yet prove that unlimited glossing capability is ineffective; it simply means that one can comprehend L2 texts without the whole text being glossed.

One would think that more glossing would result in more comprehension. However, this is still uncertain based on the lack of CALL studies that have made use of an unlimited glossing capability. Further, perhaps most surprising is the finding that the middle category, the group of studies with 5-10% of glossing, had a much smaller effect size of .20, which is barely within the parameters of having any practical importance. In fact, there is only one study in the 5-10% category that has an effect size of more than 1.0 (Cheng & Good, 2009). The other two groups have several studies with effect sizes over 1.0. The difference here is so striking that it begs the question of whether there is an important threshold in play here. That is, can there be too much glossing, and, if so, is it possible that too much glossing can become a distraction, especially if it is used with L2 readers for whom a text is fairly easy? It depends on how visible glossing is, however, what matters is too much looking up of glossing rather than necessarily the amount of glossing. Traditional glossing that is excessive may have a detrimental effect on L2 reading because glossing is not always needed and actually may have taken the learners' attention away from concentrating on extracting meaning from the text (Taylor, 2010). In other words, while glossing may assist the L2 reader in "noticing the gap," it may also distract the L2 reader from paying enough attention to textual meaning. Another explanation suggested by the data is that the studies with a greater percentage of glossing (i.e., 5-10%) were all paper-based: had those studies been CALL studies, the percentage effect may have been higher, perhaps even proportionate, to the amount of glossing.

CALL Studies

It is interesting that three out of the top four studies (Goyette, 1995; Guidi, 2009; Salem, 2006; Stoehr, 1999; Yanguas 2009) with the highest effect size (all from the 5% or less category) are fairly recent studies that incorporated CALL glossing on a limited basis. As we continue to look at the top six effect sizes, the next two largest effect sizes (after the first four) are also CALL studies, namely, Stoehr (1999) and Goyette (1995). Interestingly, if we had included the Bowles' (2004) CALL study (we were unsure whether she randomly assigned participants to the experimental or control groups), it would have yielded an effect size of 1.73 which would have been the third highest effect size in the present meta-analysis. Thus, if Bowles (2004) had been included in our analysis, six of the top seven effect sizes would have been from CALL studies. This is an important finding because we can also confirm that there is no CALL study in the 5-10% glossing category. Thus, perhaps the discussion should be more about CALL contexts and the usefulness of CALL glossing, rather than percentage of text glosses. Or, perhaps they both play significant roles in L2 reading comprehension.

As previous research has posited (e.g., Taylor, 2006, 2009) CALL glossing may not be as large of a distraction as other types of glossing because the glosses can be easily hidden with a hyperlink such as in the Yanguas (2009) study, which describes how glosses are used: "As a result of the pilot test 21 words were glossed. In the experimental conditions, the words were hyperlinked. When the participants clicked on them a box appeared above the word with a definition in English..." (p. 54). It may be important to note that the gloss appeared *above* the word. Thus, there may be much less distraction than if the eyes need to be rerouted to the margin or to the bottom of the page. This is a benefit of CALL glosses; they can be less distracting because not only is the item consulted if needed, it also can appear relatively nearer to where the reading is actually occurring. Of course there may be factors other than the placement of the glosses that may be influencing the results.

Salem's experiment (2006), the study with the largest effect size in our meta-analysis, used a slightly different way of glossing to that of Yanguas: the L1 glossed item came up on the left-hand side of the page if the L2 readers clicked on the hyperlink. Salem's argument was that the gloss could be viewed while the text was consulted and compared (Salem, 2006). This type of glossing may be more conducive in helping students notice the gap between their own language and more native-like usage as suggested by Schimdt (1990, 1994). It appears that both the Yanguas and the Salem type of glossing are effective in CALL contexts, with perhaps the Salem idea of glossing being more effective. In both cases, we should remember that CALL glosses do not appear unless the L2 reader wants them to appear. There may be something very significant about this aspect of CALL L1 glossing. In other words, CALL glosses may be more conducive to L2 reading because they can become useful when attentional resources are brought into play under the reader's own control (Taylor, 2006, 2009, 2010).

Salem's (2006) study is intriguing on several levels, not just on the finding of CALL glossing. He also found that only having L1 glosses was the most effective way of improving reading comprehension, even when adding additonal glosses such as L1 textual glosses plus audio glosses, or textual, audio pictorial glosses, or even a combination of textual, audio, pictorial and writing down the gloss consulted. Fascinatingly, as more glosses were added to the treatment, the lower the scores became. In other words, the highest mean was simply for L1 textual glosses and the means got progressively lower as Salem added more and more glossing helps for each item. This seems counterintuitive, since one would assume that more types of glossing might help with more retention, perhaps relating to different types of learning styles, but the data does not appear to bear out this idea. Guidi's (2009) findings were similar as were the instructions in the study:

> In this task, you will read an excerpt talking about Argentineans, and their culture and traditions. On the upper section of the screen, you will see the reading passage. Some words and phrases of the text have been underlined, and translated for you. **The English translation of these words and phrases appear at the bottom of the screen. Each time you find an underlined word or phrase in the text, please find the referenced translation at the bottom of the screen (p. 24, emphasis added).**

Thus, Guidi (2009), whose study had the third highest effect size of our 28 study reports, found that glossing works. Again, we suggest that it is possible that L1 glossing works in Guidi's study because the items are not consulted unless the L2 reader actually clicks on the words in question in which case the learner's attention is actively brought to the item. Interestingly, Stoehr (1999), albeit from the 50% or more access group of studies, obtained similar results in her study which demonstrated that CALL L1 glosses were superior to CALL L2 glossing, CALL L2 paraphrasing, as well as traditional paper glossing. It is important to note in the Stoehr (1999) study that not all glosses were actually consulted; only those that L2 readers elected to consult by clicking on a link.

Is there a threshold of glossing that is generally more helpful than other levels? The analysis suggests that a glossing amount greater than 5% may not always be helpful to the students, especially in CALL contexts. It should also be pointed out that CALL glosses, even if they have 50% or more access for the L2 reader, do not mean that students will actually use the glosses that much. Very likely, the L2 reader only accesses the amount of glossing absolutely necessary for a basic understanding of the text. Thus, even if we were to see that a CALL study has unlimited glossing, this does not mean that all items will be consulted. Of course, Stoehr's (1999) study demonstrates that there is a significant difference between groups with L1 glosses and no glossing in terms of how much time was spent ($p < .001$). Salem's (2006) study showed that L2 readers in the L1 glossing group did indeed spend more time on reading than the no glossing group, but less time than the other groups which included extra glossing helps. However, these extra items and extra time spent did not result in superior comprehension than the L1 textual glossing group in Salem's (2006)

study. Such results suggest that noticing, as defined by spending time, while important and sometimes significant, does not always result in superior reading comprehension.

CONCLUSION AND PEDAGOGICAL IMPLICATIONS

So what do these results mean? They may mean that abundant glossing is over-rated. Since even in CALL unlimited glossing form, there is little doubt that glosses are not consulted in every case, it leaves open the possibility that glosses may be helpful for L2 readers who already have at least one semester of L2 learning (Joyce, 1997). It could be suggested that, perhaps most significantly, CALL glossing is less intrusive because items do not appear on the page until they are actually consulted and attentional resources are diverted towards them. Proof of this is seen in the data which uncovers the counterintuitive conclusion that the largest effect sizes have limited CALL glossing. Of course, it may not be reasonable to expect the L2 reader to consult a text that is so difficult such that more than 10% glossing is needed. Very high effect sizes were found in CALL studies that only had targeted glossing (Yanguas, 2009) or large amounts of glossing that the learning could select (Stoehr, 1999). In both the Yanguas (2009) and Stoehr studies, however, I suspect that a similar actual amount of glosses were consulted, since the Stoehr (1999) study had instant lookup glosses whereas the Yanguas study had links the participant could consult. It seems reasonable that a student would not consult more glosses than necessary to basic understanding of an L2 text.

Glossing should assist L2 reading comprehension, and as our study has shown, this is not necessarily accomplished with large volumes of glossing. With large amounts (50% or more) of glossing available, it would no doubt be difficult to read the average L2 text if all available glosses were consulted every time; furthermore, even if an entire text is hyperlinked, it would be extremely laborious and distracting to click on every item while attempting to read the text. Thus, while CALL glossing is very likely effective, the very flexibility that it offers can be detrimental if consulted too often with a text that may be too difficult for the L2 reader.

LIMITATIONS AND FUTURE RESEARCH

The present meta-analytic study is limited in that we have only directly addressed one (albeit an important one) of the variables, the percentage of text glossed, that may be influencing L1 glossing studies. Since the present study has done much of the work of finding the studies and extracting the effect sizes, further research should study whether the location of the appearance of the L1 glossed item can significantly improve the level of L2 reading comprehension. Worthy of further study is the idea that perhaps glosses appearing right over the item in question may be more effective than glosses that appear to the left or right of the margin. Perhaps covering the glossed item in CALL glossing would be less effective than having the item appear on the bottom or side of the page. Further studies, with a similar body of research to that in the present meta-analysis, could make use of at least some of the findings and data to take future research in new directions, such as studying the effects of how L2 reading comprehension is measured in studies, and how the results can influence the results of studies conducted on L1 glossing and L2 reading comprehension. Other topics may include the type of text chosen to test the L2 readers. It is possible that authentic texts may be more conducive to the use of L1 glossing than adapted texts or vice versa. In general, it is likely that the most influential variables in L1 glossing studies have something to do with the type of glossing or the type of testing. Thus, the way in which the independent and dependent variables are manipulated likely account for the greater part of the variance in scores across studies.

NOTES

[1] Bowles (2004) did not randomly assign participants to their respective groups so we cannot be assured of whether there was a preexisting significant difference between the groups according to their competency level. Lomicka (1997) used L1 and L2 glosses at the same time. Palmer (2003) and Jacobs (1994) were not included because they contained over 10% glossing which no other study had in the present meta-analysis. Farvardin and Bira (2012) did not have a control group. Coriano Velázquez (2001) did not randomly assign the participants nor were the participants told about the glossing in the study. In general, there were many studies that used multiple glosses (see Abraham, 2007 for a review). Prichard and Matsumoto (2011) did not randomly assign nor pretest reading comprehension. Chun & Plass (1996) included several types of glossing at the same time.

[2] The studies included in our meta-analysis are marked with an asterisk in the References section below.

REFERENCES[2]

Abraham, L. B. (2007). Second language reading comprehension and vocabulary learning with multimedia. *Hispania, 90,* 98-109.

*Al-Jabri, S. S. (2009). The effects of L1 and L2 glosses on reading comprehension of recalling ideas by Saudi students. *Umm Al-Qura University Journal of Social Sciences*, *1*, 10-27. Retrieved from http://libback.uqu.edu.sa/hipres/MAGZ/3500001-6.PDF

*Aweiss, S. (1994). Situating learning in technology: The case of computer-mediated reading supports. *Journal of Educational Technology Systems, 23,* 63-74. doi: 10.2190/FWEF-TH4H-495C-YF45

*Azari, F. (2012). Gloss or no gloss? EFL learners' preference. *American International Journal of Social Science*. *1*, 75-85.

*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. doi: 10.2307/346199

Bernhardt, E. B. (1983). Testing foreign language reading comprehension: The immediate recall protocol. *Die Unterrichtspraxis, 16,* 27-33. doi: 10.2307/3530598

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. doi: 10.1093/applin/16.1.15

Bowles, M. A. (2004). L2 glossing: To CALL or not to CALL. *Hispania, 87,* 541-552. doi: 10.2307/20063060

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

*Cheng, Y., & Good, R. L. (2009). L1 glosses: Effects on EFL learners' reading comprehension and vocabulary retention. *Reading in a Foreign Language*, *21*, 119-142.

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

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. doi: 10.3138/cmlr.57.3.402

*Coriano Velázquez, A. (2001). Vocabulary acquisition through reading: A study of the effectiveness of different CALL based annotations. *Master's Abstracts International*, *40*, 561. (UMI No. 1406590)

*Davis, J. N. (1989). Facilitating effects of marginal glosses on foreign language reading. *The Modern Language Journal, 73,* 41-48. doi: 10.1111/j.1540-4781.1989.tb05308.x

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.

Farvardin, M. T., & Biria, R. (2012). The impact of gloss types on Iranian EFL students' reading comprehension and lexical retention. *International Journal of Instruction, 5*, 99-114.

Gass, S. M., & Selinker, L. (1994). *Second language acquisition: An introductory course.* Earlbaum: New Jersey.

*Goyette, E. S. (1995). *The effects of dictionary usage on text comprehension* (Doctoral dissertation, McGill University, Montreal, Quebec, Canada). Retrieved from http://digitool.library.mcgill.ca/R/QLPEN4Q98ALTDL54UK5LX2672IKRSBCPTE3VKIISRGENVJBY6Q-05129?func=results-full

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

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, Philadelphia, PA.

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. doi: 10.1111/j.1467-9817.1994.tb00049.x

*Joyce, E. E. (1997). Which words should be glossed in L2 reading materials? A study of first, second and third semester French students' recall. *Pennsylvania Language Forum*, 58-64. (ERIC Document Reproduction Service No. ED 427508)

Kern, R. G. (1994). The role of mental translation in second language reading. *Studies in Second Language Acquisition, 16,* 441-461. doi: 10.1017/S0272263100013450

*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. doi: 10.1111/j.1540-4781.1994.tb02043.x

*Ko, M. H. (1995). *Glossing in incidental and intentional learning of foreign language vocabulary and reading comprehension* (Unpublished master's thesis). University of Hawai'i, Manoa, Oahu.

*Ko, M. H. (2005). Glosses, comprehension, and strategy use. *Reading in a Foreign Language*, *17*, Retrieved 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 master's thesis). University of Hong Kong, Pokfulam, 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., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*, 15-30.

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, Philadelphia, PA.

*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, Washington, DC.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63,* 59-82. doi: 10.3138/cmlr.63.1.59

*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, Indiana, PA.

Prichard, C., & Matsumoto, Y. (2011). The effect of lexical coverage and dictionary use on L2 reading comprehension. *The Reading Matrix, 11,* 207-225.

*Salem, E. B. (2006). *The influence of electronic glosses on word retention and reading comprehension with Spanish Language Learners* (Unpublished doctoral dissertation). University of Kansas, Lawrence, KS.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 1,* 129-158. doi: 10.1093/applin/1.2.129

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.

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language. In R. Day (Ed.), *Talking to learn* (pp. 237-326). Rowley, MA: Newbury House.

*Stoehr, L. E. (1999). *The effects of built-in comprehension aids in a CALL program on student-readers' understanding of a foreign language* (Unpublished doctoral dissertation). University of Texas, Austin, TX.

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.

Taylor, A. (2002). *A meta-analysis on the effects of L1 glosses on L2 reading comprehension* (Unpublished doctoral dissertation). Purdue University, West Lafayette, IN.

Taylor, A. (2006). The effects of CALL versus traditional L1 glosses on L2 reading comprehension. *CALICO Journal, 23*, 1-10.

Taylor, A. (2009). CALL-Based versus paper-based glosses: Is there a difference in reading comprehension? *CALICO Journal, 27*, 147-160.

Taylor, A. (2010). Glossing is sometimes a distraction: Comments on Cheng and Good (2009). *Reading in a Foreign Language,* 22, 353-354.

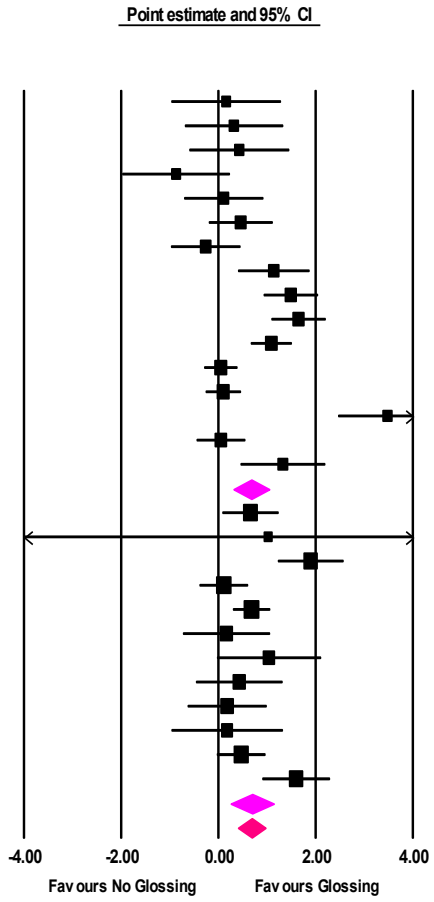
Taylor, A. (2013). CALL versus paper: In which context are L1 glosses more effective? *CALICO Journal, 30*, 63-81. doi: 10.11139/cj.30.1.63-81

*Yanguas, I. (2009). Multimedia glosses and their effects on L2 text comprehension and vocabulary learning. *Language Learning and Technology, 13,* 49-67.

APPENDIX A

Published versus Non-published Studies

Group by Published?	Study name	Statistics for each study: Point estimate	Lower limit	Upper limit	p-Value
No	Baumann, 1994 Beginning level; Bicycle text	0.160	-0.977	1.297	0.783
No	Baumann, 1994 Beginning level; Breakfast text	0.320	-0.699	1.339	0.538
No	Baumann, 1994 Intermediate level; Bicycle text	0.430	-0.609	1.469	0.417
No	Baumann, 1994 Intermediate level; Breakfast text	-0.870	-1.987	0.247	0.127
No	Joyce, 1997 101 level	0.110	-0.713	0.933	0.793
No	Joyce, 1997 102 level	0.460	-0.206	1.126	0.176
No	Joyce, 1997 201 level	-0.260	-0.985	0.465	0.482
No	Lou, 1993	1.140	0.395	1.885	0.003
No	Stoehr, 1999	1.490	0.922	2.058	0.000
No	Guidi, 2009	1.650	1.082	2.218	0.000
No	Huang, 2003	1.090	0.659	1.521	0.000
No	Ko, 1995	0.050	-0.303	0.403	0.781
No	Kwong-Hung, 1995	0.100	-0.272	0.472	0.599
No	Salem , 2006	3.480	2.455	4.505	0.000
No	Martinez- Fernandez, 2010	0.052	-0.458	0.561	0.843
No	Goyette, 1995	1.327	0.445	2.209	0.003
No		0.667	0.300	1.035	0.000
Yes	Aweiss, 1994	0.660	0.072	1.248	0.028
Yes	Azari, 2012	1.024	-6.295	8.343	0.784
Yes	Davis, 1989	1.900	1.214	2.586	0.000
Yes	Jacobs, et al., 1994	0.110	-0.400	0.620	0.672
Yes	Knight, 1994	0.682	0.290	1.074	0.001
Yes	Al Jabri, 2009	0.166	-0.740	1.072	0.719
Yes	Cheng, 2009 Level 1	1.040	-0.038	2.118	0.059
Yes	Cheng, 2009 Level 2	0.430	-0.472	1.332	0.350
Yes	Cheng, 2009 Level 3	0.180	-0.643	1.003	0.668
Yes	Cheng, 2009 Level 4	0.180	-0.976	1.336	0.760
Yes	Ko, 2005	0.470	-0.040	0.980	0.071
Yes	Yanguas, 2009	1.600	0.894	2.306	0.000
Yes		0.686	0.244	1.128	0.002
Overall		0.675	0.392	0.957	0.000



AUTHOR BIODATA

Alan Taylor conducts research in applied linguistics and teaches French at BYU-Idaho. His primary interests include L2 reading comprehension and meta-analytic methods. He has published several large-scale meta-analyses in the *CALICO Journal*.

AUTHOR ADDRESS

tayloral@byui.edu