VOCABULARY GLOSSING: A META-ANALYSIS OF THE RELATIVE EFFECTIVENESS OF DIFFERENT GLOSS TYPES ON L2 VOCABULARY ACQUISITION

by **Vahideh Sadat Vahedi**

Ferdowsi University of Mashhad, International Campus,

Azadi Sq., Mashhad, Khorasan Razavi, Iran

Vahedi2010 @ yahoo.com

Behzad Ghonsooly (Corresponding author) and **Reza Pishghadam**

Ferdowsi University of Mashhad,

Azadi Sq., Mashhad, Khorasan Razavi, Iran

Ghonsooly @ um.ac.ir, Pishghadam @ um.ac.ir

Abstract

In recent years there has been a growing interest to incorporate hypertext glosses into L2 reading materials and accordingly, it has provoked researchers to uncover to what extent and under which moderator variables a specific type of gloss yields more effective outcomes than other types of glossing. The present meta-analysis attempts to examine the magnitude of the effect of different gloss types (single vs. multiple glosses) on L2 vocabulary acquisition along with identification of the contextual factors that influence between-study variation through synthetizing 34 primary articles which satisfy the inclusion criteria. The overall effect size is found to be + 0.83 ($p<0.05$), indicating that multiple glosses (text+visual) has a large, positive effect on learners' vocabulary acquisition than single mode of glossing (text-only). Moderator analyses further suggest that intensity of the program and L2 proficiency level are potential moderators influencing the heterogeneity between effect sizes, whereas the learning context, sample size and research design do not have such impact.

Keywords: hypertext gloss, vocabulary acquisition, meta-analysis, effect size, heterogeneity

1. Introduction

Nowadays, computers have turned to be a familiar sight in the 21st century classrooms and technology is used to enrich many educational tasks (Ortega, 1997). Since the 1960s, specialists in the field of educational technology have tried to accelerate the process of technology-pedagogy integration through developing programs on the basis of Computer-Based Instruction (CBI) to drill, instruct and evaluate students. The dream of technology

revolution in educational settings envisaged through technology-pedagogy integration is now almost five decades old. “Some envisage a day when computers will serve all children as personal tutors: a Socrates or Plato for every child of the 21st century” (Kulik & Kulik, 1991, p. 75).

In the field of ESL/ EFL, many language educators have also tried to increasingly incorporate CBI programs into their classes as a supplement or replacement for traditional teaching methods. In fact, due to the versatility, flexibility and adaptability of computer technology, it is viewed as a promising language learning tool. In this vein, because of the importance of vocabulary acquisition as a key ingredient in the process of language learning, considerable attention has been directed to investigation of techniques used to facilitate the acquisition of new target words.

Among all, glossing has gained popularity and proved to be facilitative in the process of L2 vocabulary acquisition. The application of glosses has recently become a common approach in enriching L2 reading materials. A hypermedia or hypertext gloss refers to “short definitions or explanations with nonlinearly linked data associated with text, graphics, audios, and videos in computerized text” (Kommers, Grabinger & Dunlap, 1996 cited in Yun, 2011). Compared to dictionary definitions, glosses seem to be more preferable because readers are not interrupted during the reading process when they consult with the available definitions in the text (Yanguas, 2005). Moreover, based on Schmidt’s (1995) *Noticing Theory*, one of the prerequisites of learning is conscious attention to target items and attempts should be made to make the target items noticed. In this regard, glossing is considered to be effective for increasing noticing and probably improving vocabulary acquisition among ESL/EFL learners (Nation, 2002; Yoshii, 2006). Glosses also enhance the learners ‘autonomy during the flow of reading (Nation, 2002).

Accordingly, since there is no doubt about the advantages of glossing for incidental vocabulary acquisition, many language specialists have shifted their focus from gloss effect to gloss type. The review of the studies on the effects of different gloss types (text, visual, audio, etc.) has brought inconsistent results, some have revealed no differences between them (e.g. Lomicka, 1998; Ben Salem & Aust, 2007) while others have shown the advantages of one gloss type over the other (Chun & Plass, 1996; Al-Seghayer, 2001) in vocabulary acquisition. Clearly, there exists no consistency among the results of the research studies investigating multiple and single gloss types, and the effect of different

modes of presentation of glosses on vocabulary acquisition still remains an open research area that needs further exploration.

Regarding this dilemma, when there is a need to draw general conclusions, as Kulik, Kulik, and Shwalb (1986) emphasize it, it is necessary to integrate the results from a variety of settings under different conditions and to apply tools of research synthesis to the results of the individual studies. Taking benefit of a research method known as *Meta-Analysis* (Glass, 1977; Hedges and Olkin, 1985), it is possible to examine the effectiveness of different gloss types – single vs. multiple glosses – on vocabulary acquisition by integrating and analyzing the results of numerous studies, taking into account all the variables which might be effective. Such a systematic review can also provide a map of the past and current research in the field of textual glosses and vocabulary acquisition, showing the pathway for future studies. Owing to the conflicting findings reported by existing research in this area, and the fact that very little, if any, comprehensive systematic analysis of the effects of different types of glosses has been done, the present study attempts to provide a quantitative review of the effectiveness of single textual glosses as compared to multiple glosses examining methodological and substantive features which are considered to be influential on the overall effectiveness of textual glosses for enhancing vocabulary acquisition.

The following research questions guided this research study:

1) What is the overall effectiveness of multiple gloss type (text+visual) on vocabulary acquisition?
2) Is there a statistically significant heterogeneity among effect sizes of the studies investigating the effect of different gloss types on vocabulary acquisition in the present meta-analysis?
3) In the case of statistically significant heterogeneity among effect sizes, what are the potential moderators to systematically account for the heterogeneity among the effect sizes of primary studies investigating the effect of different gloss types on vocabulary acquisition?

2. Literature review

In recent years, the application of glosses has increased tremendously, fueling a debate over whether or not they are an effective means of improving students' vocabulary acquisition. Even though a large body of research exists investigating the effectiveness of

hypertext glosses (Ariew, 2006; Nagata, 1999; Akbulut, 2007; Al-Seghayer, 2001; Chun & Plass, 1996; Lomicka, 1998; Robin, 2007; Nikolova, 2004; Khan, 1997; Plass, Chun, Mayer & Leutner, 1998), there remains one important question, namely whether the meanings, when provided, should be presented in a single mode (text-only) or with a combination of multimedia-embedded features such as video, picture, sound, etc.

The results of many research studies concerning whether different types of hypertext glosses enhance vocabulary acquisition of L2 learners have been somewhat inconclusive. Plass et al. (1998), for instance, investigated the effect of different modes of glossing on vocabulary acquisition in a multimedia environment. It was found that subjects who had selected verbal and visual glosses performed significantly better than those who had selected one mode of glossing.

In a similar study, Kost, Fost and Lenzini (1999) examined the effect of textual and pictorial glosses on vocabulary acquisition in printed texts. They found that learners who had access to textual and pictorial glosses outperformed their counterparts who had access to textual glosses alone. They hypothesized that this finding is the result of different degrees of cognitive effort employed by learners to process information. They remarked that "the mapping of pictures onto the mental model provides and stronger bond than the mapping of words due to the different representations of their information (analog vs. symbolic)" (p. 94).

In another study conducted by Al-Seghayer (2001), the effect of different gloss types on L2 vocabulary retention was examined. Results of the study showed that subjects' performance on text-plus-video was significantly better than their performance on the single mode of glosses.

Yoshii & Flaitz (2002) further investigated the effect of different gloss types on L2 vocabulary acquisition in a multimedia reading environment. They found that the text plus picture group significantly performed better than the group who had access to textual glosses alone.

Yeh and Wang (2003) examined the effect of three modes of multimedia glosses: text only, text plus picture, and text, picture, and sound. They reported that text plus picture was the most effective gloss type for vocabulary acquisition.

More recently, Akbulut (2007) compared the performance of subjects who had access to text-only glosses with those with access to text+visual glosses on vocabulary test. The findings of the study revealed a significant difference between text-only and

text+visual gloss groups on vocabulary test; however, no significant difference was found on reading comprehension performance of the two groups.

Despite such positive outcomes of application of different modes of glosses, some researchers report no significant differences between the performance of the groups using text-only versus text+visual glosses on L2 vocabulary tests.

In a study conducted by Ben Salem and Aust (2007), it was found that learners who used glosses had significantly higher reading comprehension and vocabulary acquisition scores than non-gloss users; however, no significant difference was reported between the performance of text-only and text+visual gloss users.

Furthermore, the results of the study conducted by Lomicka (1998) revealed no significant difference between the vocabulary scores of those who consulted with text+visual glosses as compared to those who had access to text-only mode of glosses.

To sum up contradictory evidence provided by existing studies in this area and reach general conclusions, conducting the current meta-analytic review seems to be promising. The results of the present meta-analysis may also shed light on the variation between primary studies' results more explicitly through examining methodological and substantive features considered to be influential on the overall effectiveness of single and multiple glosses for enhancing vocabulary acquisition.

3. Method

3.1. Literature accumulation

A literature search was run using Scopus, Web of Science, and Google research databases in order to identify almost all relevant studies examining the relative effectiveness of glosses either in single (text-only) or multiple (text+visual) mode of presentation on enhancing learners' vocabulary acquisition, published until 2014 when this meta-analysis was conducted. After obtaining a preliminary set of articles including 95 studies, references used in the selected journals and article citations were also examined. Concurrently, specialists in the field of CALL were also consulted for articles that had not been gathered by the primary search procedures. The whole process resulted in the identification of 105 articles which were of potential relevance to the present meta-analysis. Subsequently, all gathered articles were scrutinized to see whether they satisfied the following inclusion criteria to be included in the final analysis.

3.2. Inclusion and exclusion criteria

Studies which are included in the present meta-analysis data set were required to satisfy the following inclusion and exclusion criteria:

- Studies should have a quasi-experimental or experimental design, investigating experimentally the effect of single and multiple hypertext glosses on vocabulary acquisition in an ESL/ EFL context.
- Studies written in English and published until 2014 will be included in the data set.
- Studies should include sufficient data for calculating the effect size like combinations of means, standard deviations, t-test or ANOVA statistics, group sizes, etc.
- The study must be available in its full text form.
- Uncontrolled experiments and anecdotal reports or self-assessment of improvement will be excluded.
- Studies should be free from serious methodological weaknesses like a) unfair teaching of the content assessed in the posttest to one of the groups (experimental or control); b) non-randomly assigning participants to control and experimental groups; c) neglecting the pre-existing differences between the groups, etc.

Following these criteria, the abstracts of the gathered articles are scrutinized in the first instance to determine if the experiments involved the application of text+visual and text-only glosses for vocabulary acquisition. Unsuitable abstracts (no textual gloss or no vocabulary acquisition, 38 articles), non-research articles (5 articles), publications in languages other than English (4 articles) are not further reviewed. The full texts of all remaining articles were comprehensively checked based on the identified inclusion and exclusion criteria to assess their qualification to be included in the present meta-analysis. Out of these, articles lacking the minimum information (e.g. sample size, mean, standard deviation, etc.) were excluded from the analysis (24 articles). In total, 34 primary articles satisfied all the inclusion criteria and went under the systematic analysis.

3.3. Reliability of coding

After gathering the relevant studies and excluding those that did not satisfy the pre-determined criteria for being included in this meta-analysis, the next step was to prepare a coding table, based on which the common features among different studies can be identified and organized. Cooper (2010) has proposed some common features that should

be included in every coding scheme. According to the researcher, "every coding sheet should include the variables of investigation, participants' characteristics, research design and statistical information of every individual study" (Alsadhan, 2012, p. 43).

To increase the reliability of coding, two independent coders coded each individual study involved in the data set. Then, the two coders jointly reviewed their coding forms and, in the case of some disagreements, they re-evaluated the points of dispute.

3.4. Data analysis

3.4.1. Effect size

If continuous measures are used in meta-analysis, standardized difference between the means of the experimental and the control group, often called Hedges's *g* (Hedges & Olkin, 1985, p. 78) is typically used as a measure of effect size to summarize the finding of each individual study. The underlying reason is that the quantitative results from the different primary studies are converted into a standardized metrics that allows for meaningful numerical comparison across studies.

$$g = \frac{\bar{X}_1 - \bar{X}_2}{S_p} \times J$$

Where $\bar{X}_1$ is the mean of the experimental condition (text+visual) and $\bar{X}_2$ is the mean of the controlled condition (text-only), J is a correction factor for small sample bias, and S_p is the pooled standard deviation calculated based on the following formula where n_1 and S_T stand for experimental group sample size and standard deviation and n_2 and S_C for control group sample size and standard deviation respectively. The reason underlying such correction is that effect size of a treatment for small samples tends to be overestimated. To avoid this, the formula is multiplied by J factor to correct the upward bias.

$$s_p = \sqrt{\frac{(n_1 - 1)s_T^2 + (n_2 - 1)s_C^2}{n_1 + n_2 - 2}}$$

$$J = 1 - \frac{3}{4(n_1 + n_2 - 2) - 1}$$

Effect sizes are evaluated with reference to their 95% confidence interval and related *p* values. A 95% confidence interval indicates that if the same study is replicated several times, the effect size estimate would be within that range 95% of the time. True effect size is estimated to be included in this range. Confidence interval ranges that include zero show that the obtained effect sizes are considered to be statistically non-significant (Norris & Ortega, 2000); if the confidence interval ranges do not contain the value zero, they are considered as statistically significant. As the confidence interval becomes

narrower, more trustworthy effect size is revealed. In this meta-analysis, Cohen's (1988) rules were used to interpret the magnitude of effect size: .20 is considered a small effect, .50 a moderate effect and .80 a large effect.

3.4.2. Estimation and test of residual heterogeneity

In primary studies, the dispersion among scores is quantified based on standard deviation and variance of the scores. In a meta-analysis, the process of identifying and quantifying the *heterogeneity* in effect sizes is quite similar since it is described based on the standard deviation and variance. The only factor that makes the process more complicated is that in a meta-analysis the variance is intended to be found between *true* effect sizes, while the variation that we actually observe is a combination of both *(true) heterogeneity plus random error*.

In the cases that all studies share the same true effect sizes, the (true) heterogeneity is zero. In such cases, if we deal with different observed effect sizes, it may be due to the within- study error. In a random- effects model, we assume that the *true effect size does vary from one study to the next.* In this case, if we deal with different observed effect sizes, it may be due to the (true) heterogeneity in effect sizes as well as within-study error.

For comparability reasons, I^2 value seems to be more informative, revealing the percentage of heterogeneity between effect size estimates that is due to the between-study level variation rather than random error alone (Higgins et al., 2003; Higgins & Thompson, 2002), having $I^2 = 100\% - (1 - df/Q)$. While Q-value is used to test the null hypothesis that there is no dispersion across effect sizes, I^2 quantifies this dispersion and can be compared across various meta-analyses with different numbers of studies and different set of moderator variables. To interpret the degree of the heterogeneity between effect sizes, the following rule can be used: $I^2 = 0\%$ stands for no heterogeneity, $I^2 = 25\%$ stands for low heterogeneity, $I^2 = 50\%$ stands for moderate heterogeneity and $I^2 = 75\%$ stands for high heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003).

3.4.3. Outlier analysis

If the variances observed between the effect sizes in the data set seem to be approximately large as compared with the error variances, one possible underlying reason for such heterogeneity may be due to the existence of outliers. To detect any outlier in the data set, the stability of the results can be investigated through leaving out some studies one at a

time and re-doing each meta-analysis for the reduced data set. Accordingly, outliers which have particular impact on effect size estimates can be identified and analyzed.

3.4.4. Publication bias

The existence of publication bias in any meta-analysis is usually assessed using funnel plots. A funnel plot provides scatter plots of the treatment effects obtained from individual studies against a measure of study size or standard error (Sterne, Becker, & Egger, 2005). In cases where publication bias does not exist, the effect size estimates will be symmetrically distributed around the overall mean effect size. In such cases, more precise effect size estimates tend to cluster more closely around the mean effect size. Ignoring the studies which report the most positive or negative effect sizes and even non-significant ones due to publication bias would undoubtedly result in the asymmetry in the relationship between effect sizes and related standard errors or sample sizes.

3.5. Instrument

The software used for the systematic analysis of data in the current study was *Comprehensive Meta-Analysis, version 2*. As a program developed by a team of the most famous experts in the field of meta-analysis, it offers various options for data entry, analysis, and display. Data in more than 100 formats, for example, can be entered for calculating the effect size estimate. Moreover, multiple study designs can be included in the same analysis. That is, data from studies that used independent groups, paired/matched designs can be entered into the program, which has the potential to analyze different design simultaneously.

4. Findings

The comprehensive literature search yielded 105 potentially qualified primary studies on the comparative effects of single and multiple glosses on vocabulary acquisition. Among these, 34 articles could satisfy the inclusion criteria for being included in the present meta-analysis.

4.1. The overall effect size estimate

To measure the magnitude of effects of different gloss types on L2 vocabulary acquisition, two groups – a treatment (text + visual) and a control (text-only) group – were compared

together in the present meta-analysis and effect sizes for individual studies as well as the overall effect size estimate were calculated. The overall weighted mean effect size of 34 weighted effect sizes (Hedge's *g*) was +0.839 (Table 1). According to Cohen's rule of interpretation of the effect size magnitude (1988), the reported effect size was largely positive for L2 vocabulary acquisition.

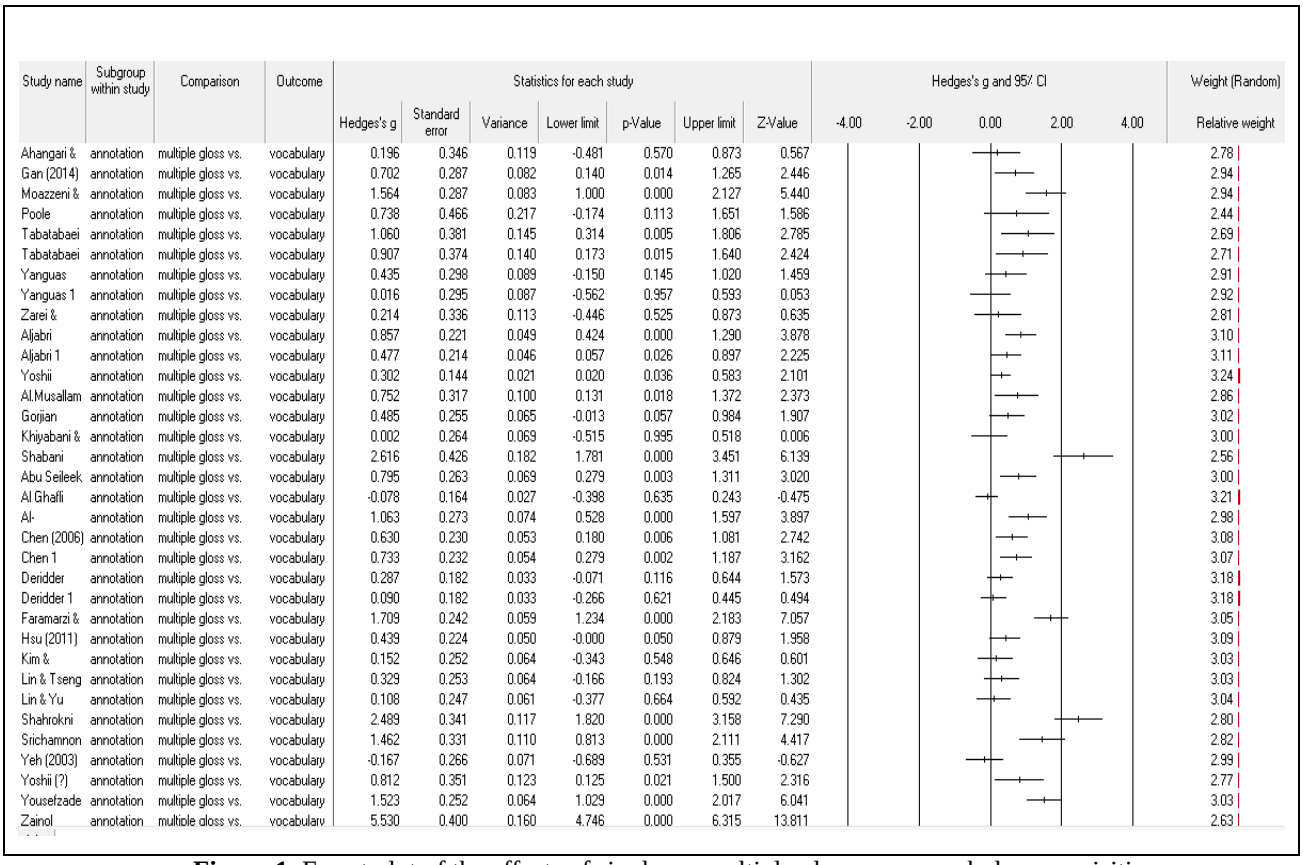
Table 1. Overall effect size estimate based on 34 primary studies.

Model	Effect size and 95% confidence interval						Test of null (2-tail)		Heterogeneity				Tau-squared		
Model	Number Studies	point estimate	standard error	variance	lower limit	upper limit	Z-value	P-value	Q-value	df(Q)	P-value	I^2	Tau Squared	variance	tau
Fixed	34	0.63	0.04	0.002	0.54	0.71	14.32	0.000	327.46	33	0.000	89.91	0.59	0.03	0.77
random	34	0.83	0.14	0.02	0.56	1.11	5.93	0.000							

*p<0.05

As reported above, a 95% confidence interval ranges from 0.562 to 1.116. This confidence interval range does not include the value zero, suggesting that multiple-gloss hypertexts (text + visual) did improve acquisition of L2 vocabulary items. That is, the group which was provided with multiple hypertext glosses (text+visual) outperformed the control group with access to a single gloss (text-only) on a vocabulary test.

To present a more informative picture of each individual study's statistics, the following forest plot (Figure 1) has been depicted. It is worth mentioning that points to the left of the line show the outperformance of the single gloss (text-only) group, whereas points to the right of the line reveal the performance of the multiple glosses (text+visual) group. The upper and lower 95% confidence interval (CI) range for each effect size is represented through the upper and lower limit of the line. At the right side of the figure, the weight assigned to each study is shown pictorially. The size of each square shows the relative weight of each study to the overall effect size estimate, where larger squares indicate greater weight.



Study name	Subgroup within study	Comparison	Outcome	Hedges's g	Standard error	Variance	Lower limit	p-Value	Upper limit	Z-Value	Weight (Random) Relative weight
Ahangari &	annotation	multiple gloss vs.	vocabulary	0.196	0.346	0.119	-0.481	0.570	0.873	0.567	2.78
Gan (2014)	annotation	multiple gloss vs.	vocabulary	0.702	0.287	0.082	0.140	0.014	1.265	2.446	2.94
Moazzeni &	annotation	multiple gloss vs.	vocabulary	1.564	0.287	0.083	1.000	0.000	2.127	5.440	2.94
Poole	annotation	multiple gloss vs.	vocabulary	0.738	0.466	0.217	-0.174	0.113	1.651	1.586	2.44
Tabatabaei	annotation	multiple gloss vs.	vocabulary	1.060	0.381	0.145	0.314	0.005	1.806	2.785	2.69
Tabatabaei	annotation	multiple gloss vs.	vocabulary	0.907	0.374	0.140	0.173	0.015	1.640	2.424	2.71
Yanguas	annotation	multiple gloss vs.	vocabulary	0.435	0.298	0.089	-0.150	0.145	1.020	1.459	2.91
Yanguas 1	annotation	multiple gloss vs.	vocabulary	0.016	0.295	0.087	-0.562	0.957	0.593	0.053	2.92
Zarei &	annotation	multiple gloss vs.	vocabulary	0.214	0.336	0.113	-0.446	0.525	0.873	0.635	2.81
Aljabri	annotation	multiple gloss vs.	vocabulary	0.857	0.221	0.049	0.424	0.000	1.290	3.878	3.10
Aljabri 1	annotation	multiple gloss vs.	vocabulary	0.477	0.214	0.046	0.057	0.026	0.897	2.225	3.11
Yoshii	annotation	multiple gloss vs.	vocabulary	0.302	0.144	0.021	0.020	0.036	0.583	2.101	3.24
Al.Musallam	annotation	multiple gloss vs.	vocabulary	0.752	0.317	0.100	0.131	0.018	1.372	2.373	2.86
Gorjian	annotation	multiple gloss vs.	vocabulary	0.485	0.255	0.065	-0.013	0.057	0.984	1.907	3.02
Khiyabani &	annotation	multiple gloss vs.	vocabulary	0.002	0.264	0.069	-0.515	0.995	0.518	0.006	3.00
Shabani	annotation	multiple gloss vs.	vocabulary	2.616	0.426	0.182	1.781	0.000	3.451	6.139	2.56
Abu Seileek	annotation	multiple gloss vs.	vocabulary	0.795	0.263	0.069	0.279	0.003	1.311	3.020	3.00
Al Ghafli	annotation	multiple gloss vs.	vocabulary	-0.078	0.164	0.027	-0.398	0.635	0.243	-0.475	3.21
Al-	annotation	multiple gloss vs.	vocabulary	1.063	0.273	0.074	0.528	0.000	1.597	3.897	2.98
Chen (2006)	annotation	multiple gloss vs.	vocabulary	0.630	0.230	0.053	0.180	0.006	1.081	2.742	3.08
Chen 1	annotation	multiple gloss vs.	vocabulary	0.733	0.232	0.054	0.279	0.002	1.187	3.162	3.07
Deridder	annotation	multiple gloss vs.	vocabulary	0.287	0.182	0.033	-0.071	0.116	0.644	1.573	3.18
Deridder 1	annotation	multiple gloss vs.	vocabulary	0.090	0.182	0.033	-0.266	0.621	0.445	0.494	3.18
Faramarzi &	annotation	multiple gloss vs.	vocabulary	1.709	0.242	0.059	1.234	0.000	2.183	7.057	3.05
Hsu (2011)	annotation	multiple gloss vs.	vocabulary	0.439	0.224	0.050	-0.000	0.050	0.879	1.958	3.09
Kim &	annotation	multiple gloss vs.	vocabulary	0.152	0.252	0.064	-0.343	0.548	0.646	0.601	3.03
Lin & Tseng	annotation	multiple gloss vs.	vocabulary	0.329	0.253	0.064	-0.166	0.193	0.824	1.302	3.03
Lin & Yu	annotation	multiple gloss vs.	vocabulary	0.108	0.247	0.061	-0.377	0.664	0.592	0.435	3.04
Shahrokni	annotation	multiple gloss vs.	vocabulary	2.489	0.341	0.117	1.820	0.000	3.158	7.290	2.80
Srichamnon	annotation	multiple gloss vs.	vocabulary	1.462	0.331	0.110	0.813	0.000	2.111	4.417	2.82
Yeh (2003)	annotation	multiple gloss vs.	vocabulary	-0.167	0.266	0.071	-0.689	0.531	0.355	-0.627	2.99
Yoshii (?)	annotation	multiple gloss vs.	vocabulary	0.812	0.351	0.123	0.125	0.021	1.500	2.316	2.77
Yousefzade	annotation	multiple gloss vs.	vocabulary	1.523	0.252	0.064	1.029	0.000	2.017	6.041	3.03
Zainol	annotation	multiple gloss vs.	vocabulary	5.530	0.400	0.160	4.746	0.000	6.315	13.811	2.63

Figure 1. Forest plot of the effects of single vs. multiple glosses on vocabulary acquisition.

As is evident, all effect sizes reported in the present meta-analysis were not positive, indicating that the combination of different gloss types had sometimes a negative impact on student learning. This analysis also reveals that 7 (20%) of the 34 studies had an effect size equal to 0.5 or greater, showing that the effect of multiple glosses on vocabulary acquisition was approximately moderate. 14 (42%) studies had an effect size 0.2 or less and 13 (38%) effects sizes were 0.8 or larger, showing a small and large effects respectively.

4.1.2 Outlier analysis

To examine the effect of any possible outlier that may cause extraordinary changes in the weighted mean effect size, an outlier analysis was run. Administering a "one-study removal" analysis, the weighted mean effect size did not change (ES= 0.839) within the 95% confidence interval. This implies that there is no outlier among primary studies which can affect the overall effect size estimate.

4.1.3. Publication bias

To detect any publication bias inherent in the current meta-analysis, the funnel plot of the primary studies was examined. As shown in Figure 2, while smaller studies tend to spread

more widely around the mean effect size at the bottom of the funnel (ES= 0.839) due to their larger standard error , larger studies mainly spread on top of the funnel.

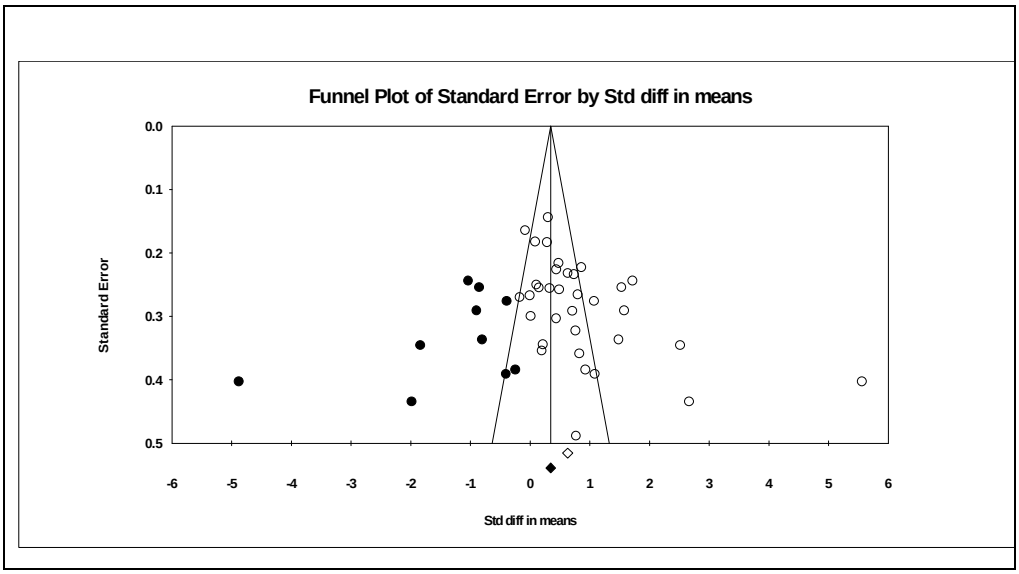


Figure 2. Funnel plots of the effectiveness of single (text-only) versus multiple (text+visual) glosses on vocabulary acquisition (● Represent imputed studies and ○ Show the observed ones included in the meta-analysis).

As it appears, the effect sizes are not symmetrically distributed around the vertical line shaping the funnel plot to an asymmetrical one. To detect the sources of bias in the present meta-analysis, the imputed studies were inserted in the funnel plot, revealing a clear tendency for the smaller studies to give more positive results than the larger studies. This can be accounted for as a sign of publication bias resulting from the publication of small studies if their results are significant and positive than if their results are negative or non- significant.

To calculate the number of unpublished studies with non-significant results which were considered to be necessary to nullify the overall effects reported in the current meta-analysis, two tests were run, Orwin's Fail-safe N tests and Classic Fail-safe N. The classic Fail-safe N test estimated that totally 2318 studies with null results are needed to nullify the overall effect size. The Orwin's Fail-safe N test also estimated that 395 missing null studies are needed to make the mean effect size closer to the critical value of 0.05.

4.1.4. Analysis of residual heterogeneity

To explore the degree of homogeneity among effect size estimates included in the present meta-analysis, the test of homogeneity of variance – the *Q*-test – (Lipsey & Wilson, 2001) was run. The resulting Q value (Q=327.14, df= 33, $p<0.05$) as shown in Table 1 rejected the

hypothesis of homogeneity and showed that variation between the effect sizes may not be due to the sampling error alone and other variables may contribute to such heterogeneity. According to the I^2 value (the percent of variance between effect sizes not caused by chance alone), 90% of the observed variance between studies is due to real differences in the effect size or between-study level variability. Only about 10% of the observed variance would have been expected based on random error. Moderator analyses that follow might be able to help clarify this unexplained variability.

Contextual factors affecting the variation between the effects of different gloss types on vocabulary acquisition

Regarding the significant heterogeneity found between effect sizes and the possible influence of some moderator variables, the following methodological and substantive variables were considered to be related to the overall effects of different types of glossing for L2 vocabulary acquisition: research design, sample size, L2 proficiency levels, the duration of the program, second vs. foreign language environments (SL and FL, respectively), and the publication year. The overall mean effect sizes (Hedges' *g*), 95% confidence interval ranges and the Q-value for all studies as well as subcategories which involved the different contextual factors were calculated (Table 2) and discussed in detail in next sections. The findings help to see which contextual factors influenced the effectiveness of multiple and single gloss types on vocabulary acquisition in these studies.

Table 2. Summary of moderator analyses

Moderator Variable level	Numb. of Effect Size (N_{es})	Effect Size (g)	Lower Confidence	Upper Confidence	Q-value	*p*-value
1. Research design					**0.920**	**0.337**
Randomized	9	1.167	0.275	2.059		
Matched	25	0.715	0.475	0.955		
2. Sample size					**0.350**	**0.554**
More than 60 (large)	21	0.881	0.513	1.249		
Less than 60 (small)	13	0.728	0.378	1.077		
3. Proficiency level					**6.533**	**0.000 ***
Advanced	2	0.816	0.343	1.290		
Intermediate	25	0.850	0.521	1.178		
Elementary	7	0.755	0.130	1.380		
4. Duration of the program					**3.897**	**0.048 ***
High	10	1.483	0.602	2.363		
Low	24	0.571	0.363	0.779		
5. Context of learning					**0.405**	**0.525**
EFL	27	0.864	0.535	1.193		
ESL	7	0.677	0.204	1.150		
6. Publication year					**3.033**	**0.082**
2000s	15	0.564	0.312	0.817		
2010s	19	1.046	0.566	1.527		

4.1.5. Research methodological characteristics

4.1.5.1. Research design

According to some researchers (e.g., Abrami & Bernard, 2006), different research designs are considered to act as potential sources of variation between effect size estimates. To further investigate this, all studies analyzed in the current meta-analysis were classified into two subcategories based on their research design; randomized experiments (N= 9) in which subjects were randomly assigned to multiple (text+visual) and single (text-only) hypertext gloss groups, and matched studies (N= 25) which were those in which subjects in both experimental (text+visual) and control (text-only) groups were matched based on key variables in the pretest.

As presented in Table 2, the results reveal that multiple hypertext glosses had a moderate impact in matched experiments (ES = .71) and lager impacts in randomized studies (ES = 1.16). Considering the two 95% confidence intervals of the two mean effect sizes, both were statistically significant since they crossed zero. However, based on the Q value (Q= 0.92, p>0.05) reported in Table 2, the difference between the two effect sizes of subcategories of research design was not statistically significant.

4.1.5.2. Sample size

To address the variation of learners working either in small groups or large ones, the weighted mean effect sizes for the subgroups were calculated. In the current meta-analysis, studies including sample size greater than 60 were grouped as large (n=21) and those who had participants less than 60 were considered to be small ones (N=13). The calculation results show that multiple glosses had large, positive effects on large studies (ES= 0.88), but a medium effect on the small subgroup (ES = 0.72). However, the difference between the large and small studies effects sizes proved to be statistically non-significant (Q= 0.35, p> 0.05).

4.1.6. Substantive features

4.1.6.1. Proficiency level

In order to check the homogeneity of effect sizes across three proficiency levels of learners, the weighted mean effect sizes for elementary (N= 7), intermediate (N=25) and advanced (N= 2) proficiency levels were calculated. The results show that the treatment had moderate effects on the subgroup of elementary learners (ES= 0.75), but large effects on the subgroup of

intermediate (ES = .85) and advanced (ES = .81). Based on the 95% confidence intervals, the overall mean effect sizes for the three proficiency levels of learners were shown to be statistically significant. Moreover, as presented in Table 2, the Q-value (Q=6.53, p<0.05) showed that heterogeneity between effect sizes across three proficiency groups is statistically significant, shedding light on the hypothesis that multiple hypertext glosses seem to highly benefit advanced and intermediate learners in comparison to their counterparts in elementary group for L2 vocabulary acquisition.

4.1.6.2. Duration of the program

To examine whether the intensity of treatment influenced the effectiveness of multiple hypertext glosses for enhancing L2 vocabulary acquisition, the effect sizes were calculated for the two subcategories of treatment duration: high and low intensity of intervention. It was revealed that the treatment had a moderate, positive effect when the treatment duration was low (ES = 0.57) and a large, positive effect when the treatment period was high (ES = 1.48); both of these findings were statistically significant.

Based on Table 2, the test of heterogeneity did show a statistically significant difference between the effect sizes of two treatment duration subcategories (Q= 3.89, p<0.05), suggesting that more technology use can probably lead to better outcomes.

4.1.6.3. Context of learning

To investigate the effect of multiple hypertext glosses in different language learning contexts, the overall mean effect sizes for second and foreign language learning conditions were calculated. The results indicate that the treatment had a medium impact when the learners learnt a language in an ESL context (*ES* = .67) and large impacts when the participants were acquiring target vocabulary items in an EFL context. Both 95% confidence intervals of the two mean effect size were statistically significant. The analysis of the Q-value (Q= 0.40, p=0.52) also showed that even though learners receiving the treatment in an EFL context performed slightly better than their counterparts in an ESL context, the difference between effect sizes in two conditions was not statistically significant.

4.1.6.4. Publication year

Some researchers (e.g. Cheung & Slavin, 2011; Fletcher-Finn & Gravatt, 1995) have hypothesized that parallel to the development of educational technology, its effectiveness improves remarkably too. To test this hypothesis, studies which were included in this meta-

analysis were subcategorized into two time span; 2000s and 2010s. The mean effect size estimates for these two time periods were found to be 0.56 and 1.54, respectively. Although the test of heterogeneity show that the difference between learning outcomes resulting from multiple hypertext gloss application was not statistically significant, the present review provides support for more positive results attained in recent years.

5. Discussion

Pre-defined inclusion criteria yielded 34 qualified studies to be analyzed. The first goal of the current meta-analysis was to examine the effects of different hypertext gloss types (multiple vs. single gloss) on L2 vocabulary acquisition. Regarding this, it was found that multiple glosses had a large positive effect (ES= 0.83) on vocabulary acquisition. Like previous meta-analyses conducted in the same trend, the results of the current study reveal that a combination of different types of glosses (text, picture, video, etc.) generally lead to better learning outcomes in comparison to traditional single gloss type. The overall effect size obtained in this study is much larger than that reported in a recent meta-analysis conducted by Yun (2011), investigating the effect of multiple hypertext glosses on different vocabulary knowledge (ES= 0.46).

The heterogeneity Q-value obtained in the present meta-analysis proved to be statistically significant, revealing that the variation between effect sizes may be due to factors other than sampling error alone. To explain the unexplained variation, it was required to conduct sub-analyses to further investigate the moderator variables which were hypothesized to influence the effectiveness of multiple hypertext glosses. To discuss the points more deeply, methodological features like research design and sample size as well as some substantive features such as learners' proficiency level, learning context, intensity of intervention, publication year have been scrutinized to uncover some potential moderators that contribute to the variation between effect sizes in the present meta-analysis.

With regard to research design, it appeared to have no statistically significant impact on the variation between studies although studies with randomized design had a better mean effect size (ES= 1.16) than matched studies (ES= 0.71). This finding is in contrast with the study conducted by Cheung and Slavin (2011), which emphasized the fact that different research designs yield different outcomes to the extent that the mean effect size for the quasi-experimental studies was revealed to be approximately twice the size of the effect for randomized studies.

When it comes to sample size of studies, some researchers like Slavin and Smith (2009) emphasize that small sample size studies are more likely to provide larger effect sizes compared to those of larger studies. Contrary to this suggestion, the present study yielded a larger mean effect size for large studies compared to that of the small ones. That is to say, studies with small sample size had less statistical power than large sample size studies: a small sample size tends to yield a small overall weight while a large sample size produces a larger weight. This finding was not in line with Liao (1999) and Cheung and Slavin's (2011) results, showing that studies with larger samples had less statistical power than those with small samples. They asserted that hypertext gloss effects on learners' achievement should be reported cautiously in cases where sample size of the studies is small or medium. Future research may shed light on confirming or rejecting this contrasting finding.

Regarding learners' proficiency levels, the findings revealed that multiple hypertext glosses had larger impact when participants were in an intermediate and advanced proficiency level. One justification may be that beginners probably lack enough L2 proficiency to search through hypertext nodes. As Loewen and Erlam (2006) claim, learners may encounter some target items which are beyond their development stages while they consult with hypertext glosses. Besides, based on the cognitive load theory (Sweller & Chandler, 1991), less proficient language learners may be unable to use all benefits of multimedia glosses in vocabulary acquisition and reading comprehension because of the high cognitive load of such multimedia tools (Plass et al., 2003; Sweller, 1994). Hence, when learners who are at the intermediate level of language proficiency are compared with beginners, they seem to benefit more from certain types of glosses. The finding of the present study is in contrast to Yun's (2011) study, reporting that out of all learners it was beginners with access to multiple types of glosses that benefited most from multiple glosses in reading. The current finding also supports the results of recent study conducted by Ben Salem (2006), showing that due to the high cognitive load inherent in multiple hypertext glosses, advanced learners took most benefit from such glosses as compared with learners at low or intermediate proficiency level.

As for the intensity of the treatment, the results showed that long duration of the programs had higher mean effect size estimates than short one. It seems to suggest that more exposure to glosses can lead to better learning outcomes.

With regard to the learning conditions, it was revealed that the treatment had a moderate impact for the subgroup of ESL context, and also large impact for the subgroup of EFL context. It is worth mentioning, however, that the sample size of some subgroups was quite small, affecting the overall effect size of the treatment. Further analyses including larger

number of samples especially in ESL/ EFL context is needed to reach a better understanding of how language learning context affect the effectiveness of different types of hypertext glosses.

As for the year of publication, the results showed that studies conducted recently in the 2010s yielded higher mean effect size estimates as compared to those conducted in the 2000s.

6. Conclusion

The current meta-analysis examined the results of 34 primary studies that compared learners' target vocabulary acquisition under two conditions; multiple glosses (text+visual) and single gloss (text-only). The overall mean effect size was found to be + 0.83 ($p<0.05$), indicating that multiple glosses (text+visual) had a large, positive effect on learners' vocabulary acquisition than single mode of glossing (text-only). Moderator analyses further suggested that intensity of the program, L2 proficiency level and sample size were potential moderators influencing the heterogeneity between effect sizes, whereas the learning context and research design did not have such an impact.

The present meta-analysis, like most others, has several limitations. First, some sources of bias such as publication bias might be found in the present meta-analysis resulting from the great number of unpublished papers and the lack of access to such unpublished sources. In addition, a great number of significant studies may be extracted in the present meta-analysis due to the lack of the availability of critical statistical data for the analysis and inaccessibility of the authors. Besides, few studies (only 9 out of 34 in this study) included a delayed posttest; therefore, the long-term effect of different gloss types on vocabulary retention is impossible to be determined. Further research is expected to be conducted having a design including both immediate and delayed post-test evaluations. Third, besides learners' proficiency level, learners' learning styles (verbalizers or visualizers) seem to be a critical moderator variable affecting the final learning outcome. Thus, the learners who prefer verbal type of glosses (text-only) tend to benefit most from a single gloss type in the form of text only gloss when given a chance to select either text-only or text + visual glosses, while the learners who prefer visual type gloss tend to benefit most from a combination of different gloss modes such as text + visual gloss. Investigating such moderator variable was not possible in this study since none of the primary studies provided the detailed information about the learners' learning preferences. Future research is needed to fill this gap.

The present study has some pedagogical implications. First, hypertext glosses can yield better learning outcomes if learners are provided with opportunities to acquire essential

skills to study individually and use their time effectively in a computer environment in which there is less teacher control as compared to traditional classrooms. Training sessions can make students gradually familiar with the CALL environment, showing how to make more effective use of glosses. Furthermore, teachers should also become aware of the potential of hypertext glosses and their different types as learning tools as well as the ways through which different gloss types can promote efficient language learning.

References

Note: Studies that were included in the present meta-analysis are marked with an asterisk (*).

Abrami, P. C., & Bernard, R. M. (2006). Research on distance education: In defense of field experiments. *Distance education*, 27(1), 5–26.

Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary acquisition and reading comprehension of advanced learners of English as a foreign language. *Instructional Science*, *35*, 499-517.

*Ahangari, S., & Abdollahpour, Z. (2010). The effect of multimedia annotations on Iranian EFL learners' L2 vocabulary acquisition. *The Journal of Applied Linguistics*, 3(1), 1-18.

*AlGhafli, M.H. (2011). The effect of mediated glosses on vocabulary retention and reading comprehension with English language learners in Saudi Arabia. Unpublished dissertation.

*Aljabri, S.S. (2009). The effect of pictorial annotations on Saudi EFL students' incidental vocabulary acquisition. *Journal of Arabic and Human Sciences*, 2(2), 43-52.

AlSadhan, R. O. (2011). *Effect of textual enhancement and explicit rule presentation on the noticing and acquisition of L2 grammatical structures: a meta-analysis.* Unpublished master's thesis, Colorado State University, Colorado.

*Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, *5*(1), 202-232.

*Al-Musallam, E., Al-Twairesh, N., & Al-Shubaily, S. (2005). Acquisition of vocabulary items through multimedia vs. still pictures: a comparative study. *Language Learning*, 6(2), 56-68.

Ariew, R. (2006). A template to generate hypertext and hypermedia reading materials: its design and associated research findings. *The Reading Matrix*, *6*(3), 195-209.

Ben Salem, Y., & Aust, R. (2007). The influence of feature-rich computerized glosses on reading comprehension and vocabulary acquisition. *Proceedings of the sixth IASTED conference on web-based education*, France, 182-190.

*Chen, Z. (2006). *The effects of multimedia annotations on L2 vocabulary immediate recall and reading comprehension: A comparative study of text picture and audio-picture annotations under incidental and intentional learning conditions*. Unpublished doctoral dissertation.

Cheung, A.C. & Slavin, R. E. (2011). *The effectiveness of education technology for enhancing reading achievement: a meta-analysis.* Best Evidence Encyclopedia.

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

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

* De Ridder, I. (2002). Visible or invisible links: does highlighting of hyperlinks affect incidental vocabulary acquisition, text comprehension, and the reading process? *Language Learning & Technology,* 6(1), 123-146.

*Faramarzi, S., Elekaei, A., & Koosha, M. (2014). On the impact of multimedia glosses on reading comprehension, vocabulary gain and vocabulary retention. *International Journal of Language Learning and Applied Linguistics World*, 6(4), 623-634.

Fletcher-Finn, C., & Gravatt, B. (1995). The efficacy of computer-assisted instruction (CAI): A meta-analysis. *Journal of Educational Computing Research*, 12(3), 219–241.

*Gan, X. (2014). Study on the effects of gloss type on Chinese EFL learners' incidental vocabulary acquisition. *Theory and Practice in Language Studies*, 4(6), 1251-1256.

Glass, G. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351-379.

*Gorjian, B. (2011). Using hypermedia annotations to teach vocabulary on the web. *ALT-C2011Conference Proceedings.*

Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando: Academic Press.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J., & Altman, D.G. (2003). Measuring inconsistency in meta-analysis. *Br. Med. J.*, 327, 557–560.

*Hsu, M-H. (2011). The effect of first language gloss on reading comprehension, lexical acquisition and retention: single gloss and multiple-choice gloss. *WHAMPOA – An Interdisciplinary Journal***,** 61, 33-52.

Khan, B. H. (1997). *Web-based Instruction*. New Jersey: Educational Technology publications.

*Khiyabani, H.R., Ghonsooly, B., & Ghabanchi, Z. (2014). Using multimedia in teaching vocabulary in high school classes. *Journal of Advances in English Language Teaching*, 2(1),1-13.

*Kim, D., & Gilman, D.A. (2008). Effects of text, audio, and graphic aids in multimedia instruction for vocabulary acquisition. *Educational Technology & Society*, *11* (3), 114-126.

Kost, C. R., Foss, P., & Lenzini, J. J. (1999). Textual and pictorial glosses: Effectiveness of incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals, 32* (1), 89-113.

Kulik, J. & Kulik, C. L. (1987). Review of recent research literature on computer-based instruction. *Contemporary Educational Psychology*, 12, 222-230.

Kulik, C. L., Kulik, J., & Schwalb, B. (1986). Effectiveness of computer-based adult education: a meta-analysis. *Journal of Educational Computing Research*, *2*(2), 235-252.

Liao, Y. C. (1999). Effects of hypermedia on students' achievement: A meta-analysis. *Journal of Educational Multimedia and Hypermedia, 8*(3), 255-277.

*Lin, C., & Tseng, Y. (2012). Videos and animations for vocabulary acquisition: a study of difficult words. *TOJET: The Turkish Online Journal of Educational Technology,* 11(4), 346-355.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical Meta-Analysis.* Thousand Oaks, CA: Sage.

Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning 19*(1), 1-14.

Lomicka, L. (1998). "To gloss or not to gloss": An investigation of reading comprehension online. *Language Learning & Technology*, *1*(2), 41-50.

*Moazzeni, Zh., Bagheri, M. S., Sadighi, F., & Zamanian, M. (2014). The effect of different gloss types on incidental vocabulary retention of Iranian EFL students. *International Journal of Language Learning and Applied Linguistics World*, 5(2), 396-415.

Nagata, N. (1999). The effectiveness of computer-assisted interactive glosses. *Foreign Language Annals, 32* (4), 469-479.

Nation, I. S. P. (2002).*Learning Vocabulary in Another Language. The Cambridge Applied Linguistics Series.* Cambridge University Press.

Nikolova, O. R. (2004). Effects of visible and invisible hyperlinks on vocabulary acquisition and reading comprehension for high-and average-foreign language achievers. *ALSIC*, *7*(1), 29-53.

Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 90*(1), 25-36.

*Poole, R. (2012). Concordance- based glosses for vocabulary acquisition. *CALICO Journal*, 29(4), 679-693.

Robin, R. (2007). Commentary: Learner-based listening and technological authenticity. *Language Learning & Technology*, *11*(1), 109-115.

Schmidt, R. (1995). Consciousness and foreign language learning. In R. Schmidt, (Ed.), *Attention and Awareness in Foreign Language Learning* (pp. 1-63). University of Hawai'i at Manoa: Second Language Teaching and Curriculum Center.

*Shbani, K. (2014). The effects of computerized instruction of vocabulary through hypertexts on L2 learners' cognitive functioning. *Procedia- Social and Behavioral Sciences*, 149(2014), 868-873.

*Shahrokni, S.A. (2009). Second language incidental vocabulary acquisition: the effect of on-line textual, pictorial, and textual pictorial glosses. *Electronic Journal for English as a Second Language*, 13(3), 25-35.

Slavin, R.E., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, 31(4), 500-506.

*Srichamnong, N. (2010). Incidental EFL vocabulary acquisition: the effects of interactive multiple-choice glosses. *Proceeding of international conference ICT for language learning.*

Stern, J. A. C., Becker, B. J. & Egger, M. (2005). *The Funnel Plot. Publication Bias in Meta-Analysis.* Chichester: John Willy & Jones.

Sweller, J. & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction*, 8, 23-34.

*Tabatabaei, O., & Shams, N. (2011). The effect of multimedia glosses on online computerized L2 text comprehension and vocabulary acquisition of Iranian EFL learners. *Journal of Language Learning and Research*, 2(3), 714-725.

*Tababtabaei, O., & Mirzaei, M. (2014). Comprehension and idiom learning of Iranian EFL learners. *Journal of Educational and Social Research*, 4(1), 45-56.

Yanguas, I. (2005). Type of multimedia gloss and L2 proficiency: A computer-based study. Paper presented at Second Language Research Forum, (SLRF), New York, NY.

*Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning & Technology*, 13(2), 48-67.

Yeh, Y., & Wang, Ch. (2003). Effects of multimedia vocabulary annotations and learning styles on vocabulary acquisition. *CALICO Journal*, 21(1), 131-144.

Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: the effect of text and picture annotation types. *CALICO Journal*, *20*(1), 33-58.

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

*Yoshii, M. (2011). Effects of gloss types on vocabulary acquisition through reading: comparison of single and multiple gloss types. *CALICO Journal*, 4 (2), 34-51.

*Yousefzadeh, M. (2011). Computer-based glosses vs. traditional paper-based glosses and L2 learners vocabulary acquisition. *International Journal on New Trends in Education and Their Implications*, 2(3), 99-102.

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

*Zainol Abidin, M., Pour-Mohammadi, M., Sharbaf Shoar, N., Cheong, S.T.H., & Jafre, A.M. (2011). A comparative study of using multimedia annotation and printed textual glossary in learning vocabulary. *International Journal of Learning & Development*, 1(1), 82-90.

*Zarei, A.A., & Mahmoodzadeh, P. (2014). The effect of multimedia glossing on L2 reading comprehension and vocabulary production. *Journal of English Language Literature*, 1(1), 1-7.