Accepted for publication in TESOL Quarterly (Wiley Blackwell)

Efficacy of Multimodal Glossing on L2 Vocabulary Learning: A Meta-Analysis

Nasrin Ramezanali, Takumi Uchihara, & Farahnaz Faez

(University of Western Ontario)

ABSTRACT

This meta-analysis examined the effectiveness of an additional gloss mode in single versus dual and dual versus triple glossing on second language (L2) learners' word learning. In total, 22 studies, providing 26 independent effect sizes, were coded, and 11 moderator variables including quality of data sample, learner variables, gloss features, text features, and methodological features were examined. The results showed that the overall effect of an additional gloss mode was medium (g = 0.46) for immediate posttests and small (g = 0.28) for delayed posttests. However, analyses of moderator variables indicated that the effect of additional gloss modes is influenced by a range of variables related to learner (e.g., proficiency), gloss (e.g., language), text (e.g., narrative versus expository), and research design (e.g., test format). Importantly, adding an additional mode to single textual gloss enhances vocabulary learning, whereas adding a mode to dual glossing does not result in significantly better vocabulary learning. The findings suggest that using more than two gloss modes is not necessary as it does not always lead to better learning of new words.

KEYWORDS

glossing, single glossing, dual glossing, multimodal glossing, multimedia glossing, vocabulary learning, meta-analysis

INTRODUCTION

Vocabulary knowledge is fundamentally necessary for mastering another language. Foreign/second language (FL/L2) learners require multiple exposures to L2 vocabulary in various contexts to assist them to manage and recall unfamiliar words promptly over the long term. Vocabulary learning techniques can help expand learners' word reservoirs. These techniques include learners' engagements with words, time spent on lexical items, intentional and incidental vocabulary learning activities, and glossing.

Several studies have examined the effectiveness of glossing for L2 vocabulary learning. Findings have revealed that glossing enhances L2 vocabulary learning and retention (Al-Seghayer, 2001; Ramezanali & Faez, 2019), avoids incorrect guessing (Ko, 2005), leads learners to learn words autonomously (Stewart & Cross, 1991), and can be a substitute for a dictionary (Yanguas, 2009). A recent meta-analysis conducted by Yanagisawa, Webb, and Uchihara (2020) suggests varying degrees of effectiveness of different formats of glosses (e.g., multiple choice, glossary, texts, audios) for L2 vocabulary learning. Furthermore, different numbers of glossing modes, including single (Boers, Warren, He, & Deconinck, 2017), dual (Al-Seghayer, 2001), and triple (Yeh & Wang, 2003), have been reported to help improve the word learning process.

Within the broad category of multimodal glossing, multimedia glossing has emerged (Al-Seghayer, 2001; Türk & Erçetin, 2014) to present explanations of target words in hypertext contexts via multiple resources such as texts, graphics, still pictures, sounds, and dynamic videos/animations in a multimedia-based learning environment. In general, adding an additional mode to glossing (e.g., definitions + pictures, definitions + pictures + audios, definitions + pictures + videos + audios) in hypertext contexts is an effective way to enhance reading comprehension (Al-Seghayer, 2001), increase vocabulary learning (Ramezanali & Faez, 2019), provide FL/L2 learners with adequate use of authentic texts (Al-Seghayer, 2001), attend to learners' preferred mode of vocabulary learning (Ramezanali & Faez, 2019), and save learners' time and effort in reading L2 texts (Khezrlou & Ellis, 2017). Nevertheless, the findings of relevant studies are inconsistent as to whether adding an additional gloss mode to single and dual modes can contribute positively to vocabulary acquisition. As an example, adding an additional gloss mode to single glossing (e.g., adding a video or picture to a textual definition) was not reported to be beneficial in some studies (Acha, 2009; Ariew & Erçetin, 2004; Boers et al., 2017), with the additional mode distracting learners (Ariew & Erçetin, 2004). Whereas, other studies (Al-Seghayer, 2001; Ramezanali & Faez, 2019) reported that dual gloss combinations of text and video or text and picture were significantly more effective for L2 learners' vocabulary learning compared with single glossing.

Given such inconsistencies, the present meta-analysis of 22 studies included either single versus dual or dual versus triple glossing modes to seek clarity about the effectiveness of different gloss combinations on L2 vocabulary learning. For the purpose of this study, single glossing mode is defined as a one gloss mode (i.e., textual definition-only). Dual glossing modes represent the combination of two glossing modes (e.g., definitions with audios, definitions and pictures, or

definitions and videos), and multimodal glossing is viewed as any triple gloss combination (e.g., definitions + audios + pictures, definition + pictures + videos).

LITERATURE REVIEW

Studies on Single, Dual, and Triple Glossing Modes

Beginning with a definition of multimedia, this section summarizes the results of major studies on different combinations of glossing (i.e., single, dual, triple) in a multimedia setting. Because part of speech and inclusion of pretests in studies included in this meta-analysis are two moderator variables, we then discuss how the choice of part of speech of the glossed words contributed to word learning, and how learners' vocabulary gains were measured for studies that did not administer pretests; finally, two relevant meta-analytic studies and their limitations and gaps are discussed.

Multimedia. Mayer (2014) defined multimedia as the presentation of both words and pictures to learners. In his definition, words signify any material presented in verbal form such as spoken texts or printed texts, and pictures represent static or dynamic pictorial materials such as illustrations, graphs, diagrams, maps, animations, or videos. Thus, multimedia learning refers to "building mental representations from words and pictures" (Mayer, 2014, p. 2).

The underlying theory that frames multimedia learning with vocabulary is the cognitive theory of multimedia learning, suggesting that word learning is best facilitated when the new information is presented through both verbal and visual representational modes rather than just a single mode. In other words, L2 learners' interactivity with multimedia input is enhanced through connecting both visual and verbal systems to written and pictorial cues in the brain. Furthermore, utilizing multimedia annotations/glossing to instruct unknown words can help learners to experience word learning through the cognitive processes of selecting relevant words and images, organizing them into verbal and visual representations, and finally integrating the words into corresponding verbal and visual stimulus (Mayer, 2014). The three assumptions of this theory are dual channels (individuals have two separate information processing channels: verbal and visual), limited capacity (individuals can process limited amount of information in each channel at one time), and active processing (there is a connection between verbal and visual modes that integrate them and account for individual's long-term memory). The theory also indicates that having two separate but interrelated verbal and visual systems allows learners to benefit more if they receive the target words through the verbal as well as visual tools.

Single versus dual glossing modes. Although numerous studies on hypertext glosses have been conducted, the results are inconclusive in regard to the effectiveness of gloss types and gains on L2 vocabulary acquisition. Findings of several studies have shown that dual glossing mode is more effective than single mode for word acquisition. Utilizing meaning recognition and meaning recall measurements, Al-Seghayer (2001) conducted a within-participants study with L2 learners. The findings revealed that glossed words with definitions and pictures and definitions and videos were learned better than those with definitions only. Also, the participants recalled words better when definitions and video clips were provided than with definitions and still pictures. The glossed words included mixed parts of speech (i.e., nouns, adjectives, and

verbs). Using dual glossing modes of text definition and pictures as well as text definition and videos with mixed parts of speech for glossed words, Akbulut's (2007) between-participants study showed that the dual glossing of definition and videos was significantly more effective than single glossing mode of text-only in three vocabulary measurements of form recognition, meaning recognition, and meaning recall.

In a between-participants design, Ramezanali and Faez (2019) reported that both meaning recognition and vocabulary recall measurements in immediate and delayed tests showed that dual glossing modes of definition and video/animation and definition and audio glossing were more effective than single glossing mode (definition-only) for many test sessions; even though single glossing mode was also effective for a few test sessions. The selected glossed words were only nouns in their study. In contrast, working with different groups of participants and language abilities and using nouns only, Acha (2009) conducted a between-participants experimental study of L2 children to examine if simultaneous verbal and visual presentation modes (e.g., definition, picture, or both) would affect learners' vocabulary acquisition in the short and long term. The results of meaning recall measurement showed that verbal annotation alone (i.e., definition) was more effective than the visual mode (i.e., picture) for word learning and recollection after 2 weeks. Likewise, administering meaning recognition and recall tests, Boers et al. (2017) found that the dual glossing modes of textual and pictorial did not help learners to learn and retain nouns any better than providing the learners with textual explanations only (single glossing mode). Therefore, according to the findings of these two studies, single glossing mode was more effective than dual modes for word learning and retention. These findings contradict studies that reported the advantage of dual modes compared with textual definition only.

Dual versus triple glossing modes. Studies on dual- and triple-gloss modes have also shown inconsistent findings. With a dual versus triple gloss combination, Yeh and Wang (2003) studied whether definition annotations only; definition and picture annotations; and definition, picture, and sound annotations were effective for learners' vocabulary learning in a multimedia setting. The findings indicated that the mode with definition and picture was the most effective type of vocabulary annotation. In contrast, Sadeghi, Khezrlou, and Modirkhameneh (2016) distinguished between three vocabulary learning conditions—explicit (i.e., learning target words before reading a text), incidental (i.e., no prior vocabulary instruction), and intentional (i.e., consulting glosses during reading) conditions. The authors then investigated the effectiveness of each condition on learners' vocabulary acquisition and reading comprehension by means of different multimodal hypertext glosses (e.g., definition and pictures; definition and audio; and definition, picture, and audio) in a multimedia environment using three word categories (nouns, adjectives, and verbs). The findings of the between-participants design showed that the latter gloss type (definition, picture, and audio) was more effective than the other gloss modes and resulted in better vocabulary gains and reading comprehension. Also, the learning conditions affected participants' performance across gloss types in terms of both immediate and delayed posttests. The dual mode of definition and picture was effective with incidental vocabulary instruction in the immediate posttest. However, the participants in explicit and intentional groups achieved higher scores in the triple glossing (definition, picture, and audio) modes.

Glossing and part of speech. Even though both nouns and verbs are among "the most common parts of speech found in natural text" (Webb, 2008, p. 234), most studies have chosen nouns for glossing. Compared with other parts of speech, it is easier to provide simple and clear word definitions, find suitable pictures, and comprehensible video animations for concrete nouns (Acha, 2009; Boers et al., 2017; Ramezanali & Faez, 2019). As Mohsen and Balakumar (2011) report, concrete words can be imaged by different visual modes because "they are tangible to the senses" (p. 153). Choosing nouns to gloss can also control for the part of speech (Jung, 2016; Laufer & Rozovski-Roitblat, 2011). There are still studies that have used verbs only (e.g., Lin, 2009) or the mixture of other parts of speech (nouns, verbs, adjectives; e.g., Akbulut, 2007). Hence, it is important to understand whether glossing has a differential impact for learning different parts of speech.

Learners' learning gains and word preknowledge. Not all the studies included in this meta-analysis have administered a pretest to measure participants' preknowledge of the words to be glossed. Instead, they have used unique ways to control for prior word knowledge of glossed words. As an example, Akbulut (2007) and Tabatabei and Shams (2011) asked the participants to read the texts and underline the words they did not know; expert review was then conducted for the suitability of words to learners' level. Salem (2006) asked language instructors to rate the words in terms of difficulty from a scale of 0 to 10. Words that received an average rating or above were selected for glossing. Finally, Boers et al. (2017) used pseudowords with meanings glossed in the margin. The variation of approaches to controlling prior knowledge of target words might be a potential confounding variable worth further exploration, as it might lead, in part, to mixed findings of previous vocabulary studies comparing the effectiveness of multimodal glossing. Thus, the inclusion of pretest is one of the moderator variables in the present meta-analysis.

Meta-analytic studies on glossing modes. A meta-analytic review on glossing modes and L2 word learning is not new (e.g., Abraham, 2008; Huang, 2010). There are two studies that compare single versus multimodal glossing that align with the purposes of this study (Vahedi, Ghonsooly, & Pishghadam, 2016; Yun, 2011).

Vahedi et al. (2016) conducted a meta-analysis of the effect of hypertext single versus dual gloss modes on L2 vocabulary acquisition, considering within- and between-study variables. Analyzing 34 studies, the researchers found that the dual glossing mode (textual and visual) had a larger effect size than single glossing mode (definition-only) on learners' vocabulary gains (the standardized mean difference in gain scores between the two groups or Hedges's $g = 0.84$). Moderator variables such as intensity of the program and L2 learners' proficiency levels contributed to the between-studies heterogeneity in effect sizes more than the variables of learning context, sample size, and research design.

In a similar vein, Yun (2011) performed a meta-analysis of 10 studies with 35 effect sizes, which included both within- and between-participants designs to compare single (text-only) versus dual (text and visual) glossing modes, examining their effectiveness on vocabulary learning through reading. The findings showed that dual glossing modes were moderately more effective than the

single counterpart for immediate posttests (Hedges's g = 0.41) and delayed posttests (g = 0.35). The variables of learners' proficiency levels and test formats were also shown to significantly moderate the effect of dual glossing modes on learning gains. As for the effect of proficiency as a moderator variable, it seems that proficiency levels were coded as described in the original studies. Learners at the beginning level tended to benefit more from dual glossing compared with single glossing (g = 0.698) than other proficiency levels (i.e., intermediate, beginning intermediate, and advanced). Looking at test formats as a moderator, the benefit of dual glossing appeared to be larger than single glossing when learning gains were measured through multiple-choice test formats (recognition) than recall or read-aloud protocol (production).

Although revealing, these two meta-analytic reviews have some limitations. First, they investigated the effectiveness of single versus dual glossing modes on L2 vocabulary learning without considering the impact of an additional glossing mode to dual glossing. It is important to specify the number of glossing modes as single versus dual or dual versus triple because of the role that learners' cognitive capacity plays in processing relevant information. In other words, the number of gloss modes may impact the learners' cognitive abilities to process new words. The rationale relies on two of the principles of multimedia learning: multimedia principle, suggesting that learners can learn more effectively if they are presented with words and pictures (i.e., dual modes) rather than words alone (i.e., single mode); and redundancy principle, indicating that people learn better from graphics and narration than from graphics, narration, and on-screen text (Mayer, 2014). Although the evidence regarding the redundancy principle is inconclusive considering its beneficial effect in L2 learning contexts, , learners may process a limited amount of information at a time due to the cognitive overload. Hence, it is important to investigate if more than two modes may be subject to overload of information for L2 learners.

Second, the two previous meta-analyses examined both within- and between-participants designs. This blended data could have provided biased effect size unless correlations between paired samples were implemented for bias correction (Plonsky & Oswald, 2015). To avoid this and due to the limited number of within-participants studies that met our selection criteria, the present meta-analysis focused exclusively on primary studies with between-participants design.

Third, the extent to which L2 word gains are retained over time is inconclusive. Although Yun's (2011) meta-analysis showed an advantage of dual over single glossing modes for delayed posttests, the result was based on a mixture of within- and between-participants designs, which might have failed to accurately represent the effect of additional gloss modes on long-term vocabulary retention. Thus, the present meta-analysis aimed to fill these gaps by examining the effectiveness of single versus dual and dual versus triple glossing modes on L2 learners' vocabulary learning for immediate and delayed posttests.

RESEARCH QUESTIONS

This study was guided by the following research questions:

> RQ1. Is adding a glossing mode to the combination of glossing annotations beneficial for L2 learners' vocabulary gains for both immediate and delayed posttests?

RQ2. How does the addition of a glossing mode compare between single versus dual and dual versus triple glossing modes for L2 learners' vocabulary gains for both immediate and delayed posttests?

RQ3. To what extent do learner variables, gloss, text, and methodological features moderate the overall benefit of an additional glossing mode for vocabulary gains for both immediate and delayed posttests?

METHOD

Literature Search

As guided by In'nami and Koizumi (2010), the following databases were searched in full-text PDFs, paper copies, and online to identify resources for inclusion in the analysis:

- The Educational Resources Information (ERIC)
- Linguistics and Language Behavior Abstracts (LLBA)
- ProQuest Dissertations and Theses
- Google
- Google Scholar
- Journals of *Language Learning*, *Language Learning and Technology*, *TESOL Quarterly*, *Applied Linguistics*, *Language Teaching Research*, and *Studies in Second Language Acquisition*
- Western University Library Catalogue.

In addition, Western University databases were checked for each of the named journals. Abstracts published from 2001 to 2018 were also explored, with combinations of the following keywords:

- Glossing
- Multimedia glossing
- Multimodal
- Vocabulary learning and glossing
- Glossing modes and word gains
- Single versus dual glossing modes
- Dual versus multi-glossing
- Hypertext context

Also, four of the earlier meta-analytic studies on glossing and L2 vocabulary acquisition (Abraham, 2008; Huang, 2010; Vahedi et al., 2016; Yun, 2011) were reviewed to retrieve relevant reports.

Selection Criteria

The following criteria were applied to include:

1. Studies that compared single versus dual and dual versus triple glossing modes.

2. Studies that adopted a between-participants design.
3. Studies that examined vocabulary gains.
4. Studies that were written in English.
5. All within-participants design studies and studies that focused on reading comprehension were excluded from this meta-analysis.

To have a comprehensive dataset, resembling true effects, this study included both published and unpublished works (e.g., two doctoral theses: Al-Ghafli, 2011; Salem, 2006). The researchers also contacted authors and gratefully received additional information to compute effect sizes. As a result, a total of 22 studies—n = 872 in experimental groups, n = 847 in control groups—were selected for subsequent coding and statistical analysis.

Coding

All 22 studies were coded as specified by the coding scheme table (see appendix). The authors encountered an issue regarding multiple effect sizes before aggregating them. Multiple effect sizes denote that the resulting mean effect size tends to be overestimated when samples come from the same participants or when they are dependent. This multiple effect size applies to the current data because the sample participants in some studies were mostly from multiple vocabulary measurements (e.g., multiple-choice vocabulary recognition and recall tests) and/or multiple glossing conditions (e.g., single, dual, and triple). Thus, the authors followed Plonsky and Oswald (2015) and earlier meta-analyses (e.g., Uchihara, Webb, & Yanagisawa, 2019) and averaged multiple effect sizes to yield a single effect size per sample (i.e., an independent pool of participants). Averaging effect sizes is advantageous over selecting one out of several overlapping effect sizes as it secures the dataset without losing valuable information. Table 1 shows characteristics of the studies included in this meta-analysis. However, single gloss mode for most studies (n = 20) included a textual gloss, except for Acha (2009) and Zarei & Mahmoodzadeh (2014), which used a single textual mode in one condition and a single pictorial mode in another condition. For purposes of consistency, the effect size from the single textual mode was used in the analysis. Furthermore, five categories of L2 moderator variables were created: (a) quality of data sample, (b) learner variables, (c) gloss features, (d) text features, and (e) methodological features. The following section explains each briefly.

Quality of data sample. The quality of the studies in this meta-analysis was coded for types of publication (e.g., journal articles, doctoral theses) and test-reliability measures. Researchers that utilized recognition test formats (e.g., multiple-choice task) often reported Cronbach's alpha (i.e., internal consistency), whereas those who implemented recall tests (e.g., translation task) reported interrater correlation.

Learner variables. Characteristics of the learners were coded for education (primary school, secondary school, university, language institute) and proficiency level (beginner/lower intermediate, intermediate, upper intermediate/advanced). It was challenging to find studies that clearly defined L2 proficiency or reported proficiency measures that could be compared across levels (Kang, Sok, & Han, 2018). Thus, we followed Yan, Maeda, Lv, and Ginther's (2016) procedure to code this variable, using reported proficiency grouping based on impressionistic

descriptors (e.g., beginner, intermediate, or advanced). Given this limitation, we adopted this approach to compare the present study with earlier meta-analyses on glossing modes that adopted the same approach (Vahedi et al., 2016; Yun, 2011).

Table 1. *Information on 26 Included Effect-Size Samples*

Study	Setting	L1	L2	ES (immediate)	ES (delayed)
Acha (2009)	FL	Spanish	English	–0.36	0.08
Al Ghafli (2011)	FL	Arabic	English	–0.07	–0.10
Akbulut (2007)	FL	Turkish	English	0.73	0.67
Salem (2006)	FL	English	Spanish	–0.09	0.09
Boers et al. (2017) sample 1	FL	Chinese	English	–0.22	Not tested
Boers et al. (2017) sample 2	SL	Malay	English	–0.39	Not tested
Boers et al. (2017) sample 3	FL	Dutch	English	–0.14	Not tested
Kost et al. (1999)	FL	English	German	0.62	0.72
Lin (2009)	FL	Chinese	English	0.71	0.31
Lin & Tseng (2012)	FL	Chinese	English	0.29	0.15
Moazzeni et al. (2014)	FL	Farsi	English	0.75	0.67
Moradan & Vafaee (2016)	FL	Farsi	English	2.48	Not tested
Plass et al. (2003)	FL	English	German	0.25	Not tested
Ramezanali & Faez (2019)	FL	Farsi	English	0.54	0.52
Sadeghi et al. (2016) sample 1	FL	Farsi	English	–0.22	Not tested
Sadeghi et al. (2016) sample 2	FL	Farsi	English	1.35	Not tested
Shahrokni (2009)	FL	Farsi	English	2.91	Not tested
Tabatabaei & Shams (2011)	FL	Farsi	English	1.33	Not tested
Türk & Erçetin (2014)	FL	Turkish	English	0.86	Not tested
Yanguas (2009)	FL	English	Spanish	0.43	0.04
Yeh & Wang (2003)	FL	Chinese	English	–0.16	Not tested
Yoshii (2006) sample 1	FL	Japanese	English	0.20	–0.03
Yoshii (2006) sample 2	FL	Japanese	English	0.47	0.48
Yoshii & Flaitz (2002)	SL	Mix	English	0.42	0.43
Warren et al. (2018)	SL	Mix?	English	0.04	Not tested
Zarei & Mahmoodzadeh (2014)	FL	Farsi	English	0.04	Not tested

Note. FL = foreign language; SL = second language; ES = effect size.

Gloss features. Two variables, gloss mode (single vs. dual and dual vs. triple) and language (L1, L2, L1&L2, and picture), were coded for gloss features. First, gloss mode serves as a primary moderator variable in this study in response to RQ2, which seeks to compare the effect of an additional gloss mode between single versus dual and dual versus triple. Second, gloss language was defined as the language used in the control condition (e.g., single mode group) in contrast to experimental conditions (single mode + additional mode).

Text features. The three variables related to text features included text type (narrative, expository), lexical coverage, and parts of speech (nouns only, verbs only, or combination). Lexical coverage was calculated by dividing the number of target words (M = 23.2, SD = 15.5) by text length (M = 658.9, SD = 317.5).

Methodological features. Three variables were coded for methodological features, including vocabulary measurement formats (form recall, form recognition, meaning recall, meaning recognition), computer use ("yes," "no"), and preknowledge control (pretest, pseudowords, other methods). In test format, recognition included tasks providing options for learners to choose from, such as a multiple-choice test with either L1 or L2 definitions (meaning recognition) and a checklist test (form recognition). Recall format involved producing L2 forms (form recall) or L1 translations (meaning recall), using translation or gap-filling tasks. In preknowledge control variable, categories include relying on students identifying unfamiliar words (Tabatabei & Shams, 2011), teachers' intuitive judgments of lexical difficulty (Salem, 2006), or the combination of both (Akbulut, 2007). For reliability, all studies were coded independently by two authors. Following Boulton and Cobb (2017), the discrepancies between the two authors' coding were tallied, and the agreement was rated at 99%. All disagreements were resolved through cross-checking the coding sheets, reviewing the relevant sections, and discussing them in several meetings.

ANALYSIS

This study used the Comprehensive Meta-Analysis Software (version 3.3) to compute the mean effect sizes and conduct moderator analyses. The researchers relied on Hedges's g as the basic unit of analysis. Hedges's g is the transformed version of Cohen's d for calculating the mean differences between experimental and control groups. In the present analysis, experimental groups were defined as groups that received a single mode + additional mode(s) (e.g., definition and pictures as dual glossing; definition, pictures, and audios as triple glossing). The control groups were single glossing modes (when compared with dual) and dual glossing modes (when compared with triple). Hedges's g was reported to offer more conservative calculations compared with Cohen's d, particularly for small sample sizes (Hedges & Olkin, 1985).

In addition, a random-effects model was employed to compute the weighted mean effect of 26 independent effect sizes, and then a mixed-effects model was used to conduct moderator analysis. In effect-size aggregation, the heterogeneity test was first conducted, using a Q statistic to examine the degree to which there was a significant variation in true effect sizes across studies. Finding a significant variation because of between-studies effects, the authors conducted moderator analysis with between-group Q statistic for predetermined moderator variables. For a continuous variable (e.g., lexical coverage), a meta-regression analysis was conducted with unrestricted maximum likelihood method. Although average mean differences were used for the majority of the moderator variables, to answer RQ2, a separate analysis of effect sizes from all glossing modes (i.e., dual and triple modes) was conducted. We followed Plonsky and Oswald's (2014) criteria to interpret the results of the effect sizes: small (0.40), medium (0.70), and large (1.00). In analyzing moderator variables, when the number of effect-size samples was smaller than three, the results based on such limited sample sizes were not interpreted. Prior to the mean

effect-size calculation and moderator analysis, outlier inspection and examination of publication bias were conducted, and the results confirmed that there was no serious issue related to publication bias or outliers (see Supplementary Information for the analysis and results).

RESULTS

The findings of the meta-analysis are presented in two sections: (b) overall impact of glossing modes for RQ1, and (c) moderator variables for RQ2 and RQ3.

Overall Benefit of an Additional Glossing Mode

To examine the overall effect of an additional glossing mode on L2 vocabulary learning, a weighted mean effect size and a 95% confidence interval (CI) were separately computed for immediate posttests (number of effect-size samples = 26) and delayed posttests (number of effect-size samples = 13). The average treatment-and-delayed-posttest intervals were 15.6 days. The findings revealed that the overall effect of an additional glossing mode was significant for immediate posttests with a medium effect (g = 0.46, CI = [0.20, 0.71], p < .001; Plonsky & Oswald, 2014). Also, the result showed that the overall effect of an additional glossing mode was significant for delayed posttests with a small effect (g = 0.28, CI = [0.12, 0.44], p = .001); however, the effect size was smaller than that of a previous meta-analysis (g = 0.35 in Yun, 2011). Furthermore, the heterogeneity test shows that the variation in the effect sizes between studies was significant for immediate posttests (Q = 163.4, p < .001), whereas the between-study variation for delayed posttests was not significant (Q = 17.8, p = .122). Thus, subsequent moderator analyses were conducted for the results of immediate posttests.
Table 2 summarizes the results of 10 categorical variables.

Table 2. *Moderator Analyses (Categorical Variables)*

Moderator	N_{ES}	Hedges's g	SE	LL	UL	Q
Quality of Data Sample						
Publication type						0.86
Journal article	23	0.50	0.14	0.23	0.77	
Doctoral thesis	3	0.13	0.38	–0.61	0.87	
Reliability						0.44
Reported	13	0.37	0.19	0.00	0.74	
Not reported	13	0.54	0.19	0.18	0.91	
Learner Variables						
Institutional level						10.80*
Primary school	1	–0.36	0.65	–1.63	0.91	
Secondary school	5	0.63	0.31	0.03	1.23	
University	17	0.28	0.16	–0.04	0.60	
Language institute	3	1.58	0.41	0.78	2.37	
						4.12

Proficiency						
Beginner/lower intermediate	7	1.02	0.27	0.50	1.54	
Intermediate	7	0.29	0.26	–0.21	0.80	
Upper intermediate/advanced	5	0.47	0.30	–0.13	1.07	
		Gloss Features				
Gloss mode						3.21
Single vs. dual	22	0.58	0.16	0.27	0.89	
Dual vs. triple	7	0.02	0.27	–0.50	0.55	
Gloss language						3.49
Additional mode added to L1 gloss	7	0.25	0.24	–0.23	0.72	
Additional mode added to L2 gloss	17	0.61	0.16	0.31	0.92	
Additional mode added to L1&L2	2	–0.12	0.44	–0.97	0.74	
		Text Features				
Text type						6.09*
Expository	8	0.82	0.21	0.40	1.23	
Narrative	13	0.15	0.17	–0.17	0.48	
Part of speech						4.35
Noun	7	–0.03	0.25	–0.52	0.46	
Verb	4	0.44	0.32	–0.18	1.06	
Mix	8	0.67	0.23	0.22	1.11	
		Methodological Features				
Test format						0.29
Form recall	9	0.44	0.24	–0.04	0.91	
Form recognition	3	0.61	0.40	–0.17	1.39	
Meaning recall	5	0.35	0.31	–0.24	0.95	
Meaning recognition	9	0.49	0.23	0.04	0.94	
Multimedia-based annotations						0.29
Yes	20	0.49	0.15	0.21	0.78	
No	6	0.32	0.28	–0.24	0.88	
Pre-knowledge control						5.34
Pretest	18	0.63	0.15	0.33	0.92	
Pseudoword use	4	–0.18	0.33	–0.83	0.47	
Other methods	4	0.26	0.33	–0.39	0.90	

Note. N_{ES} = number of effect sizes; *LL* = lower limit; *UL* = upper limit. Total number of effect sizes is not always equal because of missing data.
*$p < .05$.

Moderator Variables

To respond to RQ2 and RQ3, the moderator variables were first identified. Analyses were conducted in terms of the following five aspects to identify the factors that account for the variation in effect sizes across studies; the result of CIs crossing zero was interpreted as statistically nonsignificant ($p > .05$):

Quality of data sample. The two variables of publication type and reliability were considered important for this section:

Publication type. When type of publication was analyzed, there was no significant difference between journal articles ($g = 0.50$, CI = [0.23, 0.77]) and doctoral theses ($g = 0.13$, CI = [–0.61, 0.87]). Despite a statistically nonsignificant difference, the fact that published articles tend to provide larger effect sizes compared with unpublished theses implicates a trend toward overestimating the result (i.e., publication bias). Although the influence of publication bias was not considered serious as confirmed by previous preliminary analyses (i.e., trim-and-fill and fail-safe procedures), the result suggests that the current data might be a slight overestimate of the effect size in question.

Reliability. Although no significant difference was found between studies reporting test reliability ($g = 0.37$, CI = [0.00, 0.74]) and studies not reporting it ($g = 0.54$, CI = [0.18, 0.91]), the fact that the CI for the studies reporting test reliability included zero and a relatively large effect size for the studies not reporting test reliability is notable. It is not easy and also beyond the scope of this study to explore the possible reason why studies that reported test reliability did not yield any significant effect size (see Nakanish, 2015, for a similar finding reported by an earlier meta-analysis on extensive reading).

Learner variables. Two subcategories of institutional/educational levels and proficiency levels were analyzed.

Institutional level. As for the educational levels of learners, there was a significant variation in this category ($Q = 10.80$, $p < .05$). Learners in secondary school ($g = 0.63$, CI = [0.03, 1.23]) and language institutes ($g = 1.58$, CI = [0.78, 2.37]) gained substantial return from additional glossing modes, whereas university learners did not ($g = 0.28$, CI = [–0.04, 0.60]).

Proficiency level. Although the between-study variation did not reach statistical significance ($Q = 4.12$, $p > .05$), it is notable that lower proficiency learners (beginners) tended to benefit more from additional glossing modes ($g = 1.02$, CI = [0.50, 1.54]) than intermediate ($g = 0.29$, CI = [–0.21, 0.80]) or upper intermediate and advanced learners ($g = 0.47$, CI = [–0.13, 1.07]). Previous meta-analysis (Vahedi et al., 2016) showed that intermediate and advanced learners benefited more from an additional gloss mode than beginners, while the present meta-analysis showed the reverse.

Gloss features. Gloss features were analyzed in terms of gloss mode and gloss language.

Gloss mode. The difference between single versus dual (i.e., benefit of dual over single) and dual versus triple (i.e., benefit of triple over dual) missed reaching statistical significance ($Q = 3.21$, $p > .05$). However, according to the effect-size magnitude and confidence intervals, dual glossing mode was better than single glossing with a medium effect ($g = 0.58$, CI = [0.27, 0.89]), but the comparison between dual versus triple modes did not show a significant difference, and the effect ($g = 0.02$, CI = [–0.50, 0.55]) was negligible. These findings indicate that triple modes were not necessarily better than dual modes or vice versa, and suggest that adding another

glossing mode to a single mode may be more beneficial than adding another glossing mode to a dual mode.

Gloss language. Although a significant variation was not seen in this category ($Q = 3.49$, $p > .05$), remarkably, the effect of an additional mode was medium when learners' L2 was used in both conditions ($g = 0.61$, CI = [0.31, 0.92]). By contrast, the effect size for the L1 was small and the CI included zero or the effect was nonsignificant ($g = 0.25$, CI = [–0.23, 0.72]). This result indicates that adding another mode to an L2 textual gloss is more effective for vocabulary learning than adding another mode to an L1 textual gloss.

Text features. Three variables were analyzed as potential moderators for this category. They include text type, part of speech, and lexical coverage.

Text type. The two text types were expository and narrative. The results showed that the effect of an additional mode was larger when expository texts were used ($g = 0.82$, CI = [0.40, 1.23]) than narrative texts ($g = 0.15$, CI = [–0.17, 0.48]; $Q = 6.09$, $p < .05$).

Part of speech. Nouns, verbs, and a combination of different parts of speech were considered for this category. The findings showed that, interestingly, adding another gloss mode to the combination did not result in a significant difference between any of the speech parts for the glossed words, running counter to the expectation that the nouns could be learned easier than verbs or other word parts (Ellis & Beaton, 1993). No statistical significance was found either for verbs ($g = 0.44$, CI = [–0.18, 1.06]) or nouns ($g = –0.03$, CI = [–0.52, 0.46]), whereas the effect of additional glossing modes was significant for words with mixed parts of speech ($g = 0.67$, CI = [0.22, 1.11]).

Lexical coverage. Given that lexical coverage is a continuous variable, meta-regression analysis was conducted with unrestricted maximum likelihood method on lexical coverage as a predictor. The results showed that lexical coverage was not a significant predictor of vocabulary learning ($N_{ES} = 18$, $B = –0.03$, CI = [–0.10, 0.03]). There was a negative relationship between lexical coverage and additional glossing effects, signifying that the smaller the proportion of target words to text length, the higher the effect of additional glossing modes on L2 vocabulary learning (Figure 1). This also indicates a trend that texts with more new words are harder to comprehend for learners than those with fewer unfamiliar words, reducing the likelihood of vocabulary learning.

Methodological features. This moderator variable includes vocabulary test formats, multimedia-based annotations, and preknowledge control.

Vocabulary test format. The measurement types were form recall, form recognition, meaning recall, and meaning recognition. The findings showed that there was no significant difference across the four test formats ($Q = 0.29$, $p > .05$). With a closer examination of each test format, the effect of additional gloss modes appeared to be larger when knowledge of recognition was tested (form recognition, $g = 0.61$, CI = [–0.17, 1.39] and meaning recognition, $g = 0.49$, CI = [0.04, 0.94]) than when knowledge of recall was tested (form recall, $g = 0.44$, CI = [–0.04, 0.91] and meaning recall, $g = 0.35$, CI = [–0.24, 0.95]). The CIs including zero for all test formats except meaning recognition indicate that the positive effect of additional glossing modes is robust when knowledge of meaning recognition is tested.

Multimedia-based annotations. Although no significant difference was found regardless of whether glossed words were used in a multimedia-based learning environment or non-

multimedia-based settings, the effect appeared to be larger when multimedia-based annotations were used for target words (g = 0.49, CI = [0.21, 0.78]), compared with when a computer was not used (g = 0.32, CI = [–0.24, 0.88]). The CI including zero for noncomputerized condition indicates a nonsignificant effect for the additional glossing mode.

Preknowledge control. Although there was no significant difference across different approaches to controlling learners' preknowledge of target words (Q = 5.34, p > .05), it is notable that studies that used pretests showed a large effect of additional gloss modes on vocabulary learning (g = 0.63, CI = [0.33, 0.92]). Studies that used pseudowords showed a negligible effect (g = –0.18, CI = [–0.83, 0.47]), although caution needs to be taken to interpret this result due to the small sample size (i.e., four effect sizes) from two studies (Boers et al., 2017; Warren et al., 2018). Also, studies taking other methods (e.g., intuitive judgments of word difficulty) appeared to show a statistically nonsignificant effect of additional gloss modes (g = 0.26, CI = [–0.39, 0.90]).

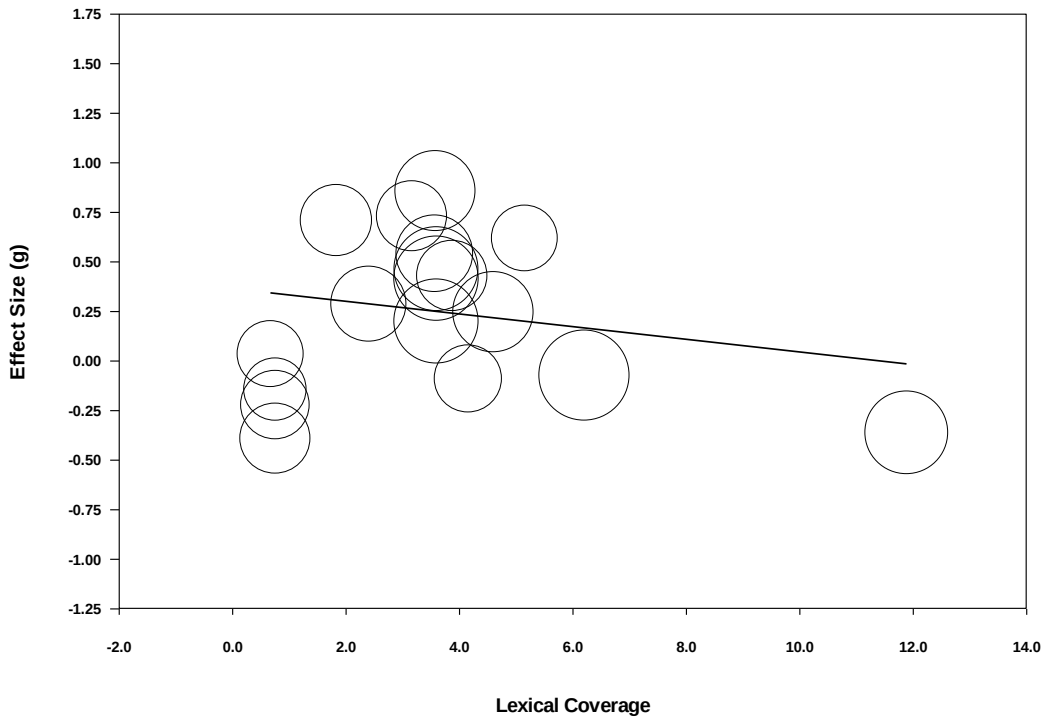

Figure 1. *The relationship between lexical coverage and effect sizes*

DISCUSSION

Overall Impact of Additional Glossing Modes

RQ1 asked if adding a glossing mode (e.g., a picture) to the combination of glossing annotations was beneficial for L2 learners' vocabulary gains. The findings showed that the overall effect sizes of additional glossing modes were medium (g = 0.46) for immediate and small (g = 0.28) for delayed posttests. Compared with other meta-analyses, reporting a moderate effect size (g = 0.41 in Yun, 2011) and a large effect size (g = 0.84 in Vahedi et al., 2016), the effect size for immediate posttest analyses (g = 0.46) here fits somewhere between the two earlier meta-analyses. We believe that this finding more accurately reflects the effect of an additional glossing mode on vocabulary learning than previous meta-analyses because the current meta-analysis focused on between-participants design, and avoided blending study designs, which may have led to biased effect sizes (Nakanish, 2015; Plonsky & Oswald, 2015). Notably, our study conducted a separate meta-analysis, exploring the effect of additional gloss modes on vocabulary retention measured through delayed posttests. The finding of a small effect size (g = 0.28) suggests that the relatively strong effect of additional modes on vocabulary uptake is not applied to vocabulary retention long term, and interpretations of the glossing effect should not be biased by the results of immediate posttests only. In vocabulary research, it is commonly accepted that glossing is effective for vocabulary learning. A good example is Vahedi et al. (2016) who reported an effect size of g = 0.84. We should note that the effect for delayed posttests is small (g = 0.28), particularly when compared with the large effect size (g = 0.84) found in Vahedi et al. (2016), following Plonsky and Oswald's (2014) effect-size criteria as small (0.40), medium (0.70), and large (1.00). Thus, the finding suggests that teachers and educators should not overestimate the efficacy of additional gloss modes on vocabulary learning for long-term retention when no follow-up activities are offered.

RQ2 asked if the addition of one glossing mode to single or dual combination would affect learners' word learning. The findings revealed that the combination of two glossing modes resulted in a better vocabulary gain for L2 learners than one; however, adding more gloss annotations (e.g., three modes) did not necessarily result in more effective word learning. The advantage of the dual over single mode in this study lends support to the available literature and cognitive theory of multimedia learning (Mayer, 2014). Previous studies suggested that bimodal glossing is more effective than single mode in promoting learners' vocabulary gains in multimedia-based settings (Abraham 2008; Al-Seghayer, 2001; Ramezanali & Faez, 2019; Yun, 2011).

Furthermore, theoretically, the rationale for the better performance in the dual glossing conditions is attributed to the fact that when learners encode target words in both visual and verbal formats, they can process the words in two channels; thus, learners can retrieve them better in more than one way (Mayer, 2014). The multimedia principle (Mayer, 2014) contributes greatly to our understanding of L2 vocabulary learning, suggesting that vocabulary gains can be enhanced by both verbal and visual exposure to information instead of either alone (Mayer, 2014).

One important point worth discussing here is why triple glossing modes were not more effective than dual glosses. One reason can possibly be attributed to learners' limited cognitive capacity to

process more than two forms of information (verbal and visual) at a time. When learners are presented with more than two modes, they could be overloaded by information (Sakar & Erçetin, 2005). However, considering the fact that we did not find any detrimental effect of additional glossing mode on learning (i.e., no statistically significant difference between dual and triple glosses), learners might have simply attended to only two modes and ignored the third one. In this sense, it is not surprising that we found that the triple glosses were as effective as dual glosses. Another reason is due to the split attention where learners are not able to construct a coherent mental connection between visual and verbal modes to process the words in memory (Mayer & Moreno, 1998).

Learner, Gloss, Text, and Methodological Variables

RQ3 asked to what extent learner variables, gloss language, text features, and methodological features moderate the overall benefit of an additional glossing mode for vocabulary gains. To address this question, the following variables were considered:

Institutional level and proficiency. With regard to institutional level, learners in secondary school (g = 0.63) and language institutes (g = 1.58) were found to benefit more from additional glossing modes in comparison to either learning in primary school (g = –0.36) or university (g = 0.28). Regarding L2 proficiency, lower proficiency learners (beginner and lower intermediate, g = 1.02) tended to learn more with the help of additional glossing modes than higher proficiency learners (intermediate, g = 0.29; upper intermediate and advanced, g = 0.47). These findings align with Yun (2011), suggesting that beginners benefited most from multi-glossing (g = 0.70). It is possible to argue that novice learners are not proficient enough to comprehend L2 texts without assistance nor to make sufficient use of contextual clues or a single mode gloss (e.g., L2 textual gloss). Therefore, such learners may need additional glossing modes like visual supports to increase the sources of input for word learning. In contrast, proficient learners might be able to collect contextual clues efficiently during reading and guess the meanings of unknown words without glosses. Consequently, when they are presented with a single glossing mode, they can learn words without multiple sources of input that denote the meaning of the target words.

Gloss language. Interestingly, we found that the effect of additional glossing modes on vocabulary learning was greater when an additional gloss mode was added to L2 textual glosses (g = 0.61, CI = [0.31, 0.92]) than when the same mode was added to L1 textual glosses (g = 0.25, CI = [–0.23, 0.72]). In other words, adding gloss modes to L2 single mode was effective, but the same degree of effectiveness was not found for the addition of another gloss mode to L1 single gloss. This result might relate to the fact that L1 textual glossing itself has a more powerful impact on vocabulary learning than an L2 textual gloss. A recent meta-analysis, investigating various types of single mode glosses (Yanagisawa et al., 2020), found that L1 glosses led to significantly larger gains than L2 glosses, regardless of learners' L2 proficiency. Because L1 single glosses themselves provide a strong support for word learning, then perhaps adding another gloss mode to L1 mode does not result in a greater word learning experience. However, this finding suggests that it is useful to add another type of gloss mode to L2 glosses.

Text type, part of speech, and lexical coverage. As for the text type, findings showed that additional glossing modes led to significantly larger gains in reading expository texts (g =

0.82) than narrative texts (g = 0.15). Compared with narrative texts, academic texts are carefully structured in terms of the logical relationship between sentences or paragraphs in a way that learners can draw on rich contextual clues to make an informed guess of the meanings of L2 words (Gardner, 2004). Thus, the positive effect of multi-glossing might be accentuated in reading expository texts. Second, the results for part of speech revealed that when target words were nouns only or verbs only, the effect of additional glossing modes was not significant (g = –0.03, CI = [–0.52, 0.46] and g = 0.44, CI = [–0.18, 1.06]), whereas it was significant when two or more parts of speech were involved (g = 0.67, CI = [0.22, 1.11]). Nonsignificant results for nouns and verbs might have been caused by relatively large standard errors for these two speech parts. Considering the widely accepted view that nouns are easier to learn (Ellis & Beaton, 1993) than other parts of speech, the negligible effect size for nouns is not surprising. Therefore, a single gloss mode might suffice for word learning for nouns without adding another mode. Third, although not statistically significant, we found a pattern of an inverse relationship between lexical coverage and additional glossing mode effects (see Figure 1), indicating that when learners read texts that contained many unfamiliar words, they tended to benefit less from additional modes of glossing. This finding points to the possibility that increasing vocabulary demand (i.e., larger unknown words in relation to text length) continues to pose challenges to word learning (Liu & Nation, 1985) even if learners' reading comprehension is supported by multi-glossing.

Test format, multimedia-based annotations, and preknowledge control. The results revealed that additional glossing modes might not be equally effective for learning different levels of word knowledge, and suggest the possibility that additional gloss modes might have a differential effect on word learning depending on aspects of word knowledge tested. Our data suggest that learning knowledge of recognition benefits more from additional glossing modes than learning knowledge of recall. Although direct comparison of this study to the previous meta-analysis is difficult due to differences in approaches to coding test formats, Yun (2011) also found medium to large effects for a range of recognition measures (g = 0.46 to 0.69) in comparison to medium-range effect sizes for production measures (g = 0.31 and 0.44). Evidence from vocabulary testing literature supports these findings as research has suggested that acquisition of recognition knowledge comes before that of recall knowledge (e.g., González-Fernández & Schmitt, 2019). In this regard, additional glossing modes might promote learning of recognition knowledge more quickly and easily than recall knowledge.

With respect to the variable of multimedia-based annotations, a relatively larger effect of additional glossing modes was found for multimedia-based conditions (g = 0.49) than non-multimedia-based settings (g = 0.32). A possible reason is that multimedia-based conditions often involve clicking or hovering a cursor over target words for textual or pictorial glosses to appear. This leads to retrieving target words (i.e., attempts to recall meanings of L2 words from memory) and, thus, promoting vocabulary learning (Webb & Nation, 2017). Yet, the situation differs in noncomputer settings where learners are presented with both target words and glosses simultaneously, not triggering lexical retrieval.

Finally, regarding preknowledge control, a larger effect size was found for studies that used pretests (g = 0.63), a negligible effect for studies using pseudowords (g = –0.18, CI = [–0.83,

0.47]), and a small but nonsignificant effect for studies using other methods (g = 0.26, CI = [–0.39, 0.90]). These findings indicate that different approaches to controlling prior knowledge have a substantial impact on the results of studies and conclusions to be made about the benefit of multimodal glossing. However, the result of pseudoword use needs to be interpreted with caution because it is based on two studies (Boers et al., 2017; Warren et al., 2018) that used non-word items limited in number (i.e., six items in both studies) and type (i.e., concrete nouns in both studies). Taken together, these findings imply that inconsistent procedures of checking learners' prior knowledge, including pretesting, using pseudowords, asking students to underline the words they did not know (e.g., Akbulut, 2007), and drawing on expert judgments (e.g., Salem, 2006), might have created unwanted measurement errors, influenced the study results, and produced inconsistent findings of effect sizes.

LIMITATIONS

There are several restrictions posed on this meta-analysis. First, methodologically, the authors were not sure how often triple glossing modes were utilized by participants in the studies that used more than two glossing modes. The absence of effectiveness of triple gloss modes in comparison to dual modes might be attributed to the possibility that, in most cases, learners do not use all three sources of information that denote word meanings. One way to explore this possibility is to investigate the extent to which each of three modes receives attention from learners by means of eye-tracking techniques (Warren et al., 2018) and how it relates to vocabulary learning gains. Second, depth of word knowledge is not researched in studies that have examined the impact of glossing modes. Given the multifaceted construct of word knowledge (Nation, 2013), the role of multimedia glossing in learning aspects of knowledge other than form-meaning connections (e.g., collocations, pronunciation) needs future investigation. Third, half of the studies included in this meta-analysis (13 out of 26/50%) have not reported test reliability (e.g., alpha). Such sporadic reporting practice of test reliability was also observed in incidental vocabulary learning research (Uchihara et al., 2019). Vocabulary researchers should be more attentive to test reliability in future research. Fourth, another limitation may concern the unbalanced numbers of published (n = 23) versus unpublished (n = 3) studies that did not provide strong argument for the results pertaining to publication bias. Fifth, the current meta-analysis did not suggest relative effectiveness of any specific combinations of gloss modes for vocabulary learning; rather, the study looked at a broadly defined gloss category, number of gloss modes (i.e., single, dual, and triple). Another study is required to investigate the effectiveness of different combinations of gloss modes, while the numbers of gloss modes are the same (e.g., audio + text vs. image + text).

PEDAGOGICAL IMPLICATIONS

This study provides important pedagogical implications for vocabulary teaching and learning. First, adding more gloss modes does not always enhance vocabulary learning. The findings of this meta-analysis showed that the presentation of two modes together using textual and aural or pictorial modes was as beneficial for vocabulary learning as that of three modes together. Given the effectiveness of dual glosses and the extra workload involved in adding a third mode, teachers can rely on using dual glosses rather than triple glosses for teaching vocabulary. Yet, the question still remains as to which gloss modes should teachers use to teach new words, single or

dual? One potential response depends on if the purpose of glossing is to understand a written text or to learn new words. For reading comprehension activities, if teachers place more weight on text comprehension, provision of single glosses (especially through L1 translation) might suffice. If more pedagogical weight is given to vocabulary learning, teachers can annotate the unfamiliar words using two gloss modes (e.g., text plus picture).

Second, it is important for teachers to realize that the effectiveness of additional glossing modes might depend on the choice of language for which textual definitions are provided. Our findings showed that additional glossing modes (e.g., pictorial gloss) were more effective when added to L2 textual glosses than L1 textual glosses. Given that L1 textual glosses are more effective for learning than L2 textual glosses (Yanagisawa et al., 2020), L1 single glosses alone are effective enough to build knowledge of form-meaning connections. However, when the goal is to build knowledge beyond form-meaning connections (e.g., knowledge of word relations), learners might benefit from the support of L2 textual glosses when accompanied with pictorial glosses. Exposing learners to definitional sentences in L2 might help develop knowledge of semantic network of words, such as synonyms, antonyms, or superordinate words (Baba, 2009).

Finally, the choice of gloss modes should be informed by L2 learners' proficiency levels. Our finding showed that additional glossing modes are beneficial for low-proficiency learners. This suggests that when teaching L2 beginners, adding another gloss mode (e.g., pictures) to a single textual gloss is highly recommended even though the use of three or more modes is not necessary. Nevertheless, the value of single textual glosses (especially, L1 glosses) should not be underestimated. L2 single glosses might not be effective for all but advanced learners (Ko, 2017); yet, L1 single glosses were found to be equally effective for learners regardless of proficiency levels (Yanagisawa et al., 2020). Teachers' efforts to assure L2 textual glosses are comprehensible to learners should also be considered. Keeping the research findings and practical considerations in mind, we recommend using L1 single textual glosses for all learners, in most cases, where possible. However, in some situations such as teaching low-proficiency learners with diverse L1 backgrounds or when a pedagogical focus is on developing learners' word knowledge beyond form-meaning connections, L2 textual glosses plus additional gloss modes are advised.

REFERENCES[1]

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning, 21*(3), 199–226. https://doi.org/10.1080/09588220802090246

*Acha, J. (2009). The effectiveness of multimedia programmes in children's vocabulary learning. *British Journal of Educational Technology, 40*(1), 23–31. https://doi.org/10.1111/j.1467-8535.2007.00800.x

*Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science, 35*(6), 499–517. https://doi.org/10.5539/elt.v6n12p48

*Al-Ghafli, M. H. (2011). *The effect of mediated glosses on vocabulary retention and reading comprehension with learners in Saudi Arabia* [Unpublished doctoral dissertation]. University of Kansas.

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

Ariew, R., & Erçetin, G. (2004). Exploring the potential of hypermedia annotations for second language reading. *Computer Assisted Language Learning, 17*(2), 237–259. https://doi.org/10.1080/0958822042000334253.

Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing, 18*(3), 191–208. https://doi.org/10.1016/j.jslw.2009.05.003

*Boers, F., Warren, P., He, L., & Deconinck, J. (2017). Does adding pictures to glosses enhance vocabulary uptake from reading? *System, 66*, 113–129. https://doi.org/10.1016/j.system.2017.03.017

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning, 43*(4), 559–617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x

Gardner, D. (2004). Vocabulary input through extensive reading: A comparison of words found in children's narrative and expository reading materials. *Applied Linguistics, 25*(1), 1–37. https://doi.org/10.1093/applin/25.1.1

González-Fernández, B., & Schmitt, N. (2019). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics*. https://doi.org/10.1093/applin/amy057

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic

[1] References marked with an asterisk indicate studies included in the present meta-analysis.

Press.
Huang, S.-F. (2010). *Effects of tasks and glosses on L2 incidental vocabulary learning: Meta-analysis* [Unpublished doctoral dissertation]. Texas A&M University.
In'nami, Y., & Koizumi, R. (2010). Database selection guidelines for meta-analysis in applied linguistics. *TESOL Quarterly*, *44*(1), 169-184. https://doi.org/10.5054/tq.2010.215253
Jung, J. (2016). Effects of glosses on learning of L2 grammar and vocabulary. *Language Teaching Research, 20*(1), 92–112.
Kang, E. Y., Sok, S., & Han, Z. H. (2018). Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research,* 1–26. https://doi.org/10.1177/1362168818776671
Khezrlou, S., & Ellis, R. (2017). Effects of computer-assisted glosses on EFL learners' vocabulary acquisition and reading comprehension in three learning conditions. *System, 65*, 104–116. https://doi.org/10.1016/j.system.2017.01.009
Ko, H. M. (2005). Glosses, comprehension, and strategy use. *Reading in a Foreign Language, 17*(2), 125–143.
Ko, H. M. (2017). The relationship between gloss type and L2 proficiency in incidental vocabulary learning, *Modern English Education, 18*(3), 47–69. https://doi.org/10.18095/meeso.2017.18.3.03
*Kost, C. R., Foss, P., & Lenzini, J. J. (1999). Textual and pictorial glosses: Effectiveness on incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals, 32*(1), 89–113. https://doi.org/10.1111/j.1944-9720.1999.tb02378.x
Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research, 15*(4), 391–411.
*Lin, C. (2009). Learning action verbs with animation. *The JALT CALL Journal, 5*(3), 23–40.
*Lin, C., & Tseng, Y. (2012). Videos and animations for vocabulary learning: A study on difficult words. *TOJET: The Turkish Online Journal of Educational Technology, 11*(4), 346–355.
Liu, N., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal, 16*(1), 33–42. https://doi.org/10.1177/003368828501600103
Mayer, R. E. (Ed.). (2014). *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.
Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312–320. http://dx.doi.org/10.1037/0022-0663.90.2.312
*Moazzeni, Z., 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.
Mohsen, M. A., & Balakumar, M. (2011). A review of multimedia glosses and their effects on L2 vocabulary acquisition in CALL literature. *ReCALL, 23*(02), 135–159.
*Moradan, A., & Vafaei, M. (2016). The effect of glosses on incidental vocabulary learning of Iranian EFL learners. *International Journal of Applied Linguistics and English Literature, 5*(6), 34–42. https://doi.org/10.7575/aiacijalel.v.5n.6p.34
Nakanishi, T. (2015). A meta-analysis of extensive reading research. *TESOL Quarterly, 49*, 6–37. https://doi.org/10.1002/tesq.157

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

*Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*, 221–243. https://doi.org/10.1016/S0747-5632(02)00015-8

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079

Plonsky, L., & Oswald, F. L. (2015). Meta-analyzing second language research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 106–178). New York, NY: Routledge.

*Ramezanali, N., & Faez, F. (2019). Vocabulary learning and retention through multimedia glossing. *Language Learning & Technology, 23*(2), 105–124. https://doi.org/10125/44685

*Sadeghi, K., Khezrlou, S., & Modirkhameneh, S. (2016). Calling Iranian learners of L2 English: Effect of gloss type on lexical retention and reading performance under different learning conditions. *Journal of Research in Reading, 39*(1), 1–21. https://doi.org/10.1111/1467-9817.12088

Sakar, A., & Erçetin, G. (2005). Effectiveness of hypermedia annotation for foreign language reading. *Journal of Computer Assisted Learning, 21*, 28–38. https://doi.org/10.1111/j.1365-2729.2005.00108.x

*Salem, B. E. (2006). *The influence of electronic glosses on word retention and reading comprehension with Spanish Language Learners* [Doctoral dissertation]. University of Kansas. Retrieved from http://www.edtech.ku.edu/research/gloss/publications/bensalem.pdf

*Shahrokni, A. (2009). Second language incidental vocabulary learning: The effect of online textual, pictorial, and textual pictorial glosses. *The Electronic Journal for English as a Second Language (TESL-EJ), 13*(3), 1–17.

Stewart, R. A., & Cross, T. L. (1991). The effect of marginal glosses on reading comprehension and retention. *Journal of Reading, 35*(1), 4–12.

*Tabatabaei, O., & Shams, N. (2011). The effect of multimedia glosses on online computerized L2 text comprehension and vocabulary learning of Iranian EFL learners. *Journal of Language Teaching and Research, 2*(3), 714–725. https://doi.org/10.4304/jltr.2.3.714-725

*Türk, E., & Erçetin, G. (2014). Effects of interactive versus simultaneous display of multimedia glosses on L2 reading comprehension and incidental vocabulary learning. *Computer Assisted Language Learning, 27*(1), 1–26. https://doi.org/10.1080/09588221.2012.692384

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning, 69*(3), 559–599. https://doi.org/10.1111/lang.12343

Vahedi, V. S., Ghonsooly, B., & Pishghadam, R. (2016). Vocabulary glossing: A meta-analysis of the relative effectiveness of different gloss types on L2 vocabulary acquisition. *Teaching English with Technology, 16*(1), 3–25.

*Warren, P., Boers, F., Grimshaw, G., & Siyanova-Chanturia, A. (2018). The effect of gloss type on learners' intake of new words during reading: Evidence from eye-tracking. *Studies in Second Language Acquisition, 40*(4), 883–906.

https://doi.org/10.1017/S0272263118000177
Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*(2), 232–245.
Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford, England: Oxford University Press.
Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing, 33*(4), 497–528. https://doi.org/10.1177/0265532215594643
Yanagisawa, A., Webb, S., & Uchihara, T. (2020). How do different forms of glossing contribute to L2 vocabulary learning from reading? A meta-regression analysis. *Studies in Second Language Acquisition.*http://doi10.1017/S0272263119000688
*Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning and Technology, 13*(2), 48–67.
*Yeh, Y., & Wang, C. (2003). Effects of multimedia vocabulary annotations and learning styles on vocabulary learning. *CALICO Journal, 21*(1), 131–144. https://doi.org/10.1558/cj.v21i1.131-144
*Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning and Technology, 10*(3), 85–101.
*Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of picture and annotation types. *CALICO Journal, 20*(1), 33–58. https://doi.org/10.1558/cj.v20i1.33-58
Yun, J. (2011). The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning, 24*(1), 39–58. https://doi.org/10.1080/09588221.2010.523285
*Zarei, A., & Mahmoodzadeh, P. (2014). The effect of multimedia glosses on L2 reading comprehension and vocabulary production. *Journal of English Language and Literature, 1*(1), 1–7. https://doi.org/10.17722/jell.v1i1.6.