Comparing the Effects of Freely Available Online Dictionaries and Printed Glosses on Vocabulary Acquisition

Dennis Laffey

(Pukyong National University)

Laffey, Dennis. (2019). Comparing the effects of freely available online dictionaries and printed glosses on vocabulary acquisition. *English Teaching, 74*(2), 27-49.

Few dictionary studies have examined the effectiveness of freely available online dictionaries (FAOD) offered by search engine companies. This experimental study examines the effectiveness of one FAOD, Naver.com, as an aid to vocabulary acquisition for a group of 87 intermediate-level Korean university students aged 19 to 28 (mean 21.9). The VKS and two multiple choice tests administered after a written recall task, and again one month later, measured vocabulary acquisition. MANOVA using treatment type (Online, Gloss, or Control) as the independent variable and test results as the dependent variables compared vocabulary acquisition. A second line of inquiry examined effects of topical familiarity on vocabulary acquisition. Findings suggest that FAODs are equally effective as a written gloss, and that both significantly outperform using context clues. Evidence is presented that unfamiliar topics may lead to slightly better vocabulary gains over reading familiar topics. Ways these findings can assist learners and teachers are discussed.

Key words: second language acquisition, computer-assisted language learning, reading, vocabulary

1. INTRODUCTION

Should language teachers allow their learners to use smartphones in the classroom as an aid to language learning? With an estimated 39 million smartphone users (Statista, 2017) in a nation with approximately 50 million citizens, Korea has one of the highest adoption rates of smartphones in the world. Koreans' passion for studying English is also well documented (Seth, 2002). It should be no wonder that many Korean university students rely on their smartphones to help them with English language learning. In particular, the

researcher has noted that the majority of his students utilize the online bilingual dictionary offered by the Korean search engine Naver (http://dic.naver.com) when encountering unknown English words or translating Korean words into English.

The effectiveness of freely available online dictionaries (FAOD) in assisting learners to comprehend texts and acquire vocabulary compared to more traditional forms of paper or electronic dictionaries have yet to be seriously investigated. Lew (2015) suggests that the lack of serious investigation into FAOD may be due to a focus on surveys of academic and expert dictionary users rather than quantitative experiments with more casual dictionary users. The slow pace of publishing research compared to the speed of innovations and developments in technology and online dictionary services may also be a factor. "[W]hile the move from print to digital dictionaries has been quite vigorous, research into dictionary use has been somewhat slow in joining the transition" (Lew, 2015, p. 232). The experiment reported here seeks to examine how effective FAODs are compared to a written gloss or reliance on context clues alone for L2 learners of English, as little if any research has yet to be done on this subject.

2. REVIEW OF THE LITERATURE

2.1. Language Learner Dictionary Use Behaviors

Past studies of dictionary use focused on effects on comprehension, production and language learning, as well as on learner behaviors and preferences (Nesi, 2014). Some studies (Lupesscu & Day, 1993; Nesi & Haill, 2002) found that certain words were prone to semantic error when learners looked them up in a dictionary, suggesting that learners may have trouble distinguishing the correct sense or grammatical function of an unknown word being looked up. Nyikos and Fan (2007), summarizing previous research, suggest that less proficient learners use dictionaries more often but less effectively, while the opposite holds for more proficient learners. Despite the drawbacks of using a dictionary while using the target language, learners in general find dictionaries useful when they encounter new words (Schmitt, 1997).

This research focuses on dictionaries as an aid to vocabulary acquisition, which is not a binary distinction. Ellis (1997) divides vocabulary knowledge into formal knowledge: knowing phonetic sounds, orthographic characters, and/or syntactical functions and connections of a word, and semantic knowledge of a word: grammatical word class, and various semantic meanings of the word. In this study, partial vocabulary acquisition is considered the gaining of formal, non-semantic word knowledge. Vocabulary acquisition is considered gaining formal, non-semantic word knowledge and at least one semantic

interpretation of the word, based on Ellis (1997).

2.2. Effects of Print Dictionaries on Language Learning

In an early study of the effects of dictionary use on reading comprehension, Bensoussan, Sim and Weiss (1984) found no difference in test scores between those that performed a reading task without a dictionary and those that used a dictionary. Luppescu and Day (1993) found that dictionary use increased the time needed to read and comprehend a passage. Knight (1994) found dictionary use contributed to reading comprehension, especially with lower proficiency learners. Loucky (2002) suggests that the early criticisms of dictionary look-up times hindering comprehension are less valid when using computerized dictionaries of any type. The consensus seems to suggest that "[d]ictionary use helps learning and comprehension, and is particularly useful for learners who do not cope well with guessing from context" (Nation, 2013, p. 417).

Luppescu and Day's (1993) study mentioned above did find benefits for vocabulary acquisition for learners using a dictionary. Knight (1994) also found that dictionary use contributed to vocabulary gains. Laufer and Hadar (1997) found that most learners could use learner dictionaries with both L1 and L2 entries more effectively than L1 only or L2 only dictionaries for vocabulary comprehension and production, with only the most advanced learners gaining equivalent results using L1 only dictionaries. Laufer and Hill (2000) suggest that learners acquire vocabulary from dictionaries best when they utilize multiple types of dictionary information during a look-up. While most researchers report dictionaries having a positive effect on vocabulary acquisition, individual learner factors such as language proficiency level, dictionary strategy use, and type of dictionary used seem to be factors affecting vocabulary gains from dictionary look-ups.

2.3. Computerized Glossing Programs

Many recent studies have examined the effectiveness of interactive glossing software and differences in learner outcomes based on these computerized glosses. Bowles (2004) found no difference in results when using a paper-based gloss and a computerized gloss, while both experimental groups performed better than the control which had no gloss. Positive benefits for vocabulary acquisition have been noted for subjects using computerized glossing software (Abraham, 2008; Fageeh, 2014; Kilickaya & Krajka, 2010; Ko, 2012). Abraham's (2008) meta-analysis of computer mediated glossing studies also reports a net positive effect of computerized glosses on reading comprehension.

While computer mediated glosses have benefits to learners, they come with a serious drawback. At the present time, many of the researcher's learners communicate and read for

pleasure via computer or smartphone, but primarily read printed text for their studies. Computer mediated glossing software may be inconvenient at best and totally impractical at worst when reading printed text. Computerized glosses are also less than ideal for assisting with verbal communication. FAOD such as those provided by most major internet search engines are easily available via computer, smartphone, or tablet. While not practical in all verbal communication situations, FAODs can also be used to query unknown words which are heard if time allows, such as during casual conversation.

Freely available FAODs suffer the same drawbacks of traditional print dictionaries, one being that they are designed for general reference rather than as an aid to language learning (Nation, 2013; Nation & Webb, 2011). Despite this, the researcher has observed that learners in Korea rely on them a great deal. Studies into the effectiveness of these dictionaries similar to the studies of computerized glossing programs mentioned above will benefit both educators and students, as the strengths and weaknesses of this form of dictionary are studied. The results may also be of interest to the search engine providers and online dictionary websites, assuming they wish to increase functionality to their users by offering learner-centric dictionary entries in addition to standard reference entries in their freely available FAODs.

2.4. Effects of Topical Familiarity

Learners' background knowledge plays a role in reading comprehension. "L2 readers who are not familiar with content schema or do not possess appropriate L2 sociocultural knowledge will have comprehension difficulties in that they cannot perceive the L2 texts in a culturally authentic way" (Erler & Finkbeiner, 2007, p. 198). Learners with greater knowledge of a topic before reading should achieve better comprehension scores than on a topic with which they are unfamiliar. For this study, topical familiarity is defined as content which can reasonably be considered common knowledge or background knowledge for a given population.

Considering vocabulary acquisition, Nation and Webb (2011) suggest that bottom-up processing may be more effective than top-down methods such as applying background schemata. Haastrup (1989, as in Nation, 2013) suggests that difficulty in guessing the meaning of a word leads to greater retention. In a study of L1 contextual learning, Nagy, Anderson and Herman (1987) found that a lack of conceptual knowledge of a target word hindered acquisition. Pulido (2009), likewise, found benefits of topical familiarity, which resulted in greater vocabulary gains from familiar texts for L2 learners. Fraser (1999) found that guessing meaning from context followed by a dictionary look-up resulted in greater gains than guessing alone. If Pulido (2009) and Fraser (1999) are correct, the combination of guessing context clues from prior knowledge combined with a dictionary

look-up from an FAOD or printed gloss should result in greater vocabulary gains from a topically familiar text than from an unfamiliar text. Both an FAOD and a printed gloss should outperform using context clues alone, regardless of topical familiarity. If the current research aligns with this hypothesis, it would suggest that reading topically familiar texts with the assistance of an FAOD may be an optimal condition for vocabulary acquisition for learners, because the majority of texts learners encounter do not contain a printed gloss.

2.5. Research Questions

If FAODs are shown to assist learners with both comprehension and vocabulary acquisition as well as traditional dictionary look-ups or glossing, educators may place a greater reliance on them. If they are shown to be deficient in some way, then educators may steer learners away from them or at least educate learners to the drawbacks of these convenient forms of dictionary. With empirical evidence, the makers of these freely available FAODs may be able to improve their programs for learners.

This study compares rates of vocabulary acquisition for Korean university students with an intermediate level of proficiency when using a popular Korean FAOD, Naver.com, a printed gloss, or no dictionary in a reading recall task. A second line of inquiry looks into any possible effects that topical familiarity may have on vocabulary acquisition from the reading task. The empirical study presented in this paper seeks to answer these two questions:

1. How well do FAODs assist L2 English learners' immediate acquisition and retention of new vocabulary from a reading task?
2. Does topical familiarity of the reading task affect comprehension and vocabulary acquisition when assisted by a FAOD, a printed gloss, or no dictionary?

3. METHOD

3.1. Participants

The participants in this study were 87 undergraduate students at a national university in Korea, attending low-intermediate, high-intermediate and advanced English conversation courses taught by the researcher and a colleague. All but one of these participants were native speakers of Korean, with one participant of Chinese nationality proficient in Korean. They ranged in age from 19 to 28, with a mean of 21.9 years old. The *Online* group ranged in age from 19 to 25, with a mean age of 21.8. The *Gloss* group ranged in age from 20 to

27, with a mean age of 21.8. The *Control* group ranged in age from 19 to 28 years old, with a mean age of 22.2.

The university does not require students to take lower level conversation classes as requirements for upper level classes, so a mix of students from various grade years can be found in each class. Fifty participants listed a TOEIC score, which ranged from a low of 600 to a high of 990, with a mean score of 850. Two participants listed TOEFL paper test scores of 770 and 940 (mean 855), while seven listed TOEFL computer test scores ranging from 61 to 93 (mean 75). Four participants listed an IELTS score, ranging from 6 to 7.5 (mean 6.5). This suggests the population was of the "independent user" [B1/B2] level of proficiency by the European Framework scale. Demographic data of the participants is located in Table 1.

TABLE 1

Participant Demographic Data

Group	Female	Male	Other Gender	2nd year student	3rd year student	4th year student	Overseas >1yr	TOEIC Mean Score
Total	65	21	1	14	51	20	9	850
(n=87)	74.7%	24.1%	1.1%	16.1%	58.6%	23.0%	10.3%	(n=50)
Online	21	8	0	3	21	5	1	831
(n=29)	72.4%	27.6%	0%	10.3%	72.4%	17.2%	3.4%	(n=12)
Gloss	25	3	1	4	19	4	4	842
(n=29)	86.2%	10.3%	3.4%	13.8%	65.5%	13.8%	13.8%	(n=17)
Control	19	10	0	7	11	11	4	825
(n=29)	65.5%	34.5%	0%	24.1%	37.9%	37.9%	13.8%	(n=21)

3.2. Design

In order to test the effects of using a free online dictionary on vocabulary acquisition, two reading passages were prepared. The first passage, *Battle of Myeongnyang*, was selected because it would be topically familiar to the participants. The second passage, *The King of the Blues*, was selected because it was likely to be topically unfamiliar to the participants. Battle of Myeongnyang was modified from the account of the battle on the Wikipedia page of Yi Sun-shin (https://en.wikipedia.org/wiki/Yi_Sun-sin, accessed March, 2017). *The King of the Blues* was modified from the official B.B. King website (http://www.bbking.com/, accessed March, 2017). Both passages were edited for passage length, sentence length, and vocabulary. Target vocabulary items for the study, discussed below, were substituted for existing words in the texts. The reading level of each passage was checked, and further modifications were made until both passages were roughly of equal length and difficulty. The details of each passage can be found in Table 2. The

passages were analyzed using the online Compleat Web VP! (http://www. lextutor.ca/ vp/comp/) and Readability Calculator (https://www.online-utility.org/english/readability_ test_and_improve.jsp).

TABLE 2

Reading Passage Comparison

Criteria	Battle of Myeongnyang	The King of the Blues
Topic Familiarity	familiar	unfamiliar
Number of Words	178	180
Number of Sentences	18	15
Average Words per Sentence	9.89	12.00
Flesh-Kinkaid Grade Level	6.70	6.72
Pausal Units	29	34
Target Words	8	8
Coverage (2,000-word level)	89.3%	90.6%
Coverage (5,000-word level)	95.5%	95.6%

Once the passages were selected, appropriate infrequent words were found that could be used within the contexts of the two passages. Eight target words were selected to insert into each passage, so that the text coverage, assuming the participants had vocabularies of 5,000-word families, would be around 95%. The assumption of a 5,000-word family vocabulary among the subjects was made based on data from a previous study (Laffey, 2016) among a similar body of students. Some of the participants in the previous study took part in this study as well. The scores of the participants who provided them suggest that the more proficient among them may be at this level while the less proficient participants would be below this level of vocabulary coverage (ETS, 2015; Gostudylink, n.d.; Milton & Alexiou, 2009). A 5,000-word family allows enough text coverage for comprehension of the texts but the unknown vocabulary could still hinder comprehension if not looked up (Nation, 2015; Nation & Waring, 1997). The remaining text was simplified as much as possible, replacing as many mid-frequency words as possible. The text coverage for a reader with a vocabulary of only 2,000-word families was around 90%.

The target words selected for the study are all of frequency band 10 or rarer as rated by the Compleat Web VP!. Because the experiment relies on commercially produced online dictionaries, pseudo-words could not be used in this study, running the risk that some of the target vocabulary may have been encountered by the participants before. The target words were checked by three groups for familiarity: an undergraduate class of Korean students of similar level to the participants, a Masters level graduate class of non-native English-speaking students, and three native speaking colleagues. Only one undergraduate, who had spent extensive time overseas, was familiar with any of the target words [altercation, conflagration, enjoin, honky tonk, lopsided]. Of the graduate students, none

claimed familiarity, but some words were guessable to them. The native English-speaking colleagues even found some of the words unfamiliar, and believed that their students would be unlikely to know most, if any, of the words. Data on the target words is found in Table 3.

TABLE 3
Target Words

Target Word	Word Class	Text	Frequency Band
altercation	noun	unfamiliar	12,000
appellation	noun	unfamiliar	13,000
approbation	noun	unfamiliar	14,000
conflagration	noun	unfamiliar	12,000
conspicuity	noun	familiar	off list
delimit	verb	familiar	11,000
enjoin	verb	familiar	11,000
girding*	adjective	familiar	13,000
honky tonk	noun	unfamiliar	off list**
inveigle	verb	familiar	17,000
lopsided	adjective	familiar	11,000
nonpareil	adjective	unfamiliar	21,000
pittance	noun	unfamiliar	12,000
sanguine	adjective	familiar	10,000
sobriquet	noun	unfamiliar	15,000
unpropitious	adjective	familiar	12,000

Note. *removed from analysis; **while the compound is off list, *honky* is band 16,000 and *tonk* is band 18,000.

Three treatment conditions of each reading passage were prepared. Version A, for the *Control* group, had instructions to read the passage for a comprehension test, but looking up words was forbidden. Participants were asked to circle any words they did not know as they read. Version B, for the *Gloss* group, contained similar instructions and also included a written gloss with each target word and its Korean translation, as found in the Naver online dictionary (http://dic.naver.com). Version C, for the *Online* group, consisted of instructions to read the passage for a comprehension test, use a smartphone to look up any unknown words using Naver, and also to circle any words that were looked up online. Naver's FAOD was selected as the dictionary of choice for this experiment because the majority of participants taking the researcher's course reported using it on a daily basis, giving them familiarity with the interface and the format of entries in the FAOD.

A written recall instrument was used to introduce the target vocabulary and ensure an adequate level of comprehension for vocabulary acquisition to occur, containing instructions in Korean on an otherwise blank sheet of paper. Three instruments were used to measure different levels of vocabulary acquisition, following Joe (1998). The first was a Vocabulary Knowledge Scale (VKS) test (Wesche & Paribakht, 1996) consisting of 16

items. Cronbach's alpha for the VKS was .86. The second was a Word Associations test intended to indicate partial semantic knowledge, consisting of 16 items and with Cronbach's alpha .78. The third was a Word Meaning test intended to indicate semantic knowledge of each target word in the sense used in the passages consisting of 16 items and with Cronbach's alpha .71.

Example test items for all three tests, in both English and Korean, are shown in Figure 1 below. Both the Word Association test and Word Meaning test were multiple choice tests designed with the target words as the stem and the correct answer, three distractors, and an 'I don't know' option as alternatives (Joe, 1998). The Word Association test listed each alternative in Korean and English. The Word Meaning test listed each alternative only in Korean. Korean translations for both multiple choice tests were taken from Google's freely available browser-based online English to Korean translation tool (https://www.google.com/). The translation and definition that the participants would see is from Naver.com, and an example is presented in Figure 2. To mitigate ordering effects, two versions of each multiple-choice test were prepared with the items in an opposite order.

3.3. Procedure

The researcher conducted the experiment during normal class hours. The students received instructions that the following activity was voluntary, and declining would not affect their grades, but participation would be rewarded with extra credit points. Every participant signed a consent form after the basic procedure was explained. Test packets were distributed randomly, and the participants were informed that each had one of three types of test: *Control* in which only context clues could be used, *Gloss* which had a gloss of the target vocabulary, or *Online* in which a smartphone could be used to access Naver's dictionary. Participants who received the *Online* packet but did not have a smartphone were asked to trade packets with another participant before beginning, however this turned out to be a non-issue as all participants had a smartphone available.

Once the packets were distributed, the researcher instructed the participants to remove only the yellow sheet from the packet, read it, and try to remember as much as they could for a written recall test later. There was no mention of a vocabulary test. In order to mitigate ordering effects, half of the participants received the familiar text on yellow paper while the other half received the unfamiliar text on yellow paper. After seven minutes, participants were asked to return the yellow sheet to the packet and remove one of the white sheets, and to write as much as they could remember from the story in Korean. Once this was finished, participants returned the written recall sheet to the packet. The activity was then repeated using the blue sheet. Participants who had the familiar text on the yellow sheet had the unfamiliar text on the blue sheet, and vice versa, so that the two reading

comprehension tasks were completed back to back.

FIGURE 1

Example Items from the VKS, Word Association and Word Meaning Tests

VKS Test Items	Word Association Test Items	Word Meaning Test Items

sobriquet
 1. I don't remember having seen this word.
 2. I have seen it, but am not sure what it means.
 3. I think I know the meaning: _____
 4. I can use the word in a sentence (also complete step 3):
 _____
	12. sobriquet
	   a. music
	   b. personality
	   c. voice
	   d. a name
	   e. don't know
		6. sobriquet
		   a. a performance name
		   b. a job title or work position
		   c. a military rank
		   d. a false identity
		   e. don't know

sobriquet
 1. 이 단어를 본 적이 없음.
 2. 본 적이 있지만 무슨 뜻인지 모르겠음.
 3. 단어 뜻을 아는 거 같음: _____
 4. 이 단어를 써서 문장을 만들수 있음.
 (또한 3 번빈칸에 단어뜻을 쓰세요.)
 _____
	12. sobriquet
	   a. 음악 music
	   b. 성격 personality
	   c. 목소리 voice
	   d. 이름 a name
	   e. 모르겠다 don't know
		6. sobriquet
		   a. 예능인의 예명
		   b. 직책
		   c. 군사 계급
		   d. 거짓 신원
		   e. 모르겠다

FIGURE 2

Example of Naver.com Dictionary Entry

After the second written recall was complete, the VKS was administered, with instructions of how to complete the test, including examples of each level of word knowledge. The VKS was administered first to avoid inflating word familiarity or word recognition scores as well as semantic knowledge scores due to familiarity from the other tests. The Word Association test was administered next, with instructions on how to

complete the test, an example problem, and advice to be honest and circle "E. I don't know" if they were not sure of an answer rather than guessing. The Word Meaning test was then administered in the same fashion. The Word Meaning test was administered last to avoid artificially inflating the number of correct responses on the Word Association test because it contained the precise definition. The entire experiment took about one hour to complete all activities.

Four weeks later, the VKS, Word Association test, and Word Meaning test were administered again as a delayed post-test. Only 72 of the original 87 participants were available to take the delayed post-test. Due to a formatting error on the original Word Association test, one of the items, *girding*, had to be removed from the analysis. When preparing the reverse-order test, the question for *girding* was accidentally omitted and another item copied in its place. For this reason, it was removed from the VKS, Word Association, and Word Meaning test instruments given for the delayed post-test, leaving 15 items on each test. Instructions and procedures for the delayed post-test were identical to the original session, with one small change. Participants were asked to record which reading treatment group they had been in on the instruments. A list of which students had been in which treatment was shown on screen for students who did not remember. Once the data was entered into a spreadsheet using LibreOffice v5.2 software, analysis was completed using SPSS v23.0.

Comprehension of the texts is operationalized as the tally of pausal units partially or completely remembered in the written recall activity (Brantmeier, Strube & Yu, 2014; Johnson, 1970). The determination of pausal units was based on recorded readings of the texts by four native speakers, and any place where two or more paused was considered an end point of a pausal unit. Split-half reliability for the scoring of pausal units had a Spearman-Brown coefficient of .542. A one-way ANOVA was performed to compare results, with treatment type (*Online*, *Gloss*, or *Control*) as the independent variable and number of pausal units recalled as the dependent variable, and an alpha level of $p < .05$. As can be seen in Table 4, all three groups had similar levels of comprehension, suggesting that any differences in vocabulary acquisition are likely to arise from the treatment type.

TABLE 4
Pausal Units Recalled Descriptive Data

Treatment	N	Mean	Std. Deviation	Std. Error of Mean
Total Pausal Units (n = 63)				
Control	29	30.28	9.99	1.86
Gloss	30	31.73	6.94	1.27
Online	31	30.65	6.75	1.21

Scores of the immediate VKS, Word Association test and Word Meaning test are used

to operationalize vocabulary acquisition, while comparison of scores on the immediate and delayed VKS, Word Association and Word Meaning tests are used to operationalize retention. The VKS was scored 0 points for an answer of "I've never seen this word before," 1 point for "I've seen this word before but I don't know the meaning," 2 points for "This word means ____" with a correct answer supplied, and 3 points for "I can use this word in a sentence. _____" with a plausible usage in a sentence provided. For students that answered 3 or 4 and provided an incorrect meaning or usage, the next lower level's point value was applied. The Word Association and Word Meaning tests were scored 0 points for an incorrect or no answer, 1 point for a correct answer. An answer of 1 on the VKS and/or a correct answer on the Word Association test was considered as formal vocabulary knowledge, while a score of 2 or 3 on the VKS and/or a correct answer on the Word Meaning test was considered as semantic vocabulary knowledge. Displaying formal knowledge was considered as partial acquisition while displaying semantic knowledge was considered as acquisition. One-way MANOVA was then used to compare the results of the three tests, with treatment type as the independent variable and scores on the tests as the dependent variables, and an alpha level of $p < .05$.

For the delayed post-test, only 78 participants were available (Control n = 27, Gloss n = 24, Online n = 27). One-way MANOVA was again used to compare the results of the three delayed post-tests, with treatment type as the independent variable and scores on the three tests as the dependent variables, with an alpha level of $p < .05$. Cases were randomly eliminated using SPSS v.23's random case selection so that each treatment group had equal numbers (n = 24).

4. RESULTS

4.1. Effects of Treatment on Vocabulary Acquisition

Three vocabulary tests were used in this experiment, and one-way MANOVA analysis of the results of all three tests performed immediately after the written recall test showed that both the *Online* group and the *Gloss* group had significant differences when compared to the *Control* group, with no significant differences between the *Gloss* and *Online* experimental groups. The multivariate tests of the MANOVA were $F(6,164) = 10.346$, $p < .000$; Wilk's $\Lambda = 0.526$, partial $\eta^2 = .275$. The effect size shows a large effect (Cohen, 1988) for the two treatment groups over the control group. Descriptive data collected in the experiment is shown in Table 5.

TABLE 5

Immediate and Delayed Post-Test Vocabulary Acquisition Descriptive Data

Treatment	N	Mean (SE)	Std. Deviation	N	Mean (SE)	Std. Deviation	% Change
		VKS			Delayed VKS		
Online	29	19.93 (1.20)	6.49	24	11.38 (1.16)	5.69	-42.9%
Gloss	29	20.07 (1.17)	6.21	24	12.25 (1.05)	5.15	-39.0%
Control	29	9.21 (0.83)	4.45	24	9.54 (0.78)	3.81	+3.6%
Total	87	16.40 (0.83)	7.67	72	11.06 (0.59)	5.01	-32.6%
		Word Association Test			Delayed Word Association Test		
Online	29	5.87 (0.63)	3.49	24	4.21 (0.52)	2.55	-28.3%
Gloss	29	6.80 (0.60)	3.28	24	5.04 (0.55)	2.68	-25.9%
Control	29	2.83 (0.39)	2.07	24	2.25 (0.28)	1.36	-20.5%
Total	87	5.30 (0.37)	3.44	72	3.83 (0.30)	2.53	-27.7%
		Word Meaning Test			Delayed Word Meaning Test		
Online	29	6.90 (0.58)	3.22	24	5.38 (0.66)	3.21	-22.0%
Gloss	29	7.40 (0.50)	2.74	24	5.63 (0.56)	2.76	-23.9%
Control	29	3.86 (0.38)	2.05	24	3.29 (0.40)	1.97	-14.8%
Total	87	6.20 (0.33)	3.11	72	4.76 (0.34)	2.86	-23.2%

One-way MANOVA results comparing the delayed post-tests in the right-hand side of Table 5 showed slightly different results. For the delayed VKS, *Online* and *Gloss* declined, while *Control*'s results were similar to the immediate post-test. No significant differences were found between the immediate and delayed VKS tests, $F(2, 75) = 2.12$, $p = .128$. On the delayed Word Association test, both treatment groups' scores declined compared to the immediate post-test while the control group performed at a similar level. However, the results show a significant difference between the treatment groups and the control group, $F(2, 75) = 11.01$, $p = .000$. The effect size (partial eta squared) was 0.227 which is a large effect (Cohen, 1988). Similarly, on the delayed Word Meaning test, both treatment groups' mean scores declined, while the control group's mean score was similar to the immediate post-test, but again there was a significant difference between the treatment groups and the control group, $F(2, 75) = 6.69$, $p = .002$. The effect size (partial eta squared) was 0.151, which is also a large effect of treatment (Cohen, 1988).

4.2. Effects of Textual Familiarity on Vocabulary Acquisition

To compare vocabulary acquisition effects of textual familiarity, an analysis was made of all participants on each of the immediate and delayed post-tests. In all cases, the mean test scores for the unfamiliar passage, *The King of the Blues*, were higher than mean scores for the familiar passage, *Battle of Myeongnyang*. Descriptive data for the familiar text vocabulary can be found in Table 6 and for unfamiliar text vocabulary in Table 7.

TABLE 6

Descriptive Data for the Familiar Text Vocabulary Items

Treatment	N	Mean (SE)	Std. Deviation	N	Mean (SE)	Std. Deviation	% Change
		VKS			Delayed VKS		
Online	29	9.10 (0.62)	3.34	24	5.71 (0.48)	2.35	-37.3%
Gloss	29	8.24 (0.55)	2.98	24	5.92 (0.56)	2.72	-28.2%
Control	29	4.86 (0.42)	2.25	24	5.17 (0.40)	1.97	+6.4%
Total	87	7.40 (0.37)	3.40	72	5.60 (0.28)	2.35	-24.3%
		Word Association Test			Delayed Word Association Test		
Online	29	2.76 (0.28)	1.53	24	1.42 (0.26)	1.25	-48.6%
Gloss	29	3.17 (0.30)	1.61	24	2.08 (0.25)	1.21	-34.4%
Control	29	1.69 (0.24)	1.29	24	1.21 (0.19)	0.93	-28.4%
Total	87	2.54 (0.17)	1.59	72	1.57 (0.14)	1.19	-38.2%
		Word Meaning Test			Delayed Word Meaning Test		
Online	29	3.38 (0.27)	1.47	24	2.58 (0.42)	2.06	-23.7%
Gloss	29	3.24 (0.28)	1.53	24	2.50 (0.31)	1.50	-22.8%
Control	29	2.41 (0.25)	1.32	24	2.08 (0.29)	1.44	-13.7%
Total	87	3.01 (0.16)	1.49	72	2.39 (0.20)	1.68	-20.6%

TABLE 7

Descriptive Data for the Unfamiliar Text Vocabulary Items

Treatment	N	Mean (SE)	Std. Deviation	N	Mean (SE)	Std. Deviation	% Change
		VKS			Delayed VKS		
Online	29	10.83 (0.75)	4.05	24	7.08 (0.73)	3.55	-34.6%
Gloss	29	11.90 (0.77)	4.17	24	7.33 (0.59)	2.90	-38.4%
Control	29	4.34 (0.52)	2.82	24	5.38 (0.46)	2.26	+24.0%
Total	87	9.02 (0.53)	4.98	72	6.60 (0.36)	3.04	-26.8%
		Word Association Test			Delayed Word Association Test		
Online	29	3.38 (0.41)	2.21	24	2.71 (0.38)	1.85	-19.8%
Gloss	29	3.72 (0.37)	2.00	24	2.96 (0.35)	1.73	-20.4%
Control	29	1.14 (0.25)	1.36	24	1.04 (0.20)	1.00	-8.8%
Total	87	2.75 (0.24)	2.20	72	2.24 (0.21)	1.77	-18.5%
		Word Meaning Test			Delayed Word Meaning Test		
Online	29	3.83 (0.37)	1.97	24	3.04 (0.34)	1.68	-20.6%
Gloss	29	4.21 (0.34)	1.82	24	3.12 (0.34)	1.68	-25.9%
Control	29	1.45 (0.23)	1.24	24	1.21 (0.25)	1.22	-16.6%
Total	87	3.16 (0.22)	2.09	72	2.46 (0.21)	1.76	-22.2%

Comparing the results between the familiar and unfamiliar texts' target vocabulary items shown in Tables 6 and 7 above, a within-subject one-way ANOVA of the immediate VKS test showed a significant difference, $F(2, 178) = 7.03$, $p = .009$. Neither the ANOVA of the immediate Word Association test, $F(2, 178) = 0.61$, $p = .436$, nor of the immediate Word Meaning test, $F(2, 178) = 0.44$, $p = .510$, showed a significant difference in vocabulary knowledge between the two texts. On the delayed post-tests, the differences in vocabulary knowledge, as analyzed by a one-way ANOVA of both the VKS, $F(2, 154) = 6.16$,

$p = .000$, and the Word Association test, $F(2, 154) = 6.38$, $p = .013$, were significant. The results of the Word Meaning test, $F(2, 154) = 0.14$, $p = .709$, were not significant. This suggests that unfamiliar texts may be better for gaining partial knowledge of words, but not necessarily better for gaining semantic knowledge of the words. The results of the analysis will be discussed below.

5. DISCUSSION

Overall, the results obtained from this study seem to support the hypothesis that FAODs are helpful to learners for purposes of vocabulary acquisition, resulting in better partial formal and semantic knowledge of vocabulary items. The analysis of subject familiarity in the text shows that unfamiliarity with the text leads to greater vocabulary acquisition, especially when assisted by an FAOD or printed gloss. The results will be analyzed in more detail here.

5.1. Vocabulary Acquisition

Both the *Gloss* and *Online* treatment groups performed significantly better than the *Control* group on all three immediate vocabulary tests, with large effect sizes. In the delayed post-tests, both experimental groups' scores declined which is to be expected, while the *Control* group's scores remained relatively unchanged. The *Control* group's scores decreased on the Word Association and Word Meaning tests by smaller amounts than those of the two experimental groups, and the *Control* group's scores actually rose slightly on the delayed VKS. The experimental groups' scores for the delayed post-tests were still significantly higher than the *Control* group's scores, despite the decline. There were no significant differences between the *Online* and *Gloss* groups, suggesting that the FAOD was comparable in effect to a printed gloss on the page. These results suggest that using a FAOD while reading does in fact assist with vocabulary acquisition, in line with most previous studies of the effects of dictionaries on vocabulary acquisition (Abraham, 2008; Fageeh, 2014; Kilickaya & Krajka, 2010; Knight, 1994; Ko, 2012; Luppescu & Day, 1993), suggesting that FAODs are useful tools for students.

5.2. Topical Familiarity

The participants of this study achieved better vocabulary acquisition in relation to the unfamiliar text, *The King of the Blues*, although the difference was not always statistically significant. It was assumed that the participants would be able to better recall vocabulary

from the more familiar passage *Battle of Myeongnyang*, especially when using an FAOD or printed gloss, but this was not the case. This runs contrary to the findings of Anderson and Herman (1987) as well as the concerns voiced by Erler and Finkbeiner (2007), that a lack of topical familiarity will hinder vocabulary acquisition and lead to problems with comprehension. Nation and Webb's (2011) proposition that bottom-up processing may play a larger role in acquisition of unknown words may help to explain the findings here. If the FAOD or printed gloss allowed the participants to gain the semantic knowledge of the words without the need to guess from context, then the participants may have been able to spend more time focused on the formal aspects of each target vocabulary item in a bottom-up fashion.

Still, the fact that the target vocabulary from the unfamiliar text was acquired better than the target vocabulary of the familiar text in all three groups is not accounted for. As the *Online* and *Gloss* groups had similar results for each text type, it would seem that dictionaries are a factor but dictionary type is not. A possible explanation for the better recall of the unfamiliar passage may be that the novelty of the unfamiliar passage encouraged greater attention to the contents of that passage and to the target vocabulary in particular (Wu & Huberman, 2007). Participants may have relied on their FAOD or gloss to a greater extent when attempting to comprehend unknown vocabulary, as context clues would have been less reliable in this case. Greater attention to the unknown words, and to the translations provided by the FAOD or gloss, may have led to the higher scores on the three vocabulary tests.

6. CONCLUSION

This study provides some evidence to support the claim that FAODs may be good aids to students seeking to learn new vocabulary. In the experiment, FAODs performed as well as a printed gloss for assisting learners in vocabulary gains, both of which performed significantly better than using context clues alone. The vocabulary gains for both the written gloss and FAOD were not especially robust over time, but were still better than the vocabulary gains attained by using context clues alone.

A second line of inquiry in this study looked at the effects of textual familiarity on vocabulary gains. Scores for vocabulary items from the unfamiliar text were consistently higher than the scores for vocabulary items from the familiar text. Scores for both the immediate and delayed VKS test showed significantly higher scores for the vocabulary from the unfamiliar text. The delayed Word Association test also showed a significant difference, but the immediate Word Association test and both Word Meaning tests did not show significant differences. This suggests that unfamiliarity with a text's topic may assist

in partial word knowledge gains such as word form familiarity, but topical familiarity does not necessarily affect the level of semantic knowledge gained about new vocabulary encountered.

This study has three main limitations. First of all, the necessity of using low frequency vocabulary items instead of pseudo-words in the reading passages required finding words that fit the meaning of the passages. This hindered the process, and the unbalanced nature of the word classes among the two reading passages may have skewed the results, as some studies have suggested that nouns may be easier to learn than adjectives or verbs (Rodgers, 1969). Although unlikely, it is possible that some participants may have been familiar with some words, as higher-level students' knowledge of low frequency words becomes less uniform as learners advance (Nation & Waring, 1997). Many participants did not mark all of the target words in the passages as unknown or looked up. However, this may be explained by participants not following instructions exactly as much as it could by assuming some participants were already familiar with some target vocabulary. In future studies of this sort, an option for participants to actively mark words as 'already known' on the tests should be included to determine which explanation is more likely, and a greater effort should be made to use a variety of word classes in each text.

The second limitation regards monitoring the use of the gloss and FAOD and how students interact with the FAOD or gloss. Studies of computer mediated glosses are easily able to check exactly which words are searched, how often they are searched, and time on task. Because this study relied on a paper-based gloss for one experimental group, and the participants' personal smartphones for the other experimental group, these data points could not be tracked in this study. It is possible that having the gloss on the page may have allowed faster readers in the *Gloss* group sufficient time to memorize those words, while the *Online* group members may or may not have looked up words repeatedly. The FAOD offered by Naver.com was also not analyzed closely. The dictionary aggregates entries from other online dictionaries, so there may be variation from entry to entry presented when a word is searched, and the primary entries in each search may not have been the most appropriate sense of the target vocabulary as used in the context of the passages.

The third limitation is that no pretest was conducted to accurately gauge the proficiency levels of the participants. The rating of students as being intermediate level comes from self-reported standardized test scores from a fraction of the participants, and observations of classroom instructors that the participants are of reasonably similar levels of proficiency. Future studies of this nature should seek to address all of these limitations to provide better data on the benefits of FAODs.

In addition to addressing the limitations of this study, future studies may wish to examine other FAODs to see if they perform equally well. Gains to other aspects of vocabulary knowledge besides receptive formal familiarity and semantic knowledge from

the use of FAODs should be tested. It is possible that Korean students' familiarity with smartphones gives them an advantage whereas learners in other contexts in which smartphones are less common may be at a disadvantage. Studies may also wish to analyze the various FAODs for the quality of their entries, ease of use, and other factors that may be relevant to learning. Studies of a computerized reading task where a computerized gloss could be employed and compared to the FAOD may also be warranted, as in the future students are expected to do a larger share of their reading from screens rather than from pages. FAODs should also be analyzed to ascertain how relevant the entries are for building semantic relationships of new vocabulary beyond a simple translation from L2 to L1.

With these results in mind, some pedagogical implications of the data seem obvious. The intermediate level participants in this study were able to easily comprehend the reading passages even without the aid of a gloss or dictionary to look up unknown words. This suggests that learners of sufficiently advanced level do not need to rely on dictionaries or glosses when reading for comprehension. Using a dictionary or gloss, however, does aid in vocabulary acquisition, which will lead to greater gains in language ability compared to simply using context clues. The use of a gloss or FAOD did not especially burden the learners, and the speed of look ups with the smartphone did not noticeably reduce reading times or affect comprehension.

Another important pedagogical implication of this study relates to the nature of FAODs. Learners in high socio-economic contexts, such as the participants in this study, have easy access to the internet through smart devices. This means they will almost always have access to an FAOD when studying. Learners in medium or low socio-economic contexts may not have such easy access to FAODs, but the fact that they are free and provide a wider range of vocabulary than cheap print dictionaries means that when such learners have access to the Internet, they gain access to a range of vocabulary that would only appear in the most expensive academic print dictionaries without needing to pay for the dictionary, which may be a hardship for learners in low socio-economic contexts. FAODs may be a better choice of dictionary for many learners, and educators may wish to take advantage of this no matter what socio-economic or educational context they may be situated in.

While there is always the risk of distractions from social media, messenger apps, and games, smartphones seem to be a convenient and familiar way for learners to access dictionaries when necessary, and can benefit learners in their effort to improve their language ability and vocabulary knowledge. While more research is needed to confirm the findings of this study, it would appear that the benefits of FAODs accessed via smartphone or tablet may outweigh the disadvantages within the language classroom. Teachers may wish to allow their learners to access this convenient tool in the language classroom after

taking appropriate steps to curb non-learning related use of the devices during the class time.

REFERENCES

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.

Bensoussan, M., Sim, D., & Weiss, R. (1984). The effect of dictionary usage on EFL test performance compared with student and teacher attitudes and expectations. *Reading in a Foreign Language*, *2*(2), 262-276. Retrieved on March 20, 2017, from the World Wide Web: https://eric.ed.gov/?id=ED232436

Bowles, M. A. (2004). L2 glossing: To CALL or not to CALL. *Applied Linguistics, 87*(3), 541-552.

Brantmeier, C., Strube, M., & Yu, X. (2014). Scoring recalls for L2 readers of English in China: Pausal or ideal units. *Reading in a Foreign Language*, *26*(1), 114-130.

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

Educational Testing Service. (2015). *Mapping the TOEIC and TOEIC Bridge tests on the Common European Framework of Reference for Languages*. Educational Testing Service. Retrieved on October 1, 2017, from the World Wide Web: https://www.etsglobal.org/Global/Eng/content/download/890/13127/version/9/file/Mapping+the+TOEIC+and+TOEIC+Bridge+Tests+on+the+CEFR+-+MAR338-LR.pdf

Ellis, N. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 122-139). Cambridge: Cambridge University Press.

Erler, L., & Finkbeiner, C. (2007). A review of reading strategies: Focus on the impact of first language. In A. D. Cohen & E. Macaro (Eds.), *Language learner strategies* (pp. 187-206). Oxford: Oxford University Press.

Fageeh, A. I. (2014). Effects of using the online dictionary for etymological analysis on vocabulary development in EFL college students. *Theory and Practice in Language Studies*, *4*(5), 883-890.

Gostudylink (n.d.). *CEFR levels explained*. Gostudylink.net. Retrieved on October 1, 2017, from the World Wide Web: http://gostudylink.net/en/blog/cefr-levels-explained

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, *19*(3), 357-377.

Johnson, R. E. (1970). Recall of prose as a function of the structural importance of the linguistic units. *Journal of Verbal Learning and Verbal Behavior, 9,* 12-20.

Kilickaya, F., & Krajka, J. (2010). Comparative usefulness of online and traditional vocabulary learning. *The Turkish Online Journal of Education Technology*, 9(2), 55-63. Retrieved on March 14, 2017, from the World Wide Web: http://files.eric.ed.gov/fulltext/EJ898003.pdf

Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal*, *78*(3), 285-299.

Ko, M. H. (2012). Glossing and second language vocabulary learning. *TESOL Quarterly*, *46*(1), 56-79.

Laffey, D. (2016). An investigation of English idiom comprehension strategies used by university students in Korea. *Studies in English Education*, *21*(2), 61-84.

Laufer, B., & Hadar, L. (1997). Assessing the effectiveness of monolingual, bilingual, and "bilingualised" dictionaries in the comprehension and production of new words. *The Modern Language Journal*, *81*(2), 189-196.

Laufer, B., & Hill, M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning & Technology*, *3*(2), 58-76.

Lew, R. (2015). Research into the use of online dictionaries. *International Journal of Lexicography*, *28*(2), 232-253.

Loucky, J. P. (2002). Improving access to target vocabulary using computerized bilingual dictionaries. *ReCALL*, *14*(2), 295-314.

Luppescu, S., & Day, R. R. (1993). Reading, dictionaries, and vocabulary learning. *Language Learning*, *43*(2), 263-287.

Milton, J., & Alexiou, T. (2009). Vocabulary size and the Common European Framework of Reference in Languages. In B. Richards, H. Daller, D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition* (pp. 194-211). Palgrave: Macmillan.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, *24*(2), 237-270.

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

Nation, I. S. P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language*, *27*(1), 136-145. Retrieved on March 1, 2017, from the World Wide Web: http://www.nflrc.hawaii.edu/rfl/April2015/discussion/nation.pdf

Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge: Cambridge University Press.

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

Nesi, H. (2014). Dictionary use by English language learners. *Language Teaching*, *47*(1), 38-55.

Nesi, H., & Haill, R. (2002). A study of dictionary use by international students at a British university. *International Journal of Lexicography*, *15*(4), 277-306.

Nyikos, M., & Fan, M. (2007). A review of vocabulary learning strategies: Focus on language proficiency and learner voice. In A. D. Cohen & E. Macaro (Eds.), *Language learner strategies* (pp. 251-273). Oxford: Oxford University Press.

Pulido, D. (2009). How involved are American L2 learners of Spanish in lexical input processing tasks during reading? *Studies in Second Language Acquisition*, *31*(1), 31-58.

Rodgers, T. S. (1969). On measuring vocabulary difficulty: An analysis of item variables in learning Russian-English vocabulary pairs. *International Review of Applied Linguistics in Language Teaching*, 7(4), 327-344.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 199-227). Cambridge: Cambridge University Press.

Seth, M. J. (2002). *Education fever*. Honolulu: University of Hawai'i Press.

Statista (2017). *Number of smartphone users in South Korea from 2015 to 2021 (in millions).* Retrieved on October 1, 2017, from the World Wide Web: https://www.statista.com/statistics/467171/forecast-of-smartphone-users-in-south-korea/

Wesche, M., & Paribakht, T. S. (1993). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, *53*(1), 13-40.

Wu, F., & Huberman, B. A. (2007). Novelty and collective attention. *Proceedings of the National Academy of Sciences*, *104*(45), 17599-17561.

APPENDIX
Reading Passages

Battle of Myeongnyang

After their first sea victory, the Japanese sailed out of Busan Harbor with 333 ships. They were **sanguine** that they could defeat Yi Sun-shin. The Koreans had only 13 ships after their **unpropitious** defeat. None of them were turtle ships.

Yi Sun-shin **inveigled** the Japanese into Myeongnyang. He sent a fast ship near the Japanese base. The Japanese fell into his trap. Ships could only enter safely one by one. The **girding** hills hid the Korean ships. Steel chains laid across it **delimited** the Japanese ships. A thick fog severely lowered **conspicuity** in favor of the Koreans. 13 Korean warships met the Japanese ships one by one. The Japanese could not fight easily against this clever plan. In this way, he defeated a force 25 times larger than his own.

Yi Sun-shin's victory helped Korea win the war. Japanese land forces near the capital were cut off and **enjoined** to pull back. Myeongnyang was Yi Sun-shin's greatest battle. This **lopsided** victory is one of the greatest achievements in naval war.

Topic: Familiar
Source: modified from content on the Wikipedia Yi Sun-shin page
https://en.wikipedia.org/wiki/Yi_Sun-sin

The King of the Blues

Riley "B.B." King defined the blues for a worldwide audience. He had the **approbation** of fans, critics and musicians. He recorded over fifty albums, many considered **nonpareil**. He was born in 1925, in Mississippi. He played on street corners for a **pittance**, in up to four towns a night.

King's big break came in 1948 on Sonny Boy Williamson's radio show. This led to shows at a local **honky tonk**, and later to his own radio show. King needed a catchy radio **sobriquet**. His nickname Beale Street Blues Boy was eventually shortened to "B.B." King.

One night, while King performed at a dance, two men had an **altercation**. They knocked over a stove, setting fire to the hall. King raced outdoors. Realizing he left his guitar inside, he ran back inside the **conflagration** to get it. He heard the cause was a woman named Lucille and wanted to remember not to make the same mistake. Ever since, King gave the **appellation** Lucille to each of his guitars until he passed away in 2015.

Topic: Unfamiliar
Source: modified from content on the B.B. King Official Website http://www.bbking.com/

Applicable levels: Tertiary

Dennis Laffey
Visiting Professor of English
Department of English Language and Literature
College of Humanities and Social Sciences, Pukyong National University
45 Yongso-ro, Nam-gu
Busan 48513, Korea
Email: dplaffey@gmail.com