Re-examining the effects of word writing on vocabulary learning

ABSTRACT

This study investigated the effects of word writing on vocabulary learning by comparing three conditions in which there was (a) limited time to write words, (b) unlimited time to write words, and (c) a non-writing word-picture pairs comparison. Non-native speakers studying English as a second language encountered 8 word-picture pairs in each condition and were administered a test measuring form recall and another measuring receptive knowledge of written form. The results indicated that there was little difference between scores on both tests when time on task was the same. However, when the participants had as much time as they needed to write and learn words they had higher scores on both dependent measures than when learning in the other two conditions. The findings suggest that an ecologically valid approach to word writing may facilitate vocabulary learning.

INTRODUCTION

Writing words is typically part of the process of vocabulary learning. Many language learning activities have a productive component that requires word writing in some form, and many deliberate vocabulary learning activities such as fill-in-the-blanks, crosswords, and sentence production involve writing targeted words in order to demonstrate learning. Moreover, when novel words are introduced in the classroom, teachers often encourage students to write them down. This suggests that word writing is recognized as contributing to learning. However, the results of previous studies investigating the effects of word writing on vocabulary learning have been inconclusive. Thomas and Dieter (1987) found that word writing had a positive effect on learning, while Barcroft (2006) found that it inhibited learning. Although these studies provide useful findings and highlight an area of vocabulary learning that merits

investigation, neither study examined an ecologically valid word writing condition. Nor did they measure learning using a test that was designed to isolate and measure knowledge of written form; the aspect of vocabulary knowledge that students' attention is directed towards in word writing. The present study was designed to clarify earlier findings by comparing three conditions (word writing with limited time, word writing with unlimited time, and a comparison condition involving viewing word-picture pairs) using tests designed to measure the aspect of knowledge that attention is directed towards in word writing (written form) and learning from word-picture pairs (form-meaning connection).

BACKGROUND

Thomas and Dieter (1987) examined the effects of word writing in three experiments. In the first experiment, all participants encountered L2 French-L1 English word pairs three times. During each encounter, half of the participants were given 10 seconds to write the L2 word twice in a word writing condition while the other participants simply viewed the word pairs in a non-writing condition. A productive recall test was used to measure learning after the treatment. On this test the participants were presented with the L1 part of the word pairs as cues and had to write the L2 words. The results indicated that the word writing condition was superior to the non-writing condition when responses were spelled completely correct or when they were almost correct (one incorrect letter). However, when a different scoring system that was more sensitive to partial knowledge (if the first two letters or last two letters, or any sequence of three letters were correct then the response was scored as correct) was used, there was no difference found between the conditions. The second experiment involved the same methods of learning, but a different test (receptive recall). The results indicated that there was little difference between the two conditions when the L1 word was the required response. In the third experiment, the learning conditions were the same but the dependent

measures changed again. A free recall test of L2 words using the same three scoring systems and a matching test were used. The results of the free recall test showed that word writing led to higher scores when scoring required correct spelling, and correct spelling except for one letter. There was no difference between the two conditions using the most sensitive measure of scoring, and on a matching test. Thomas and Dieter conclude that the benefit of copying the words may be more complete orthographic knowledge due to the increased focus on written form.

Thomas and Dieter's (1987) finding of a positive effect for word writing is supported by studies comparing receptive and productive paired-associate learning conditions and transfer-appropriate processing theory. Research has shown that productive learning is likely to lead to greater productive knowledge than receptive learning (Griffin & Harley, 1996; Mondria & Wiersma, 2004; Stoddard, 1929; Waring, 1997; S. Webb, 2009a, 2009b). One reason why productive learning conditions such as word writing and paired-associate learning (in the productive direction) may facilitate learning on productive tests is that they direct attention towards written form, whereas attention is directed towards meaning in receptive learning conditions. Transfer-appropriate processing theory (TAP) also provides an explanation for these findings. TAP suggests that the similarity between learning and testing conditions is likely to have a positive effect on test performance (Morris, Bransford, & Franks, 1977). Therefore, in Thomas and Dieter's study, it might be expected that word writing would lead to superior learning on the productive tests (productive recall and L2 free recall), but not the receptive tests (receptive recall and matching) because word writing is a productive activity and the word pairs condition was a receptive activity.

Barcroft (2006) examined the effects of word writing on vocabulary learning in two experiments. In the first experiment, 20 native speakers of English learning Spanish as a second language at university attempted to learn 24 L2 Spanish target words in two learning

conditions. The target words were all low frequency concrete nouns such as *rastrillo* (rake), *embudo* (funnel), and *destornillador* (screwdriver) that were unlikely to be known to the participants. The experiment used a within-participants pretest/posttest design. Prior to the treatment, a receptive recall pretest was administered. The participants were presented with L2 target words and had to correctly translate them into English. None of the participants were able to translate any of the target words on the pretest. In a non-writing learning condition, the participants encountered 12 of the 24 target words together with their meanings conveyed by pictures. In a word-writing condition, they encountered the other 12 target word-picture pairs but were required to write these target words in an appropriate space on a sheet of paper. The 24 word-picture pairs were each presented twice for 6 seconds per trial in each learning condition. The order of the conditions and the sets of target words were counterbalanced so that the only difference between the learning conditions was the need for participants to write the target words. Immediately following the treatment, a productive recall posttest was completed; the participants were presented with the 24 pictures from the treatment as cues and had to write the corresponding target words. A delayed posttest 2 days later used the same format as the immediate posttest. A scoring protocol that was sensitive to partial knowledge of the written form of the target words was used. Responses on the posttest were scored according to the number of letters that were present in the response and whether or not they were in the correct position within the response. The closer each response was to the correct spelling of the target item, the more points they were awarded for that answer. The results showed that the word-picture pairs condition was superior to the word writing condition at both retention intervals (immediate, 2 days later).

The second experiment that was carried out had two slight changes to the original design. First, the words that had been previously written were covered to ensure that they did not influence learning, and the order of the items in the testing phase was changed. The

results of Experiment 2 supported the earlier findings; the non-writing condition led to higher scores than the word-writing condition on the posttest at both retention intervals. Barcroft (2006) suggests that the difference between the findings in these experiments and those in Thomas and Dieter's (1987) study may have in part been due to the additional time (10 seconds rather than 6) available in the earlier study. He suggests that with less time there may be insufficient processing resources to encode the written forms and link form to meaning.

Barcroft's (2006) findings were supported by an earlier study he carried out comparing writing sentences and learning with word-picture pairs (Barcroft, 2004). A similar within-participants design that involved learning the same 24 target word-picture pairs was used. However, rather than simply writing the words for half of the items, the participants had to write these words in sentences. In the first experiment, the participants viewed 12 word-picture pairs four times for 6 seconds each in a non-writing condition. In a sentence writing condition, the participants were presented with the other 12 word-picture pairs and had 48 seconds to write each target word in a L2 Spanish sentence. In the second experiment, the learning conditions were the same. However, the number of repetitions and the time on task were controlled. In each condition, the word-picture pairs were presented once for 24 seconds. A productive recall test was used to measure learning and responses were scored twice; once for correct spelling and once for partially correct spelling. In the first experiment, the non-writing condition was superior to the sentence writing condition using both scoring methods. In the second experiment, the non-writing condition contributed to higher scores when scoring for partially correct spelling. The findings were surprising as they contrast the findings of research on paired-associate learning indicating that productive learning leads to higher scores on productive tests than receptive learning (Griffin & Harley, 1996; Mondria & Wiersma, 2004; Stoddard, 1929; Waring, 1997; S. Webb, 2009a, 2009b), as well as research that has found generation to have a positive effect on vocabulary learning (Joe, 1995, 1998).

Moreover, it also contradicts Laufer and Hulstijn's (2001) Involvement Load Hypothesis and Nation and Webb's (2011) Technique Feature Analysis, which were designed to evaluate the effectiveness of vocabulary learning conditions.

One limitation of Thomas and Dieter (1987) and Barcroft's (2004, 2006) studies is that the productive recall test used to assess learning may not have provided a valid measurement of the learning contributions of writing. One purpose of writing is to improve knowledge of spelling by directing attention to written form. The productive recall test format requires participants to write the L2 target words. However, it is the ability to correctly write the target words in relation to their meanings that determines correct responses; participants may be able to write every target word correctly but if they are unable to link form to meaning and write appropriate words beside cues on the recall test, then they will be unable to score correctly. The productive recall format may be an effective measurement of the word pairs conditions because its purpose is to develop knowledge of form-meaning connection. It could be argued that the free recall test used in Thomas and Dieter's third experiment is a valid measure of written form. However, this test is primarily a measure of retention; only if learners can remember items will they be able to demonstrate their knowledge of their form. This test format may be biased towards the word pairs condition because it has repeatedly been found to be an effective method of both learning and retaining knowledge of words (Bahrick, 1984; Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Bahrick & Phelps, 1987; Thorndike, 1908; W. B. Webb, 1962). For example, Bahrick (1984) found that vocabulary learned in word pairs was still recallable after 25 years. Although word writing focuses attention on written form, the extent to which it facilitates retention of written form is not clear. For this reason, it may be more effective to use a test of written form that measures learning but is less dependent on retention. The approach taken in the present study was to

use two tests to measure the learning contributions of each condition; one that measured knowledge of form-meaning connection, and one that measured knowledge of written form.

The difference between Thomas and Dieter's (1987) and Barcroft's (2006) findings may also be explained by the focus of attention in the word writing condition in relation to time on task. When time on task was 6 seconds in the word writing condition (Barcroft, 2006), attention was directed towards the written forms of the L2 words and there was little time to focus on the word-picture pairs and link form to meaning. This led to lower scores on the test of form-meaning connection. When time on task was more than was necessary to write the L2 words (Thomas & Dieter, 1987), there may have been sufficient time to direct attention to both the written forms of items and link form to meaning. In this case, word writing had a facilitatory effect on form-meaning connection. One question that was not explored in either study was how an ecologically valid approach to word writing that involved both word writing and an unlimited time to focus on linking L2 form to L1 meaning would affect learning. Thomas and Dieter's findings suggest that it may have a positive effect on learning. This is perhaps the most important question to investigate because of its relevance to teaching and learning.

THE PRESENT STUDY

The present study was based on Barcroft's (2006) research. The same two learning conditions (word writing, word-picture pairs) were compared and the same productive recall test of form-meaning connection was used as one measure of learning. However, the design also differed from the earlier study in five ways. First, a third learning condition was added (word writing with unlimited time) to determine how a more natural word writing process may affect learning. Second, a test measuring knowledge of written form that did not require learners to demonstrate knowledge of form-meaning connection was also administered to

participants because word writing directs attention to written form and the productive recall test does not provide a valid measurement of this aspect of knowledge. Third, knowledge of form-meaning connection was measured at two levels of sensitivity rather than one because Thomas and Dieter (1987) found that the scoring method affected the difference between the learning conditions. Fourth, pseudowords were used rather than real word items to ensure that the participants did not know the target word forms, and to eliminate the need for a pretest from which they might gain knowledge of the written form of the target words. Fifth, the participants were not told that they would write the words in the tests to avoid any influence that this may have on how the participants approached the learning conditions.

The present study aimed to gain insight into the effects of word writing on vocabulary learning. Specifically, this study was designed to compare the effects of word writing with limited time, word writing with unlimited time, and learning from word-picture pairs on two aspects of vocabulary knowledge: productive knowledge of form-meaning connection and receptive knowledge of written form.

Research questions

The present study seeks to address the following research questions:

1. Does writing a word have a negative effect on learning its form-meaning connection when time on task is limited?
2. Does writing a word have a negative effect on learning its form-meaning connection when time on task is not controlled?
3. Does writing a word have a negative effect on learning its written form when time on task is limited?
4. Does writing a word have a negative effect on learning its written form when time on task is not controlled?

METHODOLOGY

Participants

The participants were 38 non-native speakers of English enrolled in an English proficiency program at a university in New Zealand (10 males and 28 females, ranging in age from 17 to 45, [*M* = 24.82, *SD* = 6.13]). The participants were classified within the program as being at an intermediate to upper intermediate level. Fourteen different native languages were represented, with the largest being Chinese (11), Korean (8), Japanese (4), and Thai (4).

Target words

Twenty-four target words were selected for the study. All of the target words were concrete nouns and their meanings could be easily conveyed in pictures. The target words were all low frequency words sampled from the 5th to 14th 1000-word levels of Nation's (2006) British National Corpus lists. Three English proficiency program teachers who were familiar with the English language level of the participants were shown a list of potential target words and indicated that the participants were unlikely to know the L2 forms of any of the target words selected for the study. Pictures that clearly conveyed the meanings of the 24 target words were then selected for the experiment. The pictures were all of a similar size (*M* = 222px in height and *M* = 196px in width) and were absent of any extraneous detail that might convey non-intended meanings.

To ensure that the participants did not know any of the target words and avoid a potential learning effect from administering a pretest that included the L2 written forms of the items, the target words were replaced with pseudoword forms. CELEX (Baayen, Piepenbrock, & Van Rijn, 1993) was used to create pseudowords that had similar orthographic features to the target words. The criteria for creating the pseudowords included matching the target words for the number of letters, English orthographic neighbours, and

English bigram frequency. The 24 pseudowords were then separated into three blocks of eight. To ensure that the composition of the blocks did not affect learning, the items in the blocks had a similar number of letters (53–54) and syllables (17–19), and they were counterbalanced between the conditions. Table 1 lists the target words and the pseudowords they were matched with in each of the three blocks.

Insert Table 1 about here

Design

The experiment used a within-participants design. All participants attempted to learn the same 24 pseudoword-picture pairs in three conditions: no writing, writing, and writing with unlimited time. In each condition, the participants attempted to learn a block of 8 pseudoword-picture pairs. The order of the conditions and the block of pseudoword-picture pairs presented in each condition were counterbalanced. Each condition was presented a similar number of times in the first, second, and third position in the learning sequence. Each block of word-picture pairs was learned a similar number of times in each condition at each point in the sequence. The word-picture pairs in each block were also randomized between participants.

The experiment was run in E-Prime (Schneider, Eschman, & Zuccolotto, 2002a, 2002b) on a Windows personal computer over two sessions that were 2 days apart. The first session lasted approximately 30 minutes and the second session lasted no longer than 10 minutes. Participants completed the experiment individually in an office with the same researcher present at all times. All participants were informed that the purpose of the research was to compare different approaches to learning. They were also told that they would be tested on what they learned, and they received a voucher to compensate them for their time.

However, in contrast to Barcroft (2006), the participants were not told that they would have to write the target words in the tests as this may have affected how they attempted to learn the words.

At the start of each learning condition the participants were asked to read the instructions presented on the computer screen in English. The participants had as much time as they needed to read the instructions and could ask the research assistant questions if necessary. In each learning condition, the pseudoword-picture pairs appeared on the screen twice. The pseudoword always appeared below the picture. When the learning condition began, a fixation cross appeared on the screen for 200ms followed by a pseudoword-picture pair. In the writing and non-writing conditions, each item appeared on the screen for 6 seconds. Piloting indicated that 6 seconds was enough time to write these items and that there would be little time remaining to view the pseudoword-picture pairs. Research has also indicated that 6 seconds per encounter is sufficient to learn word pairs (Hays, Kornell, & Bjork, 2010). In the writing with unlimited time condition, the participants were given as much time as they needed to attempt to learn the items. Subsequent items were presented 200ms after the participants pressed a button to end the previous item. A fixation cross appeared before each subsequent word-picture pair to signal the next item.

In the non-writing condition, the participants were instructed to try to learn the pseudoword-picture pairs simply by viewing them as they appeared on the screen. They were instructed not to say or write the words. In the writing condition, the participants were required to write the pseudowords from the pseudoword-picture pairs as they appeared on the screen in the appropriate blank on a sheet. The writing with unlimited time condition was the same as the writing condition except that the participants had as much time as they needed to write the pseudowords. The average raw time per item in this condition was 19.7 seconds (*SD*=10.82).

Dependent measures

Two tests were administered to participants at two retention intervals: immediately after completing the three learning conditions and 2 days after the treatment. One test measured productive knowledge of form-meaning connection and the other measured receptive knowledge of written form. The order of these tests was counterbalanced so that half of the participants took the test measuring knowledge of written form first and the test of form-meaning connection second, and the rest of the participants took the tests in the reverse order. Two days later each participant was administered the tests in the opposite order to which they took the immediate posttests.

The test of form-meaning connection was the same format as the one used in Thomas and Dieter's (1987) first experiment and in both experiments in Barcroft's (2006) study. The participants were presented with the pictures from the pseudoword-picture pairs they had viewed in the learning conditions on the computer screen and had to write the corresponding pseudowords in the appropriate blank on a test sheet. The participants had as much time as they needed to write their responses and then pressed a button to see the next picture. Responses were scored in two ways: strict scoring and sensitive scoring. Responses were only scored correct if they were spelled perfectly using strict scoring. Responses that were spelled correctly or misspelled but identifiable as the target pseudowords were scored correct using the sensitive scoring method. For example, the following responses were correct under the sensitive scoring protocol for the pseudoword *hirpome*: *hirpom*, *pome*, *hipomus*, *hirpoem*, and the pseudoword *louber*: *loub*, *luber*, *laber*, and *laubor*. These responses indicated that the participants had gained partial knowledge of the pseudowords. One of the researchers and a research assistant both scored all responses in the sensitive scoring method. The interrater reliability was 97%. Scoring the responses at two levels of sensitivity indicates two levels of knowledge for these word types: partial and full.

The test measuring receptive knowledge of written form used the multiple choice format created for Webb's (2005, 2007) studies of depth of vocabulary knowledge. Participants were asked to circle the correctly spelled pseudowords from five choices: the correct spelling and four distracters. The distracters were created to be similar to the target words both phonetically and orthographically. The following examples are for the pseudowords *hirpome* and *louber*:

1.	a. hirporm	2.	a. loubore
	b. hipome		b. lubore
	c. hirporme		c. loubor
	d. hirpome		d. louber
	e. hiporm		e. luber

RESULTS

The descriptive statistics (means and standard deviations) of vocabulary knowledge scores for the dependent measures are reported in Table 2. The independent variables were learning condition (no writing/ writing/ writing with unlimited time) and retention interval (immediate and delayed post-tests). The outcome variables were the number scored correct (out of 8) on the test of written form, the number scored correct (out of 8) on the test of form-meaning connection with strict scoring, and the number scored correct (out of 8) on the test of form-meaning connection with sensitive scoring.

Insert Table 2 about here

The results of the immediate and delayed posttests were first analysed together because this provides the most robust analysis, and then separately to look at how the results may vary at

different retention intervals. Repeated measures analysis for each outcome variable (three analyses) was used (see Table 3). Both condition and session significantly predicted the number of correct answers on the test of written form, with $p < .001$ for both variables. Post-hoc analysis using the sequential Bonferroni method of adjusting for multiple comparisons showed that writing with unlimited time contributed to a significantly higher mean number of correct responses than either non-writing ($p < .001$, $\Delta = 0.93$) or writing ($p < .001$, $\Delta = 1.46$). Writing and non-writing did not differ from each other ($p = .180$, $\Delta = 0.25$). On the productive test of form-meaning connection with strict scoring (target words must be spelled correctly), both condition and session significantly predicted the number of correct answers ($p < .001$ for both variables). The post hoc Bonferroni analysis showed that writing with unlimited time contributed to a significantly higher mean number of correct responses than either non-writing ($p = .001$, $\Delta = 0.95$) or writing ($p < .001$, $\Delta = 1.24$). Writing and non-writing did not differ from each other ($p = .416$, partial $\eta 2 = .24$). On the productive test of form-meaning connection in the sensitive scoring system, both condition and session significantly predicted the number of correct answers, with $p < .001$ for both variables. The Bonferroni method showed that there was a significantly higher mean number correct for writing with unlimited time than non-writing ($p = .002$, $\Delta = 0.82$) and writing ($p < .001$, $\Delta = 1.16$). Again there was no difference between the non-writing and writing conditions ($p = .113$, $\Delta = 0.37$). Table 4 presents the results of the multiple comparisons.

Insert Tables 3 and 4 about here

A post hoc Bonferroni test was completed to locate the site of significant differences between the learning conditions at the two retention intervals (see Table 5). The analysis revealed that writing with unlimited time was superior to the writing condition on the immediate ($p < .001$,

Δ = 1.20) and delayed ($p < .001$, Δ = 1.17) posttests. Writing with unlimited time also contributed to higher scores than the non-writing condition on the immediate posttest (p = .004, Δ = 0.82) and on the delayed posttest ($p < .001$, Δ = 0.98). Again, no significant difference was found between the writing and non-writing conditions on the immediate (p = .313, Δ = 0.29) and the delayed (p = 1.000, Δ = 0.16) posttests. The results were similar when using the sensitive scoring method. Writing with unlimited time contributed to higher scores than the writing condition on both immediate ($p < .001$, Δ = 1.01) and delayed posttests ($p <$ .001, Δ = 1.27). It was also superior to the non-writing condition immediately after the treatment (p = .011, Δ = 0.66) and 2 days later ($p < .001$, Δ = 0.85). There was little difference between writing and non-writing on the immediate (p = .086, Δ = 0.39) and delayed (p = .313, Δ = 0.32) posttests.

Insert Table 5 about here

DISCUSSION

The experimental design of the present study expanded upon earlier methodologies in two ways. First, by examining the effects of word writing on two aspects of vocabulary knowledge (form-meaning connection and written form), the research provides more accurate results. Second, by comparing word writing conditions that vary in time on task, the research may better indicate the value of word writing in different situations.

In answer to the first research question, the results indicated that although word writing led to lower scores than learning word-picture pairs on the test of form-meaning connection, there was little difference between the two conditions when time on task was controlled. On the immediate posttest of form-meaning connection with strict scoring the mean scores for the writing and non-writing conditions were 1.0 (13%) and 1.4 (18%),

respectively. Two days later the scores on the writing and non-writing conditions were 0.7 (9%) and 0.9 (11%), respectively. When responses were scored for correct and partially correct spelling on the test of form-meaning connection on the immediate posttest, the mean scores were 1.7 (21%) and 2.5 (31%) for the writing and non-writing conditions, respectively. On the delayed posttest the mean scores for the writing and non-writing conditions were 1.6 (20%) and 1.2 (15%).

The results do not provide support for Barcroft's (2006) findings indicating that word writing has a negative effect on vocabulary learning. The difference in findings between this study and Barcroft (2006) might be due to design features that may have reduced the effects of word writing in the earlier study (learning effect from pretest, participants' knowledge of the posttest format prior to the treatment). The results also do not provide support for a facilitatory effect for word writing on knowledge of form-meaning connection in comparison to learning word-picture pairs when time on task is controlled in contrast to Thomas and Dieter's (1987) earlier findings and research indicating that productive learning may lead to superior learning on measures of productive recall than receptive learning (Griffin & Harley, 1996; Mondria & Wiersma, 2004; Stoddard, 1929; Waring, 1997; S. Webb, 2009a, 2009b). The reason for this may be due to the limited time that the participants in this study could direct attention to the aspect of knowledge measured on this test: form-meaning connection. Word writing alone should not be expected to lead to gains in form-meaning connection because it directs attention to written form and diverts attention away from linking form to meaning. Any additional time that is available to focus attention on form-meaning connection may contribute to knowledge of this aspect. However, because the amount of attention focused on form-meaning connection in the two tasks is less in the word writing condition, it is unlikely to be more effective.

Interestingly, when partially correct responses were also scored as correct, the boost in scores was higher for the non-writing condition at both retention intervals. This may indicate that the effectiveness of the non-writing condition may have been limited by smaller gains in knowledge of written form; participants were only able to demonstrate greater knowledge of form-meaning connection when the scoring method was sensitive to partial knowledge. This finding is supported in part by Thomas and Dieter's (1987) study that found that their word pairs condition was as effective as the writing condition when the scoring system was sensitive to partial knowledge of written form but less effective when greater knowledge was required.

In answer to the second research question, word writing with unlimited time had a positive effect on learning form-meaning connection. Scores were significantly higher for writing with unlimited time than for the other two learning conditions regardless of the method of scoring and the retention interval. The mean scores for writing with unlimited time on the posttest of form-meaning connection with strict scoring were 2.7 (34%) immediately after the treatment and 1.9 (24%) 2 days later. When partially correct responses were also scored as correct, the mean scores were 3.6 (45%) and 3.0 (38%) at the two retention intervals. This result suggests that the common process of writing down new words and their meanings in the classroom and focusing on linking form to meaning has a positive effect on vocabulary learning. However, it is important to note that this result cannot be attributed to writing alone. Word writing, directed attention to form-meaning connection through viewing word-picture pairs, and an increased time on task all likely contributed to this result. This finding suggests that when learners write new words and attend to linking form and meaning, there is a positive effect on learning form-meaning connection. However, because the time on task was more than three times as long as the other conditions, it is not clear whether word writing is an efficient method of vocabulary learning.

In answer to the third research question, the results indicated that learning word-picture pairs did not contribute to greater knowledge of written form than word writing when time on task was controlled. The mean scores on the test of written form for the writing and non-writing conditions on the immediate posttest were 4.0 (50%) and 4.3 (54%) respectively, and 3.3 (41%) and 3.6 (45%) for those conditions 2 days later. The relatively large gains in knowledge of written form made over the short learning period (12 seconds total) indicate that both conditions were effective for learning this aspect of knowledge. However, it might have been expected that the productive task of spelling a word would lead to greater knowledge of written form than the receptive task of learning word-picture pairs. Perhaps the reason this did not occur is that the test of written form had a receptive format. Transfer-appropriate processing theory would predict superior performance on the word-picture pairs condition than the word writing condition because receptive learning occurred in the former but not the latter. This result also highlights the benefits of learning with word-picture pairs. The word-picture pairs condition should be considered a weak form of paired-associate learning due to the absence of recall in the task. Barcroft (2007) found that requiring learners to recall target L2 words when cued with pictures conveying their meanings led to greater learning than when word-picture pairs were presented together. In paired associate learning attention is directed towards linking form to meaning. However, the attention placed on L2 form can lead to efficient gains in knowledge of written form as well (S. Webb, 2007, 2009a).

In answer to the fourth research question, word writing with unlimited time contributed to significantly higher scores on the test of written form than for the other two learning conditions at both retention intervals. The mean scores for writing with unlimited time on the immediate posttest were 6.1 (76%) and 4.8 (60%) 2 days later. The relatively large gains for the writing condition with unlimited time indicate that word writing with time to reflect on the shape of the word is an effective method of learning the spelling of words.

Taken as a whole, the results of this study indicate that time on task plays a key role in the effectiveness of word writing. When learners have the time to write words and direct attention to form-meaning connection, word writing is likely to have a positive effect on vocabulary learning. This finding is supported by the superiority of the word writing condition in Thomas and Dieter's (1987) study. In contrast, when there is limited time for word writing and little time to attend to form and meaning, word writing may not be an effective method of vocabulary learning with gains primarily limited to knowledge of written form. This explains the negative effects of word writing in Barcroft's (2006) study. It is important to note that word writing may also have the advantage of directing learners' attention towards the written items during review and promoting further study.

This study also indicated that students are likely to spend much greater time initially trying to learn words than is required to write them. The difference in time on task between the writing (6 seconds) and writing with unlimited time (19.7 seconds) conditions also suggests a lack of ecological validity in the former; although 6 seconds may be sufficient to write words, learners are likely to spend much more time trying to learn these words if the time is available. The higher scores on the tests for the writing with unlimited time condition would suggest that this extra time may be used to focus attention on linking form to meaning, and that the additional focus on form and meaning contributes to greater learning of written form. Although it is intuitively logical that greater time on task will lead to greater learning, this is not always the case in vocabulary learning (Karpicke & Roediger, 2008). Nor was it certain that the participants would spend much more time in the writing with unlimited time condition. However, because of the widespread use of word writing when learning vocabulary and the inconsistent findings, it was important to look at learning in both controlled and ecologically valid situations.

Although one purpose of this study was to improve on the ecological validity of the research, there are two aspects of the design that might be improved. First, in this research and many other studies of deliberate vocabulary learning, a comparison is made between conditions that involve initial learning of the form-meaning connection of target items while completing an activity. The justification for this approach is methodological; it is important to have comparable sets of items and this is most easily done using unknown target words. This occurs in research examining the effects of the keyword technique (e.g., Avila & Sadoski, 1996; Ellis & Beaton, 1993; Levin, McCormick, Miller, Berry, & Pressley, 1982), paired-associate learning (e.g., Prince, 1996; S. Webb, 2007, 2009a, 2009b; W. B. Webb, 1962), sentence writing (Barcroft, 2004; Pressley, Levin, & Miller, 1982; S. Webb, 2005), glossed sentences (Laufer & Shmueli, 1997; S. Webb, 2007), semantic elaboration (Barcroft, 2002; Pressley et al., 1982), structural elaboration (Barcroft, 2002), and learning with pictures (Levin et al., 1982). There is room to argue the degree of ecological validity about these different activities. Perhaps some teachers may provide sets of unknown words for students to study in paired associate learning or the keyword technique rather than teach the words to some extent beforehand. However, the extent to which learners may do these two parts of the activity at the same time in the classroom needs to be questioned. Perhaps the more natural learning process is for students to learn the form-meaning connection of new words first, and then complete activities that involve the recognition or use of these items afterwards to consolidate their knowledge. Thus, the activities might typically be consolidation rather than initial learning activities. One way to improve on the ecological validity of research in studies examining deliberate vocabulary learning would be to examine the effects of variables a short time after initially presenting participants with the form and meaning of target words.

For example, in this study the participants could be presented with one repetition of all 24 word-picture pairs followed by a short distracter task. They might then complete the two

learning conditions. This procedure would still allow the selection of comparable sets of unknown target words, but it would make the learning condition a consolidation activity that is closer to what naturally occurs in the classroom.

A second aspect of the design that might be improved upon is the number of times target words were written. Earlier studies examined the effects of writing words two (Barcroft, 2006) or three times (Thomas & Dieter, 1987). However, these figures may not accurately reflect initial word writing in the classroom, where words are most likely written only once. The consequence of writing words more than once in the research is that the writing process may become easier in subsequent encounters allowing participants to write words more quickly and direct greater attention to viewing the word-picture pairs. This may account for the similarity in average raw scores between the word writing and non-writing conditions on the test of form-meaning connection. It would be useful for future studies investigating the effects of word writing to limit the number of encounters to one to see how this affects learning.

It should also be noted that none of the conditions were very effective in developing knowledge of form-meaning connection. This was due in part to the design of the study. Learning 24 words in two brief encounters is unlikely to lead to large gains in knowledge. Research on paired-associate learning has indicated that increasing the number of encounters (Crothers & Suppes, 1967; Karpicke & Roediger, 2007, 2008; Nakata, 2017), including a recall component (Barcroft, 2007; McNamara & Healy, 1995), and increasing spacing (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Landauer & Bjork, 1978; Nakata, 2015) all have a positive effect on learning. This suggests that it would be possible to increase learning of word-picture pairs relatively easily. However, in the word writing condition learning might be limited without repeated study. Thus, while word writing with unlimited time was most effective in this study, it may not be most effective in relation to other approaches to

vocabulary learning. A useful follow-up to this study would be to compare other methods of learning with word writing with time on task paced by participants in each condition to determine the relative value of word writing. Another question which remains to be answered is how writing both L2 form and L1 meaning or L2 definitions together affects learning. The writing of both form and meaning together may be a more representative approach to writing newly encountered words, and thus deserves investigation.

CONCLUSION

The present study indicated that word writing does contribute to vocabulary learning. It can lead to relatively large gains in knowledge of written form when time is limited, and when there is sufficient time to write words and focus on their forms and meanings, it promotes knowledge of this aspect as well. The research also demonstrated the effect that time on task can have on findings. When time on task was limited to 6 seconds, there was little difference between the writing and non-writing conditions. However, when time on task was controlled by participants, word writing led to significantly greater knowledge on both tests at each retention interval with medium to large effect sizes in every comparison. These findings are supported by earlier research showing that varying time on task may alter results (S. Webb, 2005), and suggest that researchers should carefully consider the degree of similarity between classroom learning conditions and those in experimental interventions. Major changes to learning conditions in research need to be acknowledged or conclusions may be misleading.

REFERENCES

Avila, E., & Sadoski, M. (1996). Exploring new applications of the keyword method to acquire English vocabulary. *Language Learning*, 46(3), 379–395.

https://doi.org/10.1111/j.1467-1770.1996.tb01241.x

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). The CELEX lexical database [CDROM].

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, 113, 1–37. https://doi.org/10.1037/0096-3445.113.1.1

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, 4(5), 316–321. https://doi.org/10.1111/j.1467-9280.1993.tb00571.x

Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 13(2), 344–349.

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, 57(1), 35–56. https://doi.org/10.1111/j.1467-9922.2007.00398.x

Barcroft, J. (2006). Can writing a new word detract from learning it? More negative effects of forced output during vocabulary learning. *Second Language Research*, 22(4), 487–497. https://doi.org/10.1191/0267658306sr276oa

Barcroft, J. (2004). Effects of sentence writing in second language lexical acquisition. *Second Language Research*, 20(4), 303–334. https://doi.org/10.1191/0267658304sr233oa

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, 52(2), 323–363. https://doi.org/10.1111/0023-8333.00186

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

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

Crothers, E., & Suppes, P. (1967). *Experiments in second language learning*. New York, NY: Academic Press.

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

Griffin, G. F., & Harley, T. A. (1996). List learning of second language vocabulary. *Applied Psycholinguistics*, 17, 443–460. https://doi.org/10.1017/S0142716400008195

Hays, M. J., Kornell, N., & Bjork, R. A. (2010). The costs and benefits of providing feedback during learning. *Psychonomic Bulletin & Review*, 17(6), 797–801. https://doi.org/10.3758/PBR.17.6.797

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, 19(3), 357–377. https://doi.org/10.1093/applin/19.3.357

Joe, A. (1995). Text-based tasks and incidental vocabulary learning. *Second Language Research*, 11(2), 149–158. https://doi.org/10.1177/026765839501100206

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. https://doi.org/10.1126/science.1152408

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57(2), 151–162. https://doi.org/10.1016/j.jml.2006.09.004

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.

Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal*, 28(1), 89–108. https://doi.org/10.1177/003368829702800106

Levin, J. R., McCormick, C. B., Miller, G. E., Berry, J. K., & Pressley, M. (1982). Mnemonic versus nonmnemonic vocabulary-learning strategies for children. *American Educational Research Journal*, 19(1), 121–136. https://doi.org/10.3102/00028312019001121

McNamara, D. S., & Healy, A. F. (1995). A generation advantage for multiplication skill training and nonword vocabulary acquisition. In A. F. Healy & J. L. E. Bourne (Eds.), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 132–169). Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781483326887.n5

Mondria, J. -A. & Wiersma, B. (2004). Receptive, productive, and receptive + productive L2 vocabulary learning: what difference does it make? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 79–100). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.10.08mon

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Nakata, T. (2017). Does repeated practice make perfect? The effects of within-session repeated retrieval on second language vocabulary learning. *Studies in Second Language Acquisition*, 39, 653–679. https://doi.org/10.1017/S0272263116000280

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition*, 37, 677–711. https://doi.org/10.1017/S0272263114000825

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

Nation, I. S. P., & Webb, S. (2011). *Researching and Analyzing Vocabulary*. Boston, MA: Heinle.

Pressley, M., Levin, J. R., & Miller, G. E. (1982). The keyword method compared to alternative vocabulary learning strategies. *Contemporary Educational Psychology*, 7, 50–60. https://doi.org/10.1016/0361-476X(82)90007-8

Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *Modern Language Journal*, 80, 478–493. https://doi.org/10.1111/j.1540-4781.1996.tb05468.x

Schneider, W., Eschman, A., & Zuccolotto. (2002a). *E-Prime Reference Guide*. Pittsburgh, PA: Psychology Software Tools Inc.

Schneider, W., Eschman, A., & Zuccolotto. (2002b). *E-Prime User's Guide*. Pittsburgh, PA: Psychology Software Tools.

Stoddard, G. D. (1929). An experiment in verbal learning. *Journal of Educational Psychology*, 20, 452–457. https://doi.org/10.1037/h0073293

Thomas, M. H. & Dieter, J. N. (1987). The positive effects of writing practice on integration of foreign words in memory. *Journal of Educational Psychology*, 79(3), 249–53. https://doi.org/10.1037/0022-0663.79.3.249

Thorndike, E. L. (1908). Memory for paired associates. *Psychological Review*, 15, 122–138. https://doi.org/10.1037/h0073570

Waring, R. (1997). A study of receptive and productive learning from word cards. *Studies in Foreign Languages and Literature* (Notre Dame Seishin University, Okayama), 21(1): 94–114.

Webb, S. (2009a). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal*, 40(3), 360–376. https://doi.org/10.1177/0033688209343854

Webb, S. (2009b). The effects of pre-learning vocabulary on reading comprehension and writing. *Canadian Modern Language Review*, 65, 441–470. https://doi.org/10.3138/cmlr.65.3.441

Webb, S. (2007). Learning word pairs and glossed sentences: The effects of a single sentence on vocabulary knowledge. *Language Teaching Research*, 11, 63–81. https://doi.org/10.1177/1362168806072463

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27, 33–52. https://doi.org/10.1017/S0272263105050023

Webb, W. B. (1962). The effects of prolonged learning on learning. *Journal of Verbal Learning and Verbal Behavior*, 1, 173–182. https://doi.org/10.1016/S0022-5371(62)80026-7