

---

# *The Effect of Type of Written Exercise on L2 Vocabulary Retention*

---

**KEITH S. FOLSE**

*University of Central Florida  
Orlando, Florida, United States*

The present study used a within-subjects design to examine the effect of the type of written exercise on L2 vocabulary retention. Using input for the meaning and usage of the new words from a specially prepared minidictionary, university intensive English program students ( $n = 154$ ) practiced target vocabulary in three types of written exercises conditions: one fill-in-the-blank exercise, three fill-in-the-blank exercises, and one original-sentence-writing exercise. An unannounced posttest using a modified version of the vocabulary knowledge scale tested the meaning of the word (L1 translation or L2 synonym) and usage of the word in a student-written sentence. A repeated measures ANOVA revealed that mean scores for the three exercise types were significantly different from each other, with words practiced under the three fill-in-the-blank exercises condition retained much better than those practiced under either of the other two exercise conditions. The findings suggest the important feature of a given L2 vocabulary exercise is not depth of word processing but number of word retrievals required. This result has implications for language teachers, curriculum designers, and, in particular, materials writers of traditional workbooks and CALL materials.

**E**SL learners soon discover that their lack of vocabulary knowledge impedes their ability to comprehend or express themselves clearly in English. Research highlights the importance of vocabulary knowledge for second language (L2) learners in reading (Haynes & Baker, 1993; Huckin & Bloch, 1993), speaking (Hincks, 2003; Joe, 1998), in listening (Elley, 1989; Ellis, 1994), and writing (Hinkel, 2001; Laufer & Nation, 1995; Lee, 2003; Leki & Carson, 1994; Walters & Wolf, 1996). Schmitt (2000) notes that L2 students need approximately 2,000 words to maintain conversations, 3,000 word families to read authentic texts, and as many as 10,000 words to comprehend challenging academic texts. It is not surprising that learners see English vocabulary acquisition as their greatest language problem (Green & Meara, 1995) and express a strong desire for explicit vocabulary instruction (James, 1996). ESL teachers,

curriculum developers, and materials writers have responded with more vocabulary teaching materials, including even whole vocabulary series (Byrd, Reid, & Schuemann, in press; McCarthy, O'Dell, & Shaw, 1997).

The past decade has also seen a dramatic increase in works on key areas in L2 vocabulary learning, including the learners, the words, and the teacher (Folse, 2004; Nation, 2001; Schmitt, 2000). Research has looked at which vocabulary learning strategies learners use (Kojic-Sabo & Lightbown, 1999; Lessard-Clouston, 1994; Nassaji, 2003; Prince, 1995; Sanaoui, 1995; Schmitt & Schmitt, 1993) and how L2 learners' vocabularies develop (Laufer, 1998; Laufer & Paribakht, 1998; Parry, 1993; Schmitt, 1998). Research has also considered how many words L2 learners need to know (Hulstijn, Hollander, & Greidanus, 1996) and which words learners need to know (Coxhead, 2000; Liu, 2003). Research on the methodology of vocabulary teaching finds that explicit teaching of vocabulary results in better retention than incidental learning from natural text-based input such as book passages or dictionary entries (Chun & Plass, 1996; Hulstijn, Hollander, & Greidanus, 1996; Knight, 1994; Laufer, 1997; Laufer & Shmueli, 1997; Nassaji, 2003; Paribakht & Wesche, 1997; Prince, 1995; Zimmerman, 1997) and that the keyword method, the lone vocabulary teaching method extensively studied, consistently produces L2 vocabulary gains (Brown & Perry, 1991; Jones, 1995).

L2 vocabulary research has looked at the learner, the words, and the method of presentation. Surprisingly little work has been conducted on the next step in learning vocabulary in a formal classroom setting, namely the written practice activities that follow the presentation of any new vocabulary. This study fills this gap by comparing two of the most commonly used types of written vocabulary exercises: filling in the blank (i.e., *cloze*) and original sentence writing. Through empirical results, this study contributes valuable information on vocabulary teaching to L2 learners, teachers, and textbook and curriculum designers.

## **L2 VOCABULARY EXERCISES**

Vocabulary exercises vary greatly according to which aspect of word knowledge is being practiced. Besides a word's meaning, exercises also focus on other aspects of word knowledge, including the word's spelling, part of speech, morphology, meanings in a specific context, connotation, usage, synonyms, antonyms, and collocations. In addition, the directions for a given exercise vary considerably. These two factors affect the amount of time required, degree of difficulty, the type of retrieval (active or passive), and the number of vocabulary retrievals.

## The Value of Written Vocabulary Exercises

Native speakers learn the bulk of L1 vocabulary from natural input, with reading playing a large role. For L2 learners, the role of reading in vocabulary acquisition is important; the actual learning that takes place, however, is slow and unpredictable. According to Paribakht and Wesche (1999),

From the perspective of a language teaching program which aims at developing learners' reading proficiency and related receptive vocabulary, a reading-based, incidental learning approach may be adequate, but for programs which aim at developing learners' production skills, rapid vocabulary expansion and some measure of influence over what is learned, such an approach would appear insufficient. (p. 3)

This inadequacy is evident in empirical studies (Paribakht & Wesche, 1999; Zimmerman, 1997) showing that L2 vocabulary retention is higher for students who complete written vocabulary activities after a reading task than for students who complete another reading assignment (with the same target vocabulary) after the reading task. The question that then follows is, what is the most effective type of written exercise?

### Factors Affecting the Efficacy of Written Exercises for L2 Vocabulary Retention

Two overlapping, important factors in exercise design are attention and noticing, both of which have been extensively debated in L2 acquisition studies (Schmidt, 1990; Truscott, 1998). The design of a certain type of exercise might serve to make a particular L2 word more salient by drawing attention to the word, potentially resulting in the student noticing the word. For example, in the common vocabulary exercise sometimes called odd-man out, learners' attention is drawn to individual vocabulary items as they determine the one word that does not belong in the group: *steal*, *avalanche*, *thief*, *robbery*. Likewise, completing a cloze exercise with target words or writing original sentences draws learners' attention to those words.

In a study of second-semester L2 Spanish learners, Jourdenais, Ota, Stauffer, Boyson, & Doughty (1995) found that a very simple textual modification of learning material resulted in more noticing and use of the target language items. In this study, the researchers used an unenhanced and an enhanced version of a text. In the enhanced version, the target items (i.e., verb forms) were underlined and printed in a different font. Some target items were also put in bold font and others were shadowed.

The fact that the target items looked different from the rest of the text caused them to be noticed. Because learners in this study tended to notice and pay more attention to forms that merely looked different in the text, it can be assumed that a vocabulary exercise that draws learners' attention to the target items would also have similar success.

A third factor assumed important in the efficacy of an exercise is depth of processing ( Craik & Lockhart, 1972; Craik & Tulving, 1975). In this model, "the durability of memory traces is a direct consequence of encoding, with deeper and more elaborate encoding leading to more durable memory traces: the deeper the processing, the better the learning" (Schouten-Van Parreren, 1995, p. 190). In theory, this model favors exercises such as writing original sentences over matching or cloze exercises because writing original sentences supposedly requires a deeper level of processing.

Laufer and Hulstijn (2001), however, point out two problems with this model. First, it is impossible to operationalize *depth of processing*. Second, it is impossible to identify one activity as *deeper* than another. For example, one might assume that writing an original sentence with a word requires deeper processing than answering a multiple-choice question or matching a word to its correct meaning. However, research has not validated this assumption. Although there is general consensus that deeper processing results in better learning, it remains unclear which factors specifically influence depth and therefore cannot be operationalized. Furthermore, no research has produced results that allow one exercise type to be categorized as requiring more depth of knowledge than another.

## **Extant Research on Types of Vocabulary Exercises**

In a study of Dutch EFL learners, Hulstijn (1998) explored the question of whether writing 10 target words is more effective than just encountering them in a reading passage. In Condition 1, students read a letter to the editor featuring the target words. In Condition 2, students read the letter and then did a gap-filling exercise with the target words. In Condition 3, students wrote a letter to the editor with the target words. Students showed the highest retention in Condition 3, original letter writing. Of a maximum score of 10, Condition 1 had an average retention score of 4.3, Condition 2 had 5.9, and Condition 3 had 7.1. Students who did longer original writing retained more vocabulary; their subsequent learning job will be easier because they will then face less unknown material.

Hulstijn (1998) also reported results on a similar study in which students who read a weather report on a computer were able to look up the meaning of any unknown word and obtain an L1 translation. One group only read the passage; the other group read the passage and then had

to write their own weather report. Afterward, tests were created based on which words the students had looked up. The mean for reading-only was 3.8 (of 10); for the reading plus writing, it was 4.6.

In a similarly designed study conducted with advanced EFL learners in both the Netherlands and Israel, Laufer and Hulstijn (1998) found that using the target words resulted in better incidental vocabulary acquisition than receiving comprehensible input containing the same target words. In Condition 1, students read a letter to the editor in which the 10 target words were printed in bold and glossed in L1 in the margin, and then they answered some comprehension questions. Students in Condition 2 did the same task; however, instead of appearing in bold print, the 10 target items were deleted from the passage and students had to fill in the blanks. The 10 target items, along with five words that did not appear in the original text at all (i.e., they served as distractors), were printed with both an L1 translation and an L2 explanation. In Condition 3, students were given a list of the 10 target words with their part of speech, L2 definition, and an example sentence. Students then had to use the words in a letter to the editor regarding a controversial issue. On an unannounced vocabulary test with a maximum score of 10, the average retention scores were 2.25 for Condition 1, 3.40 for Condition 2, and 5.90 for Condition 3. Clearly, Condition 3 produced the best vocabulary retention.

Writing original language with the vocabulary likely produced better vocabulary results because students were pushed to produce output instead of just receiving input (Condition 1) or simply filling in blanks (Condition 2). Although having to produce more output with each word may have been a significant factor in retaining the word, a more likely explanation for the superiority of Condition 3 is that in writing a letter (as opposed to working with separate, unconnected sentences), students often choose a word, write a sentence, and then erase that sentence because they want to use the word in another sentence. Constructing the letter (or paragraph) therefore requires learners to interact with any single word multiple times, a factor that has been shown to play a key role in ultimate vocabulary retention (Atkins & Baddeley, 1998; Atkinson, 1975; Baddeley, 1990; Bahrack, Bahrack, Bahrack, & Bahrack, 1993; Cobb, Spada, & Zahar, 2001).

In addition to type of task, another factor that could have influenced the results is time on task. The time on task in the three conditions was different. Condition 1 took about 40 minutes, Condition 2 took 50 minutes, and Condition 3 took 70 minutes. There is almost a twofold difference in time between Conditions 1 and 3. Thus, we are left with the question whether the exercise type or time on task was a more important factor in the resultant L2 vocabulary acquisition. The purpose of the current study is to determine which of these two factors is more important in L2 vocabulary retention.

## **METHOD**

### **Participants**

Individuals in this study were 154 ESL students attending intensive programs at four U.S. universities. All of these programs are designed for international students who plan tertiary academic study in the United States, so the curricula and the student populations are very similar. Students' proficiency levels ranged from lower intermediate (50) to upper intermediate (51) to advanced level (53). Fourteen different native languages were represented, with the largest being Spanish (58), Arabic (22), Japanese (22), and Korean (11).

### **Research Design**

The focus of the study was to determine whether type of written exercise had a significant effect on L2 vocabulary retention. Three types of written exercises were tested: one fill-in-the-blank sheet, three fill-in-the-blank sheets, and writing original sentences. In this within-subjects design, all 154 participants practiced the same 15 target words. However, the words were divided into three equivalent groups (A, B, C) of five words each, and participants practiced each group of words under one of the three exercise conditions. Thus, each participant practiced all three groups but with a random assignment of practice condition to each group.

### **Three Written Exercise Conditions**

This study examined the efficacy of three written exercise conditions. In Condition 1, students did one completion exercise for five of the words, that is, one group. Students saw six words in a box at the top of the page, followed by five unrelated sentences, each of which had a uniform size blank indicating where one of the five words should be filled in. (The sixth word was provided as a distractor, or filler, to minimize reliance on guessing.) The instructions for this exercise read, "Fill in the blanks with one of these words. Each word should be used just one time."

Condition 2 was the same as Condition 1, but the words in Condition 2 were practiced in not one but three different exercises. Each of the three exercises was similar to the exercise in Condition 1.

In Condition 3, students had to write an original sentence with the given word. The instructions for this exercise read, "Write a sentence with each of these words. The sentence should show the teacher that

you understand the meaning of the word.” There was an example of a correct sentence and an incorrect sentence as well.

Conditions 1 and 3 obviously take different amounts of time to complete because of what these practice tasks require of learners. (A pilot test that I conducted shows that writing sentences takes approximately three times as long as finishing a completion exercise for the same words.) As a result of this pilot study, Condition 2 consisted of three cloze exercises because it was believed that students would take the same amount of time to do Condition 2 as they would for Condition 3. Though comparing Conditions 1 and 3 would address the effectiveness of the two different types of exercises, a comparison of Conditions 2 and 3 would not only address the effectiveness of these completion versus sentence-writing tasks but would also take into account the question of time on task, something that previous studies had not addressed.

## MATERIALS

### Target Words

This study used these 18 words in a within-subjects design:

<i>bolster</i>	<i>plead</i>
<i>bruise</i>	<i>plummet</i>
<i>burst</i>	<i>ponder</i>
<i>fib</i>	<i>squander</i>
<i>forge</i>	<i>startle</i>
<i>fret</i>	<i>swipe</i>
<i>launch</i>	<i>thrive</i>
<i>linger</i>	<i>toil</i>
<i>loathe</i>	<i>trigger</i>

Fifteen of the words are actual target words. The three words beginning with *f*, *fib*, *forge*, *fret*, serve as distractors in the completion exercises to reduce guessing. That is to say, the words that begin with *f* were never correct answers in any of the three conditions and therefore should not have drawn much attention or learning.

An overriding concern in selecting the target words for this study was that they be unknown to the participants. To this end, I perused two issues of *Newsweek* to identify potential target words that even very advanced ESL students might not know. When I saw that a large number of these words were verbs, I then eliminated all words that were not verbs. Most studies deal with 10 to 15 target words of various parts of speech. In this study, however, the part of speech was taken into account because the part of

speech of a word affects its difficulty. Ludwig (1984) notes, "Psychological research shows differential performance on tasks involving nouns, verbs, adjectives, and adverbs, indicating that the form class of a word is a reasonably potent variable in verbal tasks" (p. 554). Although the exact ranking of the parts of speech in terms of difficulty is not clear, Laufer (1990) sums it up best: "It is sometimes argued that certain grammatical categories are more difficult to learn than others. Nouns seem to be the easiest; adverbs—the most difficult; verbs and adjectives—somewhere in between" (p. 298).

Limiting the syntactical category reduces the role of this extraneous variable and increases the chance that any true effect of type of exercise can be observed. I then asked four ESL students (who spoke Spanish, Japanese, Arabic, and Korean) to identify words from the list that might have cognates in their native languages.

Finally, to ensure that the words were unlikely to be known by any of the actual participants, I conducted a pilot test with a group of very advanced ESL students ( $n = 11$ ) whose proficiency level was much higher than that of any of the actual participants. Results of this pilot test revealed that one word was known by one student, and another word was known by two students. Given this low level of knowledge, the words were deemed appropriate for this study.

The resultant list of verbs was alphabetized and divided into three groups of similar difficulty. To ensure this, the following steps were taken: (1) the frequency count from the Brown Corpus (Francis & Kucera, 1982) for each word was considered; (2) all words were the same part of speech, that is, verbs; (3) all words were single words; no phrasal verbs such as *call off* or idioms such as *wreak havoc* were chosen; (4) words that had any kind of affix that might provide a morphological clue to the meaning of the word or serve as an anchor in remembering the word were eliminated (e.g., *retrieve*, *overcome*); (5) as much as possible, only words that I thought were worth learning were used so as to increase the pedagogical value of the study for the participants because they were currently enrolled students; (6) the words were distributed in the three groups so that initial letters were equally represented (i.e., all groups have exactly one word that begins with the letter *b*); (7) the number of syllables per word was matched between groups of words in an attempt to control for word length; (8) as much as possible, words with a grammar were not included (e.g., the word *cling* is followed by the word *to*, the word *indulge* is often followed by a reflexive pronoun) because it was highly unlikely that students would pick up on this small syntactic marker in an incidental learning study when the presentation was focused on meaning; and (9) words that were deemed unique (e.g., the word *plummet* has the letters *mm* together, the word *squander* begins with *squ*) and therefore more memorable or learnable were distributed



equally among the three groups. Based on these nine factors, the words were clustered into groups that were equivalent in difficulty. Group A included *bolster*, *fib*, *launch*, *plead*, *squander*, and *toil*. Group B included *bruise*, *forge*, *linger*, *plummet*, *startle*, and *thrive*. Group C included *burst*, *fret*, *loathe*, *ponder*, *swipe*, and *trigger*.

## Pretest, Posttest

A modified version of the vocabulary knowledge scale (VKS; Paribakht & Wesche, 1997) was used for both the pretest and posttest. This modified version of the VKS (Table 1) includes three levels of word knowledge that could detect even partial gains in degrees of knowledge. Scoring on this modified VKS awarded one point if a correct meaning was demonstrated (as evidenced by an acceptable English synonym, English definition, L1 translation, or L1 definition) and one additional point if a correct example sentence with the word was provided. Thus, each word could receive a score of 0, 1, or 2.

## Minidictionary

Input for the meaning of the words was provided by a mini-dictionary that was created especially for this study. The minidictionary was four pages long and contained pertinent information regarding the meaning of each of the 18 words. These words were arranged alphabetically with ample space between the entries to facilitate easy reading and use.

Each dictionary entry began with the word printed in bold font. This was followed by part of speech (verb in all cases). The definition is the result

**TABLE 1**  
**Modified Vocabulary Knowledge Scale**

<p>1. I don't know what this word means.</p> <p>2. I know this word. It means _____. (provide an English synonym or a translation in your native language)</p> <p>3. I can use this word in a good example sentence. Write your sentence here:</p> <p>_____</p> <p>(If you do #3, you must do #2 also.)</p>
---

*Note.* Adapted from Paribakht & Wesche, 1997.

of consultations with both native speaker and ESL learners' dictionaries. The definition was written in simple language so that all participants in this study, regardless of their proficiency level, could comprehend it. The definition was followed by two simple but illustrative sentences containing the vocabulary word. In both sentences, the word was underlined. Here is the entry for the word *bolster*:

**bolster** (verb) To support; to make something feel strong again.

1. When Joe was in the hospital, I sent a nice card to *bolster* his spirits.
2. The players practiced very hard to *bolster* their chance of winning.

## Practice Booklets

A practice booklet consisted of six pages. The first page offered general directions. Condition 1 was one page, Condition 2 was three pages, and Condition 3 was one page. To control for the effect of any interaction between the individual words and the type of practice, the exercise condition was randomly assigned to the three groups of words. There were six possible combinations of condition (1, 2, 3) and word group (A, B, C) (e.g., 1A, 2B, 3C; 1A, 2C, 3B; 1B, 2A, 3C; 1B, 2C, 3A; 1C, 2A, 3B; 1C, 2B, 3A). To control for the *recency effect* (i.e., the first and last words learned are more easily remembered), the sequence of the three groups of words and the sequence of the exercise types were taken into account. Thus, within any one of these six combinations, there are six possible orderings. For example, 1A, 2B, 3C can also be 1A, 3C, 2B; 2B, 1A, 3C; 2B, 3C, 1A; 3C, 1A, 2B; or 3C, 2B, 1A. Based on these 36 combinations of condition, word group, and ordering, 36 different practice books were created, and the 154 participants were randomly assigned one of the 36 booklets.

Creating the fill-in-the-blank sentences for the target words involved several steps. First, sample sentences were checked in several English dictionaries. A pool of three sentences was written for each word. To avoid bias from one person writing all the example and exercise sentences, original sentences were obtained from 35 members of the materials writers' discussion list TESLMW-L. Finally, the examples were checked to ensure that each of the examples was different from the other two, especially in terms of collocations.

The challenge was to compose sentences that sounded native-like, were at a very low level of ESL proficiency, did not use any new or difficult vocabulary, and were different from the two examples in the student minidictionary. If the examples were too similar to the minidictionary, then students might end up learning a paired association rather than the actual meaning of the word.

For *squander*, the following three sentences were created:

1. From 1994 to 1997, I was not serious about my studies. Now I'm really sorry that I \_\_\_\_\_ those three years of my life!
2. He \_\_\_\_\_ most of his salary on beer and cigarettes. I don't think this is a good thing to do.
3. This is an important opportunity for you. If you \_\_\_\_\_ this opportunity, it will be a mistake for your future.

The three collocations here are *years* (time), *salary* (money), and *opportunity*.

## PROCEDURE

On the first day, participants were told that students at four schools were taking part in a research study comparing how different L1 groups learn English. For the pretest, students completed the VKS for 24 words, which included the 18 target words as well as six easy words.

The next activity, a simple word association, was a filler activity to decrease the chance that students would remember the words that they had just seen on the pretest. Students were given a blank sheet of paper. Students listened to a word and then wrote down their first association. On the second day, students completed another word association activity before the actual treatment phase of the experiment began. For the actual treatment, students were randomly given a minidictionary and one of the 36 versions of the practice booklets.

Because time on task was an issue in this study, students were also asked to keep track of how much time they spent on each exercise. In the upper right-hand corner of each exercise sheet, a space was provided for students to write the time they started, the time they finished, and how many minutes they spent on the exercise. Students were given up to 40 minutes to complete these exercises. Immediately after the practice booklets had been collected, students were given the unannounced posttest, which was exactly the same as the pretest.

Students were allowed thirty minutes to complete the posttest. When time was called, students were instructed to turn their posttest booklets over. They were told to write the answer to this question on the back of the last sheet:

Between the time we finished yesterday and the start of today, did you do anything at all to find out the meanings of any of the words in this test booklet? For example, did you ask a friend? Consult a dictionary? Ask your teacher? If so, tell exactly what you did. Be sure to tell which words you got informa-

tion about. If you did not do anything between yesterday and today to learn the meanings of these words, then write 'I didn't do anything to learn these words' on this sheet.

## Scoring of the Pretests and Posttests

The pretest and the posttest were the same. The 18 target words along with six filler words made up the 24 words on the test. Since there was no research interest in the six filler words, they were not scored in any way.

The modified VKS used in this study awarded a score of 0, 1, or 2 for each target word. The first part of the VKS asked the learner to give an English synonym or an L1 translation, which was worth one point. The second part asked the student to write a good example sentence with the word, which was worth an additional point. Participants understood that the only way to have their sentence considered for scoring was if they first wrote an English synonym or L1 translation.

The data were scored two times, once using a strict interpretation (S) of the criteria and again using a lenient interpretation (L) of the data. Because this was an incidental learning situation, any learning that might show up was not expected to be very great or very deep. Since writing an appropriate sentence requires the use of a correct collocation with the verb, a rather fine point, there was some concern that using only S scoring would fail to capture smaller increments of learning that had taken place. For example, if a student says only that *toil* means *work* (instead of the correct answer of *work hard* or *work with great effort*), it is obvious that some learning has taken place. To award this response a 0 ignores the fact that learning a word often happens in increments; it is rarely an all-or-nothing proposition. This simple definition would receive a score of 0 in S scoring but a score of 1 in L scoring.

## Student Interviews

To ascertain whether any additional variables played a role in vocabulary acquisition, follow-up interviews were conducted with eight students. These eight students were chosen based on their level of proficiency, their L1, and their scores in this study. Of the eight students, three were in the top third of posttest scores, three were in the middle third, and two were in the lower third. Three were Arabic speakers, three were Spanish speakers, one was Japanese, and one was Chinese.

## RESULTS

### Effect of Exercise Type

The data regarding the effect of exercise condition were analyzed using a repeated measures analysis of variance (ANOVA). Each participant in this within-subjects design reported three scores that were not totally independent of each other; repeated measures ANOVA reduces error by taking this into account. The alpha level for all analyses was set at .05 for tests of significance. Descriptive statistics for the lenient (L) scoring system, which had a 0 to 10 point scale (i.e., two points for each of the five words), are presented in Table 2.

To test for a statistically significant difference between these means, a one-way repeated measures ANOVA was conducted. The results are shown in Table 3. The main effect of exercise type was statistically significant,  $F(2, 306) = 87.01, p < .0001$ . (The  $p$ -value used throughout is the more conservative Greenhouse-Geisser adjustment to correct for possible violations of the sphericity assumption.)

Since the repeated measures ANOVA revealed a significant difference between the means, with that of Condition 2 higher than that of Condition 1 and that of Condition 3, an ANOVA of contrast variables was conducted. As shown in Table 4, there was a significant difference between Conditions 1 and 2 as well as between Conditions 2 and 3. However, there was no significant difference between Conditions 1 and 3.

The effect sizes of the pairwise comparisons were calculated using Cohen's  $d$  (Cohen, 1992). *Effect size* is defined as the extent to which an

TABLE 2  
Analysis of Variance for Effect of Exercise Type on Retention (Lenient Scoring)

Source	df	Type III SS	Mean Square	F Value
Condition	2	641.66	320.83	87.01 <sup>†</sup>
Error	306	1128.34	3.69	

Note. <sup>†</sup> $p < .0001$ , based on Greenhouse-Geisser correction.

TABLE 3  
Descriptive Statistics for Retention (Lenient Scoring) by Exercise Type

Condition	$n$	Mean	Standard Deviation
1 (one completion)	154	2.18	2.36
2 (three completions)	154	4.78	2.78
3 (original sentences)	154	2.39	2.48

**TABLE 4**  
**Analysis of Variance for Effect of Exercise Type on Retention (Lenient Scoring)**

Contrast	df	Type III SS	Mean Square	F Value	Effect Size
Conditions 1 and 2	1	1038.96	1038.96	142.05 <sup>†</sup>	1.01
Conditions 1 and 3	1	6.65	6.65	0.96	.09
Conditions 2 and 3	1	879.38	879.38	111.69 <sup>†</sup>	.91
Error	153	1204.62	7.87		

Note. <sup>†</sup> $p < .0001$ .

experimental manipulation (here, exercise type) has an effect of separating two groups. A standard guide for interpreting effect sizes is that an effect size of .2 is considered *small*, an effect size of .5 is considered *medium*, and an effect size of .8 is considered *large*. As seen in Table 3, the effect sizes for both pairwise comparisons involving Condition 2 were very large.

Data analysis using strict scoring produced very similar results. All results that were statistically significant with lenient score were also statistically significant with strict scoring.

## Time on Task

Although it is true that participants learned more vocabulary doing three fill-in-the-blank exercises than one sentence writing exercise, it is also true that doing the three completion exercises took about 50% more time. To control for time on task, a post hoc analysis was conducted on a subset of 31 students who had equivalent time on task (i.e., a difference of one minute or less) for Conditions 2 and 3. With the lenient scoring system, the mean score for Condition 2 was 5.45, while the mean score for Condition 3 was 2.65; the difference was more than double. The main effect of exercise type was statistically significant,  $F(2, 60) = 16.59$ ,  $p < .0001$ . Results using the strict scoring system were also statistically significant.

## Discussion

Many educators see fill-in-the-blank exercises as a superficial or passive use of the vocabulary, especially when compared to writing original sentences. Hulstijn and Laufer (2001) would say that fill-in-the-blank exercises induce less involvement than writing original sentences as measured

by their involvement index, which rates activities as *high* (2), *medium* (1), or *weak* (0) on three essential components of learner involvement, namely need, search, and evaluation. With this index, fill-in-the-blank exercises are rated as *high need* (2), *medium search* (1), and *medium evaluation* (1), for a score of 4, while sentence-writing exercises are *high need* (2), *medium search* (1), and *high evaluation* (2), for a score of 5. Under this index, a learner-generated context garners a higher evaluation rating than teacher- or textbook-generated sentences in fill-in-the-blank exercises.

When a learner encounters a blank in a sentence in a vocabulary exercise, however, who can say that the learner's process in trying out the various words in this slot, perhaps by translating many of the words or perhaps by remembering tidbits about some of the words (e.g., "*Forge* is something bad, so it can't fit in this sentence because this sentence is about something good"), is not indeed deep processing of or high involvement with the word? A fill-in-the-blank exercise is not only deep (processing) but also highly efficient in terms of student and teacher time required. (See Hulstijn, 2001, and Hulstijn & Laufer, 2001, for a cogent discussion of the difficulty in operationalizing *depth of processing*, much less deciding whether a particular type of practice is *deeper* than another.)

Educators have commonly assumed that students retain new vocabulary better by writing original sentences than by doing a superficial completion exercise. In this study, however, the difference between the mean retention score for writing sentences (2.39) and that of doing one completion exercise (2.18) was not statistically significant (see Table 2). Therefore, we cannot say that one of these exercises is better than the other for L2 vocabulary learning.

A more important factor in the efficacy of an exercise type than depth of processing is multiple retrievals of the target word. In this study, the mean score for students who did three completion exercises (4.78) was double that for students who wrote original sentences (2.39) and more than double that for those who did one completion exercise (2.18). The difference in these means was significant at the .0001 level (see Table 3). Furthermore, the effect size between the means for comparisons involving practice Condition 2, three completion exercises, was very large.

The results of this study show the value of a vocabulary exercise that requires multiple encounters with or retrievals of the target words. These results are in line with the psycholinguistic and educational psychology research on *rehearsal* (Baddeley, 1990) and *distributed practice* (Atkins & Baddeley, 1998). The current study indicates that doing multiple target word retrievals in an exercise, no matter how superficial the exercise may seem, is a stronger and more facilitative factor in L2 vocabulary learning than the purported deeper processing or involvement load that writing original sentences with new L2 vocabulary may offer.

## Classroom Implications

When the goal is vocabulary learning and time is an important factor, assigning ESL students the task of writing original sentences is not as efficient or pedagogically sound an activity as completing multiple fill-in-the-blank tasks. Writing original sentences involves a tremendous amount of student time in looking up the meaning of the word in a dictionary, deciding if the meaning makes sense, deciding if the word can be used in a particular way in English, coming up with a good sentence with an appropriate collocation, and then deciding if the syntax of the created sentence is correct. In addition to this student time, teacher time required in marking these sentences is overwhelming. The teacher's dubious challenge is how to mark this paper so that the student can benefit from the feedback and learn the word. Unfortunately, student original sentences with new vocabulary often resemble a word heap in which sentences have missing words, misplaced words, or incorrect (though seemingly logical) collocations. These problems, especially those involving collocation and usage, are extremely difficult to correct or reformulate in such a way that the student's original meaning and tone are maintained. Throughout this correction process, teachers must balance how much time they can realistically devote to this task. Any benefit for the student not only presupposes that the teacher will give appropriate, comprehensible, and useful feedback, but also that the student will actually read, understand, and use this feedback. Writing original sentences with vocabulary words is time-consuming for the student and the teacher.

If this invested time actually produced solid retention results, the students' and the teacher's time and effort might be worth it. However, the results of this study show that writing original sentences is neither an effective nor efficient written exercise for students to do when the goal is L2 vocabulary growth and retention.

In contrast to the time required for a sentence-writing exercise, teachers can develop a basic completion exercise relatively quickly and easily. The teacher creates a sentence for each target word, deletes the target word answer from the sentence and substitutes a blank for the word, and then lists the answers (plus a distractor or two) at the top of the page. Teacher preparation time is minimal, student time is less than with writing sentences, and teacher correction can be done quickly. Perhaps more importantly, students will always end up with a correct English example sentence to study.

The current study looked at the effectiveness of sentence writing for L2 *vocabulary learning*, and these results do not support sentence-writing exercises for L2 lexical growth when time on task is controlled. However, this is not to say that writing sentences is not effective for language goals



such as writing improvement, syntax development, spelling, or overall fluency.

This study provides empirical support for L2 vocabulary CALL (computer-assisted language learning) software programs, many of which are designed in a manner that requires multiple practice exposures to target items. A common criticism of CALL software programs has been that they do not make use of so-called deeper processing activities with L2 vocabulary, such as writing original sentences or creating personalized examples with the target items. This lack of deep processing was seen as a natural limitation of the computer, which cannot accurately interpret original, individually written input at the sentence level and provide meaningful feedback. In addition, some educators were concerned that asking students to do superficial types of exercises such as filling in the blanks (or multiple choice, true-false, or matching) would limit their level of knowledge about a word and students would not be able to actually use the word. However, the test in this study required both passive knowledge (i.e., L1 translation or L2 synonym) and active usage (i.e., an original student-generated sentence) of a word.

Results from the current study show that writing original sentences, one of the supposedly deeper processing activities that the computer could not facilitate, was only about half as effective as doing three written practice encounters with target items. Multiple encounters using fill-in-the-blank activities is a task that not only can be done extremely efficiently by the computer but also produces superior retention results. Therefore, it behooves L2 vocabulary software designers to ensure that multiple encounters with the target items is an integral part of their learning software; likewise, educators should look for this feature in software that they might purchase for their learners.

## Directions for Future Research

This study found that multiple sentence-completion exercises produce better vocabulary retention than writing original sentences when time on task is controlled. In this study, only verbs were used. Future studies could investigate whether this facilitative effect holds for other parts of speech, especially adjectives and nouns, and for multiword units such as phrasal verbs (*run out of, take after*) and idioms (*a piece of cake, bite off more than you can chew*).

Future studies could use more measures to see the treatments' different effects. The current study used the VKS to measure breadth and depth of word knowledge, and this knowledge was measured only once immediately after the treatment. Future research could use multiple

instruments to ascertain vocabulary learning over time (e.g., short-term vs. long-term retention), as done by Waring and Takaki (2003).

Finally, many different kinds of written exercises exist, including completion, matching, true-false, odd-man out, error identification, short answer, sequencing, writing original sentences or examples, and multiple choice. The current study examined only two of these, namely, completion (fill in the blank) and writing original sentences. These two exercise types were chosen because they are two very commonly used exercises in L2 classrooms. Future research should look at different combinations of exercise types. Only in this way can there be empirical support for any sort of hierarchy of exercise types.

## THE AUTHOR

Keith Folse is an associate professor of TESOL at the University of Central Florida. His publications include *Vocabulary Myths* (2004), *The Art of Teaching Speaking* (2006), and *Greater Essays* (2006).

## REFERENCES

- Atkins, P., & Baddeley, A. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics*, 19(4), 537–552.
- Atkinson, R. (1975). Mnemonics in second-language acquisition. *American Psychologist*, 30, 821–828.
- Baddeley, A. (1990). *Human memory*. Hillsdale, NJ: Lawrence Erlbaum.
- Bahrnick, H., Bahrnick, L., Bahrnick, A., & Bahrnick P. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, 4, 316–321.
- Brown, T., & Perry, F. (1991). A comparison of three learning strategies for ESL vocabulary acquisition. *TESOL Quarterly*, 25, 655–670.
- Byrd, P., Reid, J., & Schuemann, C. (Eds.). (in press). *English for academic success: Vocabulary*. Boston: Houghton Mifflin.
- Chun, D., & Plass, J. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, 80, 183–199.
- Cobb, T., Spada, N., & Zahar, R. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review*, 57(4), 31.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.
- Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, 11, 671–684.
- Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology*, 104(3), 268–294.
- Elley, W. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, 24(2), 174–187.
- Ellis, R. (1994). Factors in the incidental acquisition of second language vocabulary from oral input: A review essay. *Applied Language Learning*, 5(1), 1–32.
- Folse, K. (2004). *Vocabulary myths*. Ann Arbor: University of Michigan Press.
- Francis, N., & Kucera, H. (1982.) *Frequency analysis of English usage*. Boston: Houghton Mifflin.

- Green, D., & Meara, P. (1995). Guest editorial. *Computer-Assisted Language Learning* 8(2-3), 97-101.
- Haynes, M., & Baker, I. (1993). American and Chinese readers learning from lexical familiarization in English texts. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second Language Reading and Vocabulary Acquisition* (pp. 130-152). Norwood, NJ: Ablex.
- Hincks, R. (2003). Pronouncing the academic word list: Features of L2 student oral presentations. In *Proceedings of the 15th International Congress of Phonetics Sciences* (pp. 1545-1548). Barcelona: Universitat Autònoma de Barcelona. Retrieved March 29, 2006, from [http://www.speech.kth.se/ctt/publications/papers03/icphs03\\_1545.pdf](http://www.speech.kth.se/ctt/publications/papers03/icphs03_1545.pdf)
- Hinkel, E. (2001). Giving personal examples and telling stories in academic essays. *Issues in Applied Linguistics*, 12, 149-170.
- Huckin, T., & Bloch, J. (1993). Strategies for inferring word meaning in context: A cognitive model. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary acquisition*, 153-178. Norwood, NJ: Ablex.
- Hulstijn, J. (1998, March). *There is no learning without attention*. Paper presented at the annual meeting of the American Association of Applied Linguistics, Seattle, WA.
- Hulstijn, J. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). New York: Cambridge University Press.
- Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80, 327-339.
- Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51, 539-558.
- James, M. (1996). *Improving second language reading comprehension: A computer-assisted vocabulary development approach*. Unpublished doctoral dissertation, University of Hawaii, Manoa.
- Joe, A. (1998). What effect do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, 19, 357-377.
- Jones, F. (1995). Learning an alien lexicon: A teach-yourself case. *Second Language Research*, 11(2), 95-111.
- Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. (1995). Does textual enhancement promote noticing? A think-aloud protocol analysis. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1-63). Manoa: University of Hawaii Press.
- Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal*, 78, 285-299.
- Kojic-Sabo, I., & Lightbown, P. (1999). Students' approaches to vocabulary learning and their relationship to success. *The Modern Language Journal*, 83, 176-192.
- Laufer, B. (1990). Why are some words more difficult than others? Some intralexical factors that affect the learning of words. *International Review of Applied Linguistics in Language Teaching*, 28(4), 293-307.
- Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 20-34). Cambridge: Cambridge University Press.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19, 255-271.

- Laufer, B., & Hulstijn, J. (1998, March). *What leads to better incidental vocabulary learning: Comprehensible input or comprehensible output?* Paper presented at the Pacific Second Language Research Forum, Tokyo.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1–26.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307–322.
- Laufer, B., & Paribakht, S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48, 365–391.
- Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal*, 28(1), 89–108.
- Lee, S. (2003). ESL learners' vocabulary use in writing and the effects of explicit vocabulary instruction. *System*, 31, 537–561.
- Leki, I., & Carson, J. (1994). Students' perceptions of EAP writing instructions and writing needs across the disciplines. *TESOL Quarterly*, 29, 81–101.
- Lessard-Clouston, M. (1994). Challenging student approaches to ESL vocabulary development. *TESL Canada Journal/Revue TESL du Canada*, 12, 69–80.
- Liu, D. (2003). The most frequently used spoken American English idioms: A corpus analysis and its implications. *TESOL Quarterly*, 37, 671–700.
- Ludwig, J. (1984). Vocabulary acquisition as a function of word characteristics. *The Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes*, 40, 552–562.
- McCarthy, M., O'Dell, F., & Shaw, E. (1997). *Vocabulary in use*. New York: Cambridge University Press.
- Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, 37, 645–670.
- Nation, P. (2001). *Learning vocabulary in another language*. New York: Cambridge University Press.
- Paribakht, T., & Wesche, M. (1997). *Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition*. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 174–200). Cambridge: Cambridge University Press.
- Paribakht, T., & Wesche, M. (1999). Reading and “incidental” L2 vocabulary acquisition. *Studies in Second Language Acquisition*, 21, 195–224.
- Parry, K. (1993). Too many words: Learning the vocabulary of an academic subject. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 46–64). Norwood, NJ: Ablex.
- Prince, P. (1995). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *The Modern Language Journal*, 80, 478–493.
- Sanaoui, R. (1995). Adult learners' approaches to learning vocabulary in second languages. *The Modern Language Journal*, 79, 15–28.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129–158.
- Schmitt, N. (1998). Measuring collocational knowledge: Key issues and an experimental assessment procedure. *ITL—International Journal of Applied Linguistics*, 119–120, 27–47.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N., & Schmitt, D. (1993). Identifying and assessing vocabulary learning strategies. *Thai TESOL Bulletin*, 5(4), 27–33.

- Schouten-Van Parreren, C. (1995). Action psychology as applied to foreign language vocabulary acquisition. *Computer-Assisted Language Learning*, 8, 181–204.
- Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second Language Research*, 14, 103–135.
- Walters, J., & Wolf, Y. (1996). Language awareness in non-native writers: Metalinguistic judgments of need for revision. *Language Awareness*, 5, 3–25.
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15, 130–163.
- Zimmerman, C. (1997). Do reading and interactive vocabulary instruction make a difference? An empirical study. *TESOL Quarterly*, 31, 121–140.