

Proficiency Differences in CALL-Based Vocabulary Learning: The Effectiveness of Using “*Power Words*”

Chieko Kawauchi
Kurume University

Introduction

It is self-evident that vocabulary knowledge and skills are important for successful communication in a second language. Words are the units of meaning. Sentences, paragraphs, and whole texts are formed from words. Language ability is often regarded as the number of words that we know.

Many L2 learners see second language acquisition as essentially a matter of learning vocabulary, so they devote a great deal of time to memorizing lists of L2 words and rely on their bilingual dictionary as a basic communicative resource. And language teachers generally recognize the importance of vocabulary learning and are exploring more effective ways of promoting it. Vocabulary can be seen as a priority area in language teaching, requiring tests to monitor the learners' progress in vocabulary learning and to assess how adequate their vocabulary knowledge is to meet their communication needs.

A lot of research has been done recently, and new ways to assess learners' vocabulary knowledge have been also developed. In addition to well-known tests like multiple-choice (choose the correct answer), completion (write in the missing word), translation (given the L2 equivalent of the word), and matching (match each word with its meaning), there are now some other types of vocabulary tests. For example, to investigate the depth of word knowledge, Vocabulary Knowledge Scale was developed on the basis of learners' self-report (Paribakht & Wesche, 1993). Word Associates Test was also made by Read (1993, 1995). In order to assess the size of learners' vocabulary, the Vocabulary Levels Test was developed by Nation (1990, 2004) and Laufer and Nation (1995).

Now, let me briefly explain why CALL-based vocabulary learning can be one of the effective ways in EFL situations like Japan. As you may know, Japanese people are notorious for their bad communication skills. Various reasons can be considered, but

one of them is due to the shortage of vocabulary knowledge (Kawauchi, Kamimoto & Nagasawa, 2005). English teaching in college usually consists of skill-based activities with an emphasis on reading along with ESP type textbooks. Vocabulary teaching is rarely seen (Ishikawa, 2004). However, there is an increasing demand of developing students' vocabulary level in the rapid development of the internet and communication needs in societies and companies. Vocabulary learning is now one of the important objectives in English teaching in Japan.

One of the serious problems we are facing in college English teaching in Japan is a diversity of students in their proficiency levels. Since most of the English classes in Japan are not based on proficiency levels, teachers see diverse students in the large mixed classes. Moreover, students meet once a week and have few opportunities to communicate with native speakers of English. In these situations, CALL-based vocabulary learning can be one of the ways to develop students' vocabulary.

Our university introduced a CALL-program, *PowerWords*, in 2003 for various reasons. First of all, with this program students can study based on their own vocabulary levels. Second, students' learning history is recorded and their development is easily checked. Third, since this program is composed of various tasks such as spelling and listening practices along with various games, not only more proficient but also less proficient students can find something interesting in it. Finally, students can choose their convenient time to practice, which will help them become autonomous learners. The present study reports the effects of CALL-based *PowerWords* for the students who studied it throughout 2 semesters, or 20 weeks.

The study addresses the following two research questions:

1. Is there any difference in the effects of self-study of *PowerWords* between upper and lower levels of students in two semesters?
2. Is there any difference in students' perceptions of self-study of *PowerWords* between lower and upper groups and between the first and second semesters?

CALL Program Called *PowerWords*

First, let me briefly explain the contents of *PowerWords*. This program uses "Standard Vocabulary List (SVL) 12,000" developed by the publisher ALC. ALC states that the SVL is based on 15 kinds of corpuses including British National Corpus along with a

special consideration for Japanese learners of English. There are 12 levels each of which consists of 1000 words. Each level has 50 units and each unit has 20 words. ALC says that it usually takes 3 months to finish one level. Each word is practiced in three different ways: translation from L2 to L1, L1 to L2, and spelling. Some examples are shown below.



There are five types of fun games included such as scramble games, completion, and matching exercises. When they make frequent errors for a certain word, that word is put into the black box and randomly appears to reinforce learning. The student's learning history is also saved and accessible any time by teachers.

Method

Participants

A total of 63 students initially participated in this experiment, but during the 20 weeks 9 students dropped out. Out of the remaining 54 students, 15 upper and 15 lower level students were selected based on the pre-test (see below). The study focuses on these two groups of the students who participated in the self-study with *PowerWords*. They were first-year students at Kurume University, majoring in International Politics.

Procedure

The data were collected at three stages; April, July, and December. In April, the beginning of the first semester, the participants took the diagnostic test to see the appropriate level of *PowerWords* to start with. The students were told that they should visit the university Language Lab Self-Study Room at least once a week. To encourage this self-study, students were told that if they completed the level they started with or 50 units during a semester, they would get 20 points. In order to estimate lexical gains over the period, two parallel tests called Version 1 and Version 2, which

served as a pretest and posttest, were developed by using the beginning 5,000 words out of 12,000 wordlist. Half of the students took Version 1 and the other half Version 2 as a pre-test in April before they started *PowerWords*. The post-test was given in the reverse way.

In July, or the end of the first semester, they took the post test and the students' individual learning history was examined. They were also asked to answer a questionnaire about their feedback on the use of *PowerWords*. The questionnaire was composed of 12 questions such as the degree of enjoyment, usefulness, functions of sounds and blacklist, future use, target levels, etc.

In December, or the end of the second semester, to examine the effect of the 20 week self-study with *PowerWords*, I gave them either one or the other versions of the test again. We also gave them the same questionnaire to see if there were any changes in their perceptions about this program.

Pre- and Post-Tests

As explained earlier, Levels 1 to 5 were used, and the two parallel tests were made for each level to avoid the practice effect. Following the Vocabulary Levels Test by Nation (1990), 72 content words were randomly selected from each 1000 word level list for Version 1 and Version 2. Half of them were randomly classified into test items and the other half into the distracters. Each question consisted of 3 test items and 3 distracters (see some examples below). A total of 6 questions were made for each level. Thus, 5 level tests account for 30 questions altogether along with a total of 90 test items (30 questions x 3 test items) for each version. The test items were examined by using the Japanese translations included in *PowerWords*.

The 1000 word level (Level 1)

1 church
 2 earth ---- 協会
 3 her ---- 公園
 4 hospital ---- 地球
 5 park
 6 towel

The 2000 word level (Level 2)

1 degree
 2 directions ---- 税金
 3 sunset ---- 程度
 4 score ---- 日没
 5 tax
 6 thought

The 3000 word level (Level 3)

- 1 angle
 2 carrot ---- ヤギ
 3 goat ---- 角度
 4 insurance ---- 保険
 5 passion
 6 temperature

The 4000 word level (Level 4)

- 1 bodyguard
 2. draft ---- 医師
 3 guidance ---- 下書き
 4 institute ---- 少数
 5 minority
 6 physician

As mentioned earlier, two parallel tests, called Version 1 and Version 2, were made for each level to avoid the practice effect. In order to see the reliability for these tests, the results of the two tests given in April were compared. The results are shown in Table 1. Maximum scores for each level is 18 (6 questions x 3 test items).

Table 1. The results for the two versions

	Level 1		Level 2		Level 3		Level 4		Level 5	
	Ver 1	Ver 2	Ver 1	Ver 2	Ver 1	Ver 2	Ver 1	Ver 2	Ver 1	Ver 2
mean	17.88	17.63	16.42	16.23	15.24	14.13	11.82	10.33	13.30	11.13
SD	0.55	0.62	1.92	1.77	2.24	3.15	3.41	3.63	2.86	3.20
<i>t</i> 検定	1.68		0.41		1.62		1.67		2.84	
<i>p</i> 両側	0.09 (ns)		0.68 (ns)		0.11 (ns)		0.09 (ns)		0.006*	

There are no significant differences between Version 1 and Version 2 for Levels 1 to 4, suggesting that the two versions are equivalent for these levels. However, the two tests for Level 5 yielded a significant difference, and the result showed that Version 2 was significantly more difficult than Version 1. Therefore, the study is going to look at the Level 1 to Level 4 alone.

Based on the results with the pre-test, upper 15 and lower 15 students were selected. There was a significant difference in their total scores between the two groups ($t=14.47$): Low ($x=52.2$ $SD=3.59$) < High ($x=67.33$ $SD=1.88$). Therefore, it is fair to compare them to examine the research questions.

Results and Discussion

Research Question 1: Proficiency Differences in the Effects of Self-Study of *PowerWords*

The first question was examined by looking at two aspects: the effects on the scores and the relationships between the gains of scores and the amount of practice. The one-way ANOVAs (repeated) were carried out to see if there are any significant differences between the three tests given in April, July, and December. First of all, the results for the lower group are shown in Table 2 and those for the upper group in Table 3.

Table 2. Results for Three Tests (April, July, December) in the Lower Group

	Level 1			Level 2			Level 3			Level 4			Total		
	A	J	D	A	J	D	A	J	D	A	J	D	A	J	D
<i>x</i>	17.3	17.87	17.87	14.67	15.53	16.53	12.4	13.13	14.4	7.8	9.33	11	52.2	55.87	59.8
<i>SD</i>	0.98	0.36	0.35	1.18	2.64	1.64	2.26	3.36	2.20	1.86	3.60	2.27	3.59	8.03	5.43
<i>F</i>	3.56* (p=.037) April<July, April<December			3.55* (p=.037) April<December			2.17 (p=.126) Non significant			5.34* (p=.008) April<December			6.08** (p=.004) April<December		

Table 3. Results for Three Tests (April, July, December) in the Upper Group

	Level 1			Level 2			Level 3			Level 4			Total		
	A	J	D	A	J	D	A	J	D	A	J	D	A	J	D
<i>x</i>	18	18	17.9	17.7	17.2	17.7	16.8	16.27	16.27	14.8	14.8	13.67	67.3	66.3	65.5
<i>SD</i>	0	0	0.36	0.59	1.37	0.59	0.94	1.48	1.62	1.57	2.90	3.04	1.88	5.19	5.04
<i>F</i>	2.15 (p=.128)ns			1.65 (p=.20)ns			0.74 (p=.481)ns			0.95 (p=.392)ns			0.66 (p=.52)ns		

The lower group showed significant differences in Levels 1, 2, and 4. The post-hoc tests revealed that most of the differences were due to the differences between April and December. These results indicated that 20 weeks of the self-study with CALL were effective for the students of this level. In fact, most of them were judged as Level 1 in the diagnostic test, and they studied Level 1 in the first semester, and most of them proceeded to Level 2 in the second semester ($x=2.0$). Interestingly, the effect of CALL is extended to Level 4 which they did not practice yet. The total scores of L1 to L4 show a large increase. When these scores are converted to 4000 words, it can be roughly said that they learned a total of 424 new words by December. On the whole, it is fair to say that 20 weeks of self-study with CALL brings about the overall increase of vocabulary for the lower level students.

In contrast, the upper group failed to show any significant increase in any levels. The

high scores in Levels 1 and 2 in the pre-test (April) imply that the upper level students had already acquired these levels before they started this program. However, most of them were judged as Level 1 and studied it in the first semester. This suggests that the diagnostic test may be too conservative or the pre- and post-tests may overestimate the student's level. The latest level in December was 3.3, on average. In fact, 9 out of 15 students (60%) had been studying Level 4. In spite of this fact, no effects were found in Level 3 or Level 4 in December. Considering the extended effect in the lower group, however, no significant effect in this group is worth noting.

Regarding the relationship between the score gains and the amount of practice, I examined the correlations between them. To assess the amount of practice, the total number of units they practiced were used. The results yielded no significant correlations in either group. However, when the total scores of the July and December tests were compared with the total amount of practice, there was a significant correlation in the upper group ($r=.47, p=.05$).

To sum up, it can be said that the effects of this program were significant in the lower group, and the effects extended to the higher levels that the students have not studied yet. However, it seems to take 20 weeks to achieve the effects. The upper group did not show any significant increase of test scores, but the amount of practice relates to the total scores, suggesting a partial effect of *PowerWords*. The next question is how these learners felt the CALL-based practice. This is considered to be an important question because their perceptions about it are highly likely to affect their motivation to continue studying with CALL.

Research Question 2: Proficiency Differences in the Perceptions of *PowerWords*

To investigate the second research question, the participants were asked the following 12 questions to answer with the Likert scale of 1 to 5: 1 meaning "I don't agree at all" to 5 "I agree full."

- (1) *Did you enjoy PowerWords?*
- (2) *Do you feel that your vocabulary increased?*
- (3) *Do you think the amount of practice (5 units or 30 minutes per week) required was adequate?*

- (4) Do you think the object of completing one level in a semester was adequate?
- (5) Do you think the sound function of PowerWords was useful?
- (6) Do you think the blacklist is a good idea?
- (7) Do you want to continue using PowerWords to study upper levels?
- (8) Do you think it is appropriate to include the amount of practice in the final course grade?
- (9) Do you think PowerWords as a whole is a good material?
- (10) How long do you think it appropriate to study a week?
(15 min, 30 min, 45 min, 60 min, others)
- (11) Which level do you want to achieve?
(Level 3, Level 4, Level 5, Level 6, Level 7)
- (12) Write any comments about PowerWords.

Table 4 shows the average points for the questions (1) to (11) along with the results of the *t*-tests between the first semester and second semesters for each group. It also shows the results of the *t*-tests between the lower and upper groups regarding the questionnaire given in second semester.

Table 4. Results for the questionnaires given July and December

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Low 1st	3.1	3.6	3.0	3.5	3.3	3.1	3.5	3.5	4.1	35.0	4.5
Low 2nd	2.7	3.8	2.9	3.5	3.6	3.1	3.7	4.2	3.9	38.0	4.5
<i>t</i>	1.1ns	1.0ns	0.2ns	0ns	0.62ns	0.38ns	0.49ns	2.75*	1.0ns	0.90ns	0.27ns
Upper 1st	3.5	3.4	3.3	3.6	4.1	3.2	3.6	3.8	3.8	38.1	5.5
Upper 2nd	3.1	3.3	3.1	3.5	3.6	2.7	3.6	3.8	3.7	34.7	5.6
<i>t</i>	2.30*	0.47ns	0.35ns	0.35ns	2.17*	1.88ns	0.03ns	0.18ns	0.46ns	0.83ns	0.29ns
Low/Upper	0.77ns	1.52ns	0.54ns	0.0ns	0.0ns	1.25ns	0.19ns	1.27ns	0.60ns	0.61ns	2.25*
<i>t</i>											

Note. * $p < .05$; ns=non significant

First of all, regarding the differences between the two semesters, the results for the *t*-tests show that most of the questions yielded no significant differences except for a few questions. The lower level students answered that question 8, “the practice included in the final grade,” was more agreeable in the second semester. The upper level students found *PowerWords* more enjoyable in July than December and the sound function more useful in December. On the whole, the differences between the two

semesters are not so significant with regard to the feedback on *PowerWords*.

The *t*-tests between the upper and lower groups also show that there are not any significant differences except for question 12, the target level. The upper level students are hoping to acquire Level 5 to Level 6 ($x=5.6$), which is one level higher than the lower level students ($x=4.5$). As shown in Table 4, their perceptions appear to be quite positive. The results for question 9, “the overall usefulness of *PowerWords*,” received rather higher points compared to other questions. This implies that the students in both levels perceived this program very favorably.

How does each question correlate with other questions? The first question, the degree of enjoyment, seems to be a key factor. It highly correlated with “wanting to continue studying with this CALL” ($r=0.51$), “overall evaluation of *PowerWords*” ($r=0.58$), and a “target level they want to master” ($r=0.63$). The second question, “a sense of increasing vocabulary”, correlated highly with “overall evaluation of the program” ($r=0.53$). The question, “adequacy of 1 level completion as a requirement,” also significantly correlated with the question of “*PowerWords* practice as part of the final grade” ($r=0.58$). Finally, there was a strong correlation between “wanting to continue studying with CALL” and “target level” ($r=0.50$). In other words, those students who enjoyed this practice and felt their vocabulary increased found this program useful and want to continue to get higher target vocabulary levels. The requirement of 1 level completion is effective if the practice with CALL is included in the final grade. Finally, the length of practice they want to do at a time was 39 minutes on average.

Last but not least is the written comments made by the students. About 90% of them gave some comments, which is considered to be unusual for this type of question. 70% of the comments were favorable. 49% of these comments showed the enjoyment of the study, saying something like they were hooked on the program. 33% responded that they felt their vocabulary is increasing and a sense of accomplishment, saying that they felt fulfillment after finishing one level.

Negative comments account for 21%. They can be summarized into two major comments: tiredness and the volume of the exercises. The tiredness might have been caused by some students who had to catch up with the absences at the very end of the term. The second comment simply shows that some students felt there were too many words for each level. There is one comment worthy of note here. One student wrote

as follows:

“As the level goes up, it becomes less interesting and I feel I may not remember what I have learned.”

This student was in the upper level and studied nearly 90 units in the first semester and proceeded to Level 4 in the second semester. She is feeling that just practicing vocabulary like this does not give her a sense of acquisition. She is afraid that what she learned may not be stored in her vocabulary bank. This causes us to think the ways how the self-study with *PowerWords* should be well integrated into real communicative ability. Under a limited opportunity of communication in Japan, one of the possible solutions will be to combine the vocabulary learning with graded readers.

To summarize Research Question 2, we can say that there are not so distinctive differences between the two groups or between the two semesters regarding the perceptions for this program. The overall preferable feedback to this program is noteworthy, since it suggests that CALL-based vocabulary learning can lead to an effective self-study regardless of learners' proficiency levels. It also suggests that students are willing to continue learning this way for a long time, which will help them become autonomous learners.

Conclusion

This study revealed that CALL-based vocabulary learning was effective more for the students of the lower level than for those of the upper level. However, their perceptions were quite preferable and there were no significant differences between the two groups. Moreover, their perceptions remained for two semesters, suggesting that this CALL-based vocabulary learning will help them to be autonomous learners. However, some problems remain. First, we have to find some effective ways to deal with higher level students. Second, we have to create a syllabus which incorporates the learned vocabulary into higher levels of English ability. Third, there is a need to develop more reliable tests to assess learners' vocabulary size.

References

- Ishikawa, S. (2004). JACET 8000 oyobi kakushu kyouiku software-o mochiita daigaku-deno goi shido (College vocabulary teaching by using JACET 8000 and some educational software). In *JACET 8000 Katsuyo jireishu*, edited by JACET Basic Vocabulary Revision Committee, 7-33. Tokyo: JACET.
- Kawauchi, C., Kamimoto, T., & Nagasawa, S. (2005). CALL-based vocabulary learning: The effectiveness of using *PowerWords*. *Bulletin of the Institute of Foreign Language Education of Kurume University*, 12, 39-72.
- Nation, P. (1990). *Teaching and learning vocabulary*. New York: Heinle and Heinle.
- Nation, P. (2004). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, 16, 33-51.
- Paribakht, T. S., & Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, 11, 9-29.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10, 355-371.