To Push or Not to Push: A Vocabulary Research Question
Charles Browne, Aoyama Gakuin University

Abstract

In this experiment, 122 university students in 4 intact classes learned 10 new vocabulary words via three different methods representing strong theoretical perspectives on second language vocabulary acquisition; extensive reading (Input-Hypothesis), activity-based (Task-based learning), and writing words in original sentences (Pushed Output Theory and Depth of Processing Theory). Results indicate that regardless of the language level of the learner, significantly more words were learned by the Pushed-Output condition. Pedagogic implications are discussed.

Introduction

Although there seems to be a growing acceptance among SLA researchers that vocabulary is a crucial component of overall communicative competence (Schmitt and McCarthy, 1997), it is still an area that is often neglected both in and outside the classroom. Zimmerman (1997) argues that "the teaching and learning of vocabulary has been undervalued in the field of second language acquisition (SLA) throughout its varying stages and up to the present day (pg. 5)". Paribakht and Wesche (1997, p.174) point out that there still remains an enormous amount of research to be done in the area, as "it is still far from clear how learners acquire vocabulary or how it can best be taught".

In terms of pedagogy, if vocabulary is addressed at all in the ESL classroom, it is has traditionally been relegated to the role of incidental learning. Sokmen (1997) summarizes the way vocabulary has usually been handled:

> For many of us, our perspective on the teaching of vocabulary was greatly influenced by the top-down, naturalistic approaches of the 1970s and 1980s. The emphasis was implicit, incidental learning of vocabulary. We were taught the importance of directing L2 learners to recognize clues in context…Textbooks emphasized inferring word meaning from context as the primary vocabulary skill (p.237).

Incidental learning of vocabulary has to a large extent, then, become the status quo. Even when vocabulary is taught more directly in the classroom, it tends to be via supplementary workbooks containing word manipulation exercises, rather than as a main element of the curriculum. Schmitt (2000) points out that these types of "approaches to vocabulary have unsurprisingly focused on activities for the explicit study of vocabulary" (p. 146). Indirect support for the popularity of this type of approach can be found by making a quick scan of ESL/EFL vocabulary books sold by major publishing houses such as Addison-Wesley, Cambridge University Press, or Heinle & Heinle. Not only are there no course books offered which have vocabulary as an organizing or main component of the syllabus, but the supplementary vocabulary books which are offered tend to be very similar in approach, introducing new words and definitions on one page, followed by a series of activities to give practice using the new words (for example, crossword puzzles, matching activities, word searches, fill in the blank, and the like). Current examples of this type of book include: the English

Vocabulary in Use series (McCarthy, O'Dell, and Shaw 2001, Redman, 1999, Redman & Shaw, 1999, McCarthy, O'Dell & Shaw, 1997,), the American Vocabulary Builder series (Seal, 1997, 1990), Ways with Words (Kirn, 1984), Developing Vocabulary, (Keen, 1994), or Practice Vocabulary (Broukal, 2001).

The current study discusses the theoretical and methodological underpinnings of these two widely used techniques for teaching vocabulary, and then compares their effectiveness, with a third technique, learning vocabulary via simple writing tasks, which, although not widely used in the classroom for the past few decades, may actually be more representative of current thinking about the second language acquisition process, including *Depth of Processing Theory* (Craik and Lockhart, 1972), and the *Pushed Output Hypothesis* (Swain, 1985).

The Role of Input in SLA

Although the theories and research of Stephen Krashen are by no means universally accepted, since the early 1980s, he has arguably influenced the debate on the role of input in SLA more than any other researcher. According to Krashen's *Input Hypothesis* (Krashen, 1983, Krashen, 1985), human beings acquire languages in only one way – by understanding messages, or by receiving "comprehensible input". Learners progress along a natural order by understanding input that contains vocabulary and structures slightly above their current level of understanding (what Krashen calls "i+1 input"). They are able to understand language which contains unacquired grammar and vocabulary through the help of contextual clues such as extra-linguistic information, knowledge of the world, and previously acquired competence.

Krashen (1985), strongly downplays the role of output or interaction with his claim that comprehensible input is both necessary and sufficient for language acquisition. He states that:

> Speaking is the result of acquisition, not its cause. Speech cannot be taught directly, but emerges on its own as a result of building competence via comprehensible input (p. 2).

Although Krashen's assertion that language production doesn't play a role in language acquisition is a controversial one (Gregg, 1984, Sharwood Smith 1986, Ellis, 1990), his claim that comprehensible input is an essential ingredient for acquisition seems to be widely accepted within the field of applied linguistics (Long 1983, Swain, 1981, Brown, 1985, Ellis, 1985).

Extensive Reading

In the mid 80s, this belief that exposing learners to large quantities of comprehensible input was both necessary and sufficient to cause language acquisition was especially prevalent in the area of reading skills and vocabulary development. Krashen (1989) claimed that the most effective way for learners to acquire new vocabulary and develop their spelling ability is through exposure to large amounts of "comprehensible input" via extensive reading. Since free reading of materials that students like is also a low-anxiety activity, Krashen argues that such reading activities are the most efficient means by which a learner can acquire new vocabulary (Krashen 89, 93a & b).

His "Natural Approach" became the rallying cry of "learn to read by reading" (Smith, 1982). Reading proponents such as Nutall (1982, 1996), claimed that extensive reading programs were the most effective way of increasing vocabulary size, improving reading skills, and developing overall language ability. Studies, such as Robb and Susser's (1989) comparison of extensive reading with a skill building approach, Hafiz and Tudor's (1989) 3 month experimental study of extensive reading via graded readers, and Krashen's (1988) discussion of free reading programs and student's self reports of free reading all found that extensive reading worked at least as effectively as the methods it was compared with, and correlated well with success in reading comprehension.

Elley's "book flood" studies (Elley & Mangubai, 1981, 1983) looked at the effect of a "book flood" of 250 high interest children's books on 380 students, in comparison with a control group of 234 students over a 20 month period and found that the oral reading of children's stories to elementary school children learning English as a second language led to large, long term gains in vocabulary (twice the rate of the control group).

More indirect support for Krashen's *Input Hypothesis* can be found in Grabe and Stoller's (1997) case study of Grabe's attempt to learn Portuguese primarily via extensive reading of materials that were of specific interest to the subject. One of the chief goals of the study was to "explore the extent to which extensive reading practice without formal instruction would promote reading ability and vocabulary development in Portuguese. The authors found that it was "clear that Bill made reasonably good progress learning to read with the primary input being through extensive reading…" (p. 113) The study also found that "reading and vocabulary are reciprocally causal… that reading improves vocabulary knowledge and vocabulary knowledge supports reading development" (p. 119), and that the study "strongly suggests that reading and vocabulary will develop as the result of extensive reading practice" (p. 119).

The net result of Krashen's claims regarding the need for large amounts of comprehensible input, and the positive benefits of extensive reading, was that since the 1980s, extensive reading has became one of the most widely used methods for developing reading and vocabulary skills.   The Role of Output in SLA

 The Role of Output in SLA

Despite a general agreement among researchers that comprehensible input is an essential element of SLA, most now maintain that it is not sufficient. Swain's observational data of the French immersion program in Canada (Allen, Swain, Harley & Cummins, 1990) notes that although students received large amounts of "comprehensible input" over a period of many years, with many corresponding opportunities for "interaction" as Long defines it, few ever exhibited a full mastery of French.

The researchers noted that most classes tended to be quite teacher-centered with students giving only short answers to teacher questions (less that 14% of students' classroom utterances were more than a phrase in length). Swain argues that part of the problem was that these learners had had little opportunity to engage in two-way negotiated exchanges in the classroom. She hypothesized that in order for students to

achieve full competence, their linguistic resources needed to be stretched. What was missing was 'comprehensible output'. Swain (1985) points out that;

> To achieve native-speaker competence, the meaning of 'negotiating meaning' needs to be extended beyond that usual sense of simply 'getting one's message across'. Simply getting one's message across can and does occur with grammatically deviant forms and sociolinguistically inappropriate language (p.248).

According to Swain, negotiating meaning needs to incorporate the notion of being 'pushed' toward the delivery of a message that is not only conveyed, but conveyed precisely, coherently, and appropriately. This act of "pushing", Swain (1995) argues, leads learners to make more of an effort, to "stretch" their interlangauge resources, which forces them to process language more deeply, and helps them to move beyond their current stage of language development.

Since Swain first proposed the *Pushed Output Hypothesis* in 1985, several studies have given it qualified support. Pica, Holliday, Lewis and Morgenthaler (1989), for example, found that in response to requests for clarification or confirmation, learners tended to modify their output. Although the study did not show that these conversational modifications specifically led to acquisition, one of the assumptions of the Output Hypothesis is that such modifications contribute to the process of SLA. Kowal & Swain (1997), observed students working on three different types of collaborative tasks (dictagloss, cloze, and proof-reading) in a French immersion program, and found evidence of students "noticing the gap" between what they wanted to say and what they were able to say. As predicted by the Output Hypothesis, this happened primarily as students were "pushed" to produce the target language.

In an exploratory study involving 6 learners in a two way information gap task with the researcher (3 experimental and 3 control), Nobuyoshi and Ellis (1993) found evidence which suggests that pushing learners to produce output leads to better acquisition. Experimental subjects were asked for clarification requests by the researcher each time they produced an utterance that contained a past tense error, while the control group was not. One week later, two of the three experimental subjects showed improvement in their use of the past tense, while no improvement was shown by any of the control subjects.

Swain and Lapkin (1995), looked at 18 students from a grade 8 early French immersion class who were trained to use think aloud procedures while writing and article for a newspaper and found that as they were encouraged to produce their L2, students noticed gaps in their linguistic knowledge. Again, although the study did not specifically show that noticing led to acquisition, the authors argue that this type of noticing plays a consciousness-raising function which trigger cognitive processes such as hypothesis testing about how the L2 works, which have already been implicated in previous studies of second language learning (Selinker, 1972, Corder 1981, cited in Swain, 1995).

Due to their focus on acquisition of advanced grammatical structures via collaborative tasks, most studies of pushed output have been qualitative in nature, yielding little quantitative evidence that pushed output enhances acquisition. The current study is an

attempt to establish such quantitative data by focusing on the learning of new vocabulary words (rather than grammatical structures) which can be treated as discrete points in a statistical analysis, and operationalizing pushed output as something that can occur within an individual task.

The Depth of Processing Theory

The underlying mental processes which help pushed output to foster second language acquisition may be partially explained by studies of cognition and memory done in the early 1970s. In their seminal criticism of "multistore" approaches to explaining human memory, Craik and Lockhart (1972), proposed the *Depth of Processing Theory*. Until that time, most research had been concerned with the distinction between short-term and long-term memory (Broadbent, 1958, Waugh & Norman, 1965), with the assumption that the longer a new piece of information was held in short term memory, the better chance there was of it becoming part of long term memory (Baddeley, 1966). Craik and Lockhart argued that the most important factor is actually the shallowness or depth by which a new piece of information is initially processed, a process related to attention. The authors point out that shallow, sensory levels of depth might be characterized by processing the stimuli simply in terms of its visual or acoustic properties, whereas deeper, more semantic levels of processing might involve analysis of meaning, compatibility with the analyzing structures, and processing time. In a subsequent article, Craik and Tulving (1975), expand upon this aspect of the theory:

> Stimuli which do not receive full attention, and are analyzed only to a shallow sensory level, give rise to very transient memory traces. On the other hand, stimuli that are attended to, fully analyzed, and enriched by associations or images yield a deeper encoding of the event, and a longer lasting trace (p.270).

Brown and Perry (1991) looked at three vocabulary learning strategies for ESL students that were differentiated according to Craik and Lockhart's (1972) *Depth of Processing Theory*. Six intact classes at two levels of proficiency were divided into three treatment groups (keyword, semantic, and keyword-semantic), and then received 4 days of instruction. Results were consistent with predictions made by the theory, with information processed at the semantic level being remembered better than information processed at the acoustic and visual level, and information that was processed at multiple levels (keyword-semantic) remembered best of all.

Learning Vocabulary via Classroom Tasks

As mentioned in the introduction, if vocabulary is addressed at all in the classroom, it is usually by word manipulation tasks such as crossword puzzles, word searches, matching activities and the like. In addition to the many classroom texts of this type offered by the major ESL publishing houses (for example, McCarthy, O'Dell, and Shaw 2001, Redman, 1999, Redman & Shaw, 1999, McCarthy, O'Dell & Shaw, 1997, Seal, 1997, 1990, Kirn, 1984, Keen, 1994, Broukal, 2001), there are also a number of books offered to instruct the teacher in such approaches (for example, Nation, 1994, Taylor, 1992, Allen, 1983).

Several research studies have been conducted to compare the effectiveness of such tasks. In an experiment designed to look at the role negotiation of meaning had in

language acquisition, Newton (1995), examines the vocabulary gains made by an adult ESL learner while performing four communicative tasks. Although the experiment was. Although negotiation had the least affect on language acquisition of the four factors Newton examined, the study did have two important findings. First, that the actual use of the new vocabulary word in the process of completing the task was a good predictor of whether or not the word would be learned, and second, that the learners who made the greatest gains on the post-test were those who used the story vocabulary most generatively (that is in new contexts or structures).

Zimmerman (1997b), did a 10 week classroom-based study of the effect that L2 interactive vocabulary instruction would have on vocabulary gains. The experimental group received 3 hours per week of focused, interactive vocabulary instruction and were asked to do self-selected reading assignments, while the control group had the same self-selected reading assignment, but without the addition vocabulary instruction. Although it may not be surprising that the experimental group made greater vocabulary gains than the control group, considering the much greater amount of instruction the experimental group received, one interesting result was that the experimental group completed 50% more required reading than the control group which may indicate that the additional vocabulary instruction helped to make the reading assignments easier to understand and increased student motivation to read more.

Hall, (1992), compared the effects of learning mathematics vocabulary via split-information tasks with that of teacher-fronted learning and individual study. Although all groups made vocabulary gains, results indicated that learners in the split-information condition made significantly more gains than the other two conditions. Interestingly, Hall found that there was only a low correlation (.36) between number of exposure to the word and the learning of the word, while there was a much higher correlation (.93) between learning and the number of times learners actually used the words in ways that were not just repetitions of the way the word was presented in the input.

Vocabulary Size and SLA

In an article titled "The Mathematics of Language", Kucera (1982), points out that human language exhibits the somewhat contradictory characteristics of both efficiency and redundancy. For example, the English language is redundant in the sense that it has a only a limited number of permissible phonemes (33), as well as strict rules on what phonemes can occur together to form words (i.e. trip but not tlip). It is also extremely efficient in the sense that the vast majority of high frequency words are very short - 57% of the words in the one million word Brown Corpus (Kucera, 1982), are four or fewer letters, while the repeat-rate for long words is extremely low - "for every occurrence of a ten-letter word there are eight occurrences of a three-letter word, and for every occurrence of a twenty-letter word there are 3,524 occurrences of a three-letter word (Kucera, 1982, pg. 39).

The results of the numerous corpus-based word frequency studies done over the past few decades (Thorndike & Lorge, 1944, West, 1953, Kucera, 1982, Johansson & Hofland, 1989), reveal that this efficiency goes beyond word length alone. Despite research which estimates that the average 18 year old native English speaker has a vocabulary of somewhere between 16,000 (D'Anna, Zechmeister, & Hall, 1991) and 40,000 words (Nagy & Anderson, 1984), and the enormous total amount of words in

the English language (there are 128,000 words, for example, in a large Webster's dictionary), the most frequent words in the English language account for a disproportionate amount of the total number of running words readers encounter on a typical page of written text.

Table 1 summarizes the findings of Nation (1990), which is based on his own research on acquisition of technical vocabulary and vocabulary in university settings, as well as West's (1953) General Service List. According to these figures, knowledge of just the 1000 most frequent words in the English language (less than 1% of the words appearing in the Webster's dictionary) would allow a reader to understand approximately 75% of the words appearing on a page of text. Similar findings have been reported in Engels (1968), and Johansson & Hofland (1989). If readers know an additional 1000 words the percentage of coverage jumps to an impressive 87%.

Table 1: Frequency in Terms of % Coverage of Running Words in a Text

Type of Word	# of Words	% Known/Page
High Frequency Words	1000	75%
High Frequency Words	2,000	87%
University Word List	836	8% (academic texts)
Technical Words	2,000	3% (technical texts)
Low Frequency Words	123,200	2%

Nation (1990), also found that a small number of additional words beyond the first 2000 occurs quite frequently within the genre of academia. Referred to in the literature as the "University Word List", knowledge of just these additional 836 words gives learners understanding of about 8% more words on the page, for a total of 95%. Nation (1990), and others (for example, see Marshall and Gilmour 1993), have demonstrated that within specific technical fields, certain words occur quite frequently, and that mastery of the 1000-2000 specialized words associated with that field can give the learner an additional 2-3% coverage.

The Relationship Between Lexical Size and Reading Ability

The importance of these frequency figures and the high percentage of coverage of running words that they offer becomes clear as one considers the close relationship between lexical knowledge and reading ability, as well as the growing body of research on how many words a non-native speaker needs to know to interact effectively with unsimplified texts. Chall (1987), for example, found that student scores on vocabulary measures were so highly correlated with reading comprehension scores, that reading vocabulary quizzes could be substituted for paragraph meaning tests.

Although classroom teachers often hear complaints of student's over-reliance on dictionaries along with the admonishment that they need to learn to guess meaning from context, Marshall and Gilmour's (1993) study of the relationship between lexical knowledge and reading ability in ESP students found that top down reading skills such as schema activation clearly depended on students already having a large vocabulary.

In an analysis of the readability and lexical load of the top three selling Japanese high school ESL reading textbooks, Browne (1998) found that student's over reliance dictionaries and translations may be partially explained by the extremely high lexical load of the texts, which were often found to be even more difficult than typical native speaker texts.

In a study to determine what kind of threshold score was needed on the Cambridge First Certificate Exam before students could be said to be able to comprehend academic texts, Laufer and Sim (1985) found that the language base needed for students to be able to guess meaning from context was largely lexical in nature, outranking both knowledge of subject matter as well as syntactic structure.

Hirsch and Nation (1992), in their study of three short, unsimplified children's novels found that a knowledge of the most frequent 2000 words from West's (1953) General Service List gave the reader coverage of about 90% of the running words in the text. The authors point out, however, that this would still leave the reader with one out of every ten words as unknown, and cite research by Laufer (1989), and Liau and Nation (1985), which shows that about 95% coverage of text is necessary for students to be able to reach an acceptable level of reading comprehension, and to be able to guess meaning from context.

Laufer's (1992a) study, titled "How much lexis is necessary for reading comprehension", found that the minimum lexical threshold at which there were more readers than non-readers (operationalized as those students who receiving passing scores on the two standardized reading tests used in the experiment) was 3000 words. This threshold also marked the level at which students who were proficient readers in their L1 were able to transfer their reading strategies to the L2. A follow-up study done by Laufer (1992b), which looked at how L2 reading was affected by lexical knowledge and general academic ability, found that students with a vocabulary size of less than 3000 word families could not read well irrespective of their academic ability, and similarly, that students with vocabulary sizes of 5000 or larger could read well in their L2 regardless of their academic ability.

All of the above studies highlight the importance of student knowledge of the most frequent several thousand words in the English language. Non-familiarity with these words almost assures that students will not be able to use top down skills, activate schema, guess from context, score well on reading exams, or develop reading fluency. This implies that different types of learning activities could vary in their effectiveness according to the size of a learners' vocabulary. For this reason, the current study will examine lexical size as an additional independent variable.

Input Frequency and Vocabulary Learning

There are a number of studies which have looked at the relationship between the number of times a word occurs in the input, and the likelihood that the word will be learned, all of which have the not surprising conclusion that providing more encounters with the word increases the chance that it will be learned. Palmberg's (1987) study of beginning ESL students in Sweden, found that learners were more likely to remember words from the textbook which occurred most often. Elley and Mangubai's "book flood" studies (Elley and Mangubai, 1981, 1983, Elley, 1989, 1991), found that a

vocabulary word's frequency of occurrence in the input was the factor which had the highest correlation with the vocabulary gains made by the elementary school children he was studying. Sokmen (1997), in a review of current trends in teaching second language vocabulary, argues that providing a number of encounters with new words increases the likelihood that the word will be learned. Kachroo (1962, cited in Nation, 1990 p.43) studied the relationship between the number of times a word occurred in a course textbook and found that words with multiple repetitions were the most likely to be learned by students, and words which only occurred once or twice were not known by more than half of the learners.

A study conducted by Saragi, Nation & Meister (1978), found that if words were encountered six or more times, they were learned by over 93% of the learners. Zahar, Cobb and Spada (2001), however, in a review of several such studies which suggest that vocabulary acquisition through extensive reading is very inefficient, with learners typically unable to identify more than 1/14 words tested. They argue that frequency needs are related to learner level, with lower level learners needing more encounters for learning to occur.

Research Questions

The specific research questions for this study were as follows:

1) Do vocabulary learning activities which require pushed output help students to learn more new words than input-based or activity-based vocabulary learning activities?

2) Is there a differential effect for the vocabulary size of the learner as to which type of vocabulary learning activity is most effective?

Stated in terms of variables, the dependent variable in this design will be short term vocabulary learning (as measured by a post-treatment vocabulary quiz), while the independent variables will be task type (input, output or task-based), and learner level (large or small vocabulary size).

These research questions were tested through a quasi-experimental research experiment using a nonrandomized control group, pretest-posttest design. An overview of the research design, which was conducted over a three week period between January and February of 1999, follows in the next section.

METHODOLOGY

Subjects

The subjects for the study were 122 students in four intact classes (two freshman oral English classes, one sophomore reading skills class and one junior reading skills class) at Aoyama Gakuin University, a four-year, private, Christian university in Tokyo, Japan.

Students in all four classes were non-English majors, studying in a department of business administration. About 70% were male, with an average age of18. Since one

of the key research questions in this design relates to variability due to language ability and vocabulary size, no attempt will be made to control the proficiency level of the learners in each class.

In fact, as the data in the results section will reveal, despite some surface similarities, each class displays tremendous variation in the English language ability of its students.

All English classes in this department meet once a week for 90 minutes, for a total of either 23 or 24 classes per academic year (depending on the day of week the class meets, and the number of holidays which occur on that particular day of the week).

The university has approximately 20,000 students attending both undergraduate and graduate programs spread out among its 3 urban campuses. It is prestigious, private Christian university, consistently ranked as one of the top 10 private schools in the country. The university has, throughout its history, placed a very strong emphasis on English education, with the result that even students not majoring in English, such as the ones who participated in this study, often displaying English skills higher than their peers at other universities. Students in each of the four classes were randomly assigned one of three treatments.

Materials

The research experiment presented in this paper makes use of the following materials:

(1)   The Vocabulary Levels Test (Nation, 1992), to establish learner's overall vocabulary size.

(2)   A vocabulary checklist pretest to establish that the 10 new vocabulary words are unknown to the students before treatment.

(3)   A list of the 10 unknown treatment words with definitions given in both Japanese and English. This list is handed out to all three treatment groups before they begin their respective tasks.

(4)   A reading task which introduces the 10 new vocabulary words in the context of the reading.

(5)   A vocabulary activities task which gives the learners practice in using the 10 new words via typical vocabulary tasks such as crossword puzzles, matching activities, and word searches.

(6)   A writing task which simply asks the students to use each of the new words in at least two complete sentences.

(7)   Similar to number six (above), this writing task asks the students to use each of the new words in at least two complete sentences, but also provides one example usage sentence for each target word.

(8)   A 10 word, multiple-choice post-treatment quiz for establishing short term vocabulary retention

Each of these will now be discussed in more detail.

(1) Vocabulary Level Test

Nation's (1990) Vocabulary Levels Test was used to establish the approximate overall vocabulary size of each learner who took part in the experiment. This additional measure was taken since there is a possibility of a differential effect due to the size of a learner's vocabulary as to which of the 3 learning techniques would be most effective.

The two major techniques for assessing vocabulary size are to either make a test based on random sampling from a dictionary, or to make a test based on a frequency list derived from a corpus (Nation, 2001).

The Vocabulary Levels Test, an example of the latter, was chosen because it was felt that having specific information about a student's knowledge of vocabulary at each of several different vocabulary frequency levels might be able to yield more useful information than having just the one overall vocabulary size number which would be generated by the dictionary random sampling technique.

The Vocabulary Levels Test is based on a list of high frequency words known as the General Service List (West, 1953), which was derived from a large corpus of approximately 5 million words. Results from the test gives information about learner knowledge of vocabulary words at the 1000, 2000, 5000, and 10,000 word levels, as well as about their knowledge of specialized academic vocabulary.

At each of the five levels, the test gives students six blocks of six words – in each block, three of the words must be matched with three definitions. In this way, 36 randomly selected words from each level can be tested, even though only 18 words are matched. Table 2 gives an example of one 6 word block taken from the 2000 word level:

Table 2: Sample item from the Vocabulary Levels Test

---

1. original
2. private   ___   complete
3. royal   ___   first
4. slow   ___   not public
5. sorry
6. total

_____

Since only 18 words need to be matched at each level (for a total of 90 questions), the test is quick and easy to administer, requiring only about 45 minutes of class time.

Reported reliability studies on the Levels Test have been quite high: .94 in Read's (1988) report of it's use as a diagnostic pre-test and achievement post test in an intensive English course, and .97 in a report of its use as a norm-referenced instrument

to measure the vocabulary size of Japanese employees at major multinational corporations (Beglar and Hunt, 1999).

(2)   Vocabulary checklist pretest

This type of vocabulary test, also known as the "yes/no test" (Read, 1997), was used to establish that the 10 new vocabulary words were unknown to the students before the treatment.

In this type of test, students are simply presented with a list of words (no definitions are given) and asked to check off the words they know.

The yes/no test was chosen over other vocabulary tests where learners are asked to match words to definitions, in order to control for the possibility that some learning could occur due to taking a pre-test where information about the target words is given.

Although yes/no tests have a long history in L1 and L2 vocabulary research (Nation, 1990, cites studies of the yes/no test from as early as 1929 by Sims, and a 1941 study by Bear and Odbert), one of the chief criticisms of the test has always been that there is no way to control for students over-reporting their vocabulary knowledge by checking off words that they didn't really know.

Anderson and Freebody (1983) came up with an ingenious technique for addressing this possibility, by adding a number of plausible non-words to the checklist. Through the use of simple statistical procedures the validity and reliability of the test could be calculated by adjusting the numbers downward as learners reported false positives (i.e. when they placed a check next to a non-word). This study, as well as subsequent studies by Meara (1990, 1991) and Meara and Buxton (1987), have reported very high reliability coefficients for the test.

(3)   A list of the 10 unknown treatment words with definitions given in both
        Japanese and English.

Having taught English to students in this department for several years, it would not have been difficult to develop a list of 10 words unknown to the learners. Unfortunately, it would be next to impossible to find an authentic text of appropriate length which would contain all 10 these words for use with the students in the group who would be learning the words via extensive reading.

Therefore, it was thought that the best solution would be to first find an appropriate reading passage, and then choose the list of 10 unknown treatment words from the passage itself.

Once the reading was chosen (the reading is discussed in more detail in the following section), I worked with another of the university's native English teachers to develop a list of 48 words from the passage which we thought might be unknown to at least some of the learners. This list was then distributed to one of my classes not involved in the experiment (24 third year students in a reading class identical to one of the ones used in the experiment), and the students were asked to place an "X" next to any of the words which they did not know.

The 10 words which had the highest number of "X"s, were then chosen as the target words.

A one page sheet, which gave simple definitions for the words in both English and Japanese, was then developed with the help of a Japanese teacher of English.

This list is handed out to all three treatment groups before they begin their respective tasks.

(4)   A reading task which introduces the 10 new vocabulary words in the context of the reading.

The reading, "Seasonal Marketing and Timing New Product Introductions" (Radas & Shugan, 1998), was chosen because it was an authentic text from a marketing journal that had been scheduled to be used in one of our department's intermediate-level marketing course. A variety of studies have pointed out that learners are far more likely to be motivated to read a passage related to their major than one that is not (for example, Elley and Mangubai, 1981).

Learning vocabulary via a reading task is usually considered "incidental learning" since the main focus of the activity is not on the learning of new vocabulary words. Most research in this area shows, however, that a single exposure to a new word in such an activity is generally not enough to guarantee acquisition of that word (Meara, 1980, Beck, McKeown, & Omanson, 1987).

Other studies have shown that when the task has been modified to either increase the number of exposures to the word in a single passage, or that attention is drawn to the word in some way such as underlining or glossing (Hulstijn, Hollander & Greidanus, 1996, Watanabe, 1997), the chance that learners will acquire knowledge about the target word is enhanced.

Therefore, the passage containing the 10 target words was rewritten so that each target word occurred at least 2-3 times. In order to draw learners attention to the target words, every occurrence of each of the target words in the passage was also underlined.

(5)   A vocabulary activities task which gives the learners practice in using the 10 new words via typical vocabulary tasks such as crossword puzzles, matching activities, and word searches.

There are a great variety of ESL/EFL textbooks on the market which contain vocabulary exercises. After reviewing a number of them, several of the more frequently appearing exercises where adopted for use in this task. In order to control for time on task, the exercises were first piloted on a small group of learners in another class not involved in the experiment. After one or two trial runs, it was found that 3 vocabulary exercises were needed to approximate the same amount of time used in the extensive reading condition.

(6)   A writing task which asks the students to use each of the new words in at least two complete sentences.

This (deceptively) simple task requires students to write two complete sentences in English using each of the 10 target words. The reason two sentences were required is twofold: First, since this was the "pushed output" condition, it was thought that two sentences would require more thought and effort than one sentence. Second, when piloting the task it was found that the time students were on task was much shorter than in the first two conditions – adding the requirement of writing a second sentence evened out the amount of time needed to complete the task.

(7)   Similar to number six (above), this writing task asks the students to use each of the new words in at least two complete sentences, but also provides one example usage sentence for each target word.

This condition was added almost as an afterthought. As the time drew near for actually conducting the experiment on my students I found myself growing worried that the sentence writing task (without example sentences) would be far too difficult for my students to be able to complete. Past experience with giving students writing assignments taught me that Japanese EFL students tend to rely quite heavily on their dictionaries, which invariably provided them with sample sentences containing the word in question. It was felt that giving a sample sentence for each of the new treatment words would make the task easier for them to complete. For this particular writing condition, an additional reason that two, rather than one sentences were required was to make it more difficult for students to write their sentences simply by making a slight modification of the sample sentence.

(8)   A 10 word, multiple-choice post-treatment quiz for establishing short term vocabulary retention.

Although a variety of vocabulary tests which give more detailed information about students depth of knowledge on individual words were available, a multiple choice format was chosen for this experiment since this type of test is more sensitive to small gains in vocabulary knowledge. The reliability of the post test was calculated using a KR-20 analysis, with a resulting figure of .68.

PROCEDURES

Protocol Development

The purpose of the pilot study, which was conducted between November of 1998 and January of 1999, was to develop and test the major instruments to be used in the experiment: 1) a reading passage and accompanying comprehension questions of approximately 30 minutes in length which contain the 10 treatment words and would be of interest to the Business Administration students in my department, 2) to develop activities for the other two conditions (activity-based and pushed output) which would require a similar amount of time for students to complete, and 3) to develop a "Yes-No" vocabulary pre-test for the ten treatment words.

1)   development of reading passage

The reading passage was initially piloted with 3 third year students from one of my colleagues classes. It took between 45 to 57 minutes for them to read the passage and complete the comprehension questions, and all three reported that the passage was a bit hard for them as there "were too many difficult words" and too many "long sentences".

The passage was then shortened and simplified by rewriting longer sentences into shorter ones, using simpler grammatical structures, and, where possible, omitting difficult, low frequency vocabulary words (excluding, of course, the target words), and replacing them with high frequency words that the students were more likely to know.

As can be seen in table two, when the activity was piloted several weeks later with 7 other students from the same class, the task took much less time, about 30 minutes, with all 7 students reporting that the task was not overly difficult for them.

2)   development of activity and writing tasks

Since I wanted to remove time on task as a confounding variable, 30 minutes then became the target for the other two conditions,. The first version of the vocabulary activity condition used 4 different tasks, but when piloted with 2 third year students, it was found that the tasks took approximately 38 minutes to complete, considerably longer than the goal of 30 minutes. The activity was then run again, after decreasing the number of tasks from four to three. The first attempt at three tasks resulted in a time of 33:38 (piloted with just one student), and on the second attempt, with a different combination of three tasks, a more acceptable time of 30:56 (again, with one student) was achieved. A similar result of 29:39 (see table 3) was obtained when the trial was run with more students in the "official" trial.

Table 3: Average time on task for pilot study

Condition	# students	Ave. time/task	Range
Reading	7	30:11	(+/-) 9:50
Activities	7	29:39	(+/-) 8:03
Writing (combined)	6	32:42	
Writing only	3	33:05	(+/-) 6:03
Writing w/example 3		31:20	(+/-) 2:45

For the writing condition, it was found that if only once sentence was required for each of the target words, only about 19 minutes was needed to complete the task. When the requirement was raised to two sentences for each target word, however, the amount of time needed to complete the task became very close to the other two conditions, with an average time of 32:42 (reported in table two).

3)   development of the yes-no pre-test

Although an unofficial pilot was run with several students in order to choose 10 target words from the reading passage that would be unknown to the population, it was felt that there would be less chance of difficulties later on if the yes-no test were given another trial run with a larger group of learners from a similar population Thus, the yes-no test was given to the same group of 20 students that the 3 types of tasks were piloted

on. The test was given before the learners received the tasks in order to gauge their knowledge of the words before any opportunities for learning took place.

Since it is desirable for the student's score on the post-treatment vocabulary quiz to be attributable entirely to the treatment (and not to previous knowledge of the words), the best possible result on the yes-no test would be that all 20 students reported that they knew none of the target words, and that they would also not report knowing any of the nonsense words (i.e. they were telling the truth on their yes/no answers).

Unfortunately, as can be seen in table 4, the results for the pilot were not this clear cut. While 11/20 students (55%) reported knowing none of the target words and none of the nonsense words, another 9 students reported knowing at least one or more of the nonsense words. If we allow for a margin of error and include students with only one or two false positives (the number of nonsense words they report knowing), the number of students who report knowing none of the target words grows to 15 students (75%).

Table 4: Number of target and "nonsense" words known by pilot students

	NONSENSE WORDS					
TARGET WORDS	0	1	2	3	4	5
0	11	4				
1		1				
2			1			
3		1				
4					1	
5						
6					1	
7						
8						
9						
10						

Although it is worrisome that one student reports knowing 6/10 of the target words, and another student 4/10, it can be argued, that their responses are unreliable, since it is likely that the extremely high number of false positives they have (4/5 or 80%), are the result of random guessing, or rushing through the exam and checking "yes" for every word.

If, then, we consider these students to be outliers that should be excluded from analysis, the results seem more positive and clear. 15/18 students (83%) know none of the target words, and 17/18 (94%) know two target words or less. This seems to indicate that the 10 target words are unknown to the large majority of the population studied,

strengthening the claim that much of their score on a post-treatment vocabulary quiz will be due to the effects of the treatment, rather than a prior knowledge of the words.

Experiment

The experiment, which took place over a two week period in January of 1999, was conducted as follows:

Detailed Overview of Experimental Procedures

Class 1:	(5 min.)	Explain experiment
	(5 min.)	Pass out Levels Test/give directions
	(45 min.)	Conduct Levels Test
	(5 min.)	Collect papers
	(30 min.)	Regular class activities

Class 2:	(5 min.)	Explain experiment
	(5 min.)	Pass out Yes-No Test/give directions
	(10 min.)	Conduct Yes-No test
	(5 min.)	Collect papers
	(10 min.)	Pass out new vocabulary & treatments/give directions
	(30 min.)	Conduct 4 treatments simultaneously
	(5 min.)	Collect papers
	(5 min.)	Pass out Post-test/give directions
	(10 min.)	Conduct Post-Test
	(5 min.)	Collect papers

At the beginning of each class, students were told that they were participating in a research experiment comparing different ways to teach and learn new English vocabulary words, and that their scores on any of the tests they were taking would not in any way affect their grade for the class.

During the first class, students were given the Vocabulary Levels Test. The test took approximately 45 minutes to complete (plus another 15 minutes for handing out papers, giving directions, and collecting them), which left approximately 30 minutes to devote to usual classroom activities.

During the second class, students were first given the Yes-No Vocabulary pre-test, which took approximately 20 minutes to complete (including the time needed to give out the papers, give directions and collect the papers).

Then, students were randomly given one of the four treatments, along with a list of the 10 new vocabulary words, with definitions given in both English and Japanese. In order to decrease the chance of copying, students sitting next to each other were given different treatments. The treatment tasks took approximately 30 minutes to complete (plus another 15 minutes for handing out papers, giving directions, and collecting them).

Immediately following the treatment, students were given the post-test vocabulary quiz, which took approximately 10 minutes to complete (plus another 10 minutes for handing out papers, giving directions, and collecting them).

Analysis

There were two independent variables in this study. One independent variable was task type, which had three levels; a reading-based activity, a task-based activity, and a writing-based activity. The second independent variable was the overall vocabulary size of the learner, which had two levels: large or small. The dependent variable in the study was learning of the target words, which was operationalized as student score on a multiple choice vocabulary quiz given directly after the treatment.

Although the original number of subjects for this experiment was 122, a number of students had to be excluded from the analysis due to being absent during the day that either the Levels Test or the treatment was given (n=10), this left a group of 112 subjects for the first round of statistical analysis.

(1) reliability of instruments

A series of KR-20 analyses was conducted on all major instruments used in this study, in order to establish their reliability.

For the Levels Test, the KR-20 was calculated for the 112 subjects in this experiment, which resulted in a fairly high reliability figure of .82. Although somewhat lower than the reliability figures reported for the Read (1988) and Beglar/Hunt (1999) studies (.94 and .97, respectively) mentioned in the materials section, it is possible that the more homogeneous nature of this population, that of Japanese EFL college students with similar majors, was a contributing factor to the lower figure of .82.

The Yes/No test developed for this experiment consisted of 50 words including the 10 target words, 35 distractor words, and 5 non-words. A KR-20 analysis resulted in an overall test reliability of .86, with a reliability of .73 for the 10 target words.

The reliability of the post-treatment vocabulary quiz was also calculated using a KR-20 analysis, with a resulting figure of .68.

(2) comparison of treatments

A one way analysis of variance (ANOVA) was carried out to compare the four treatments (learning conditions) with respect to the outcome variable (post-test score). As can be seen in Table 5, significant difference among means was found ($p = 0.021$).

Table 6 shows that treatment 3 (writing) had the highest mean number of words learned, followed by 2 (activities), 4 (writing with example sentence given), and then 1 (reading).

In Table 7, Fishers method for multiple comparisons revealed that treatment 1 (reading) was significantly different from treatments 2 (activities) and 3 (writing). No other significant pair-wise differences were detected.

Table 5: ANOVA results for comparison of the 4 learning conditions

ANOVA Table for Posttest

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Trtmt	3	35.484	11.828	3.390	.0207	10.171	.754
Residual	108	376.792	3.489				

Table 6: Mean table and interaction bar for comparison of the 4 learning conditions

Means Table for Posttest
Effect: Trtmt

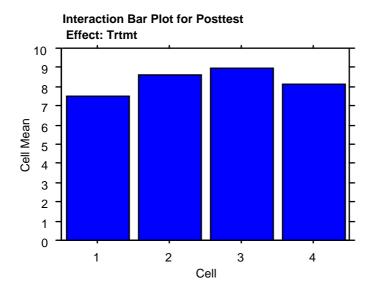|	Count	Mean	Std. Dev.	Std. Err.
1	38	7.500	2.368	.384
2	36	8.611	1.644	.274
3	19	8.947	1.311	.301
4	19	8.105	1.560	.358

Interaction Bar Plot for Posttest
Effect: Trtmt

Table 7: Fisher's results for comparison of the 4 learning conditions

Fisher's PLSD for Posttest
Effect: Trtmt
Significance Level: 5 %

	Mean Diff.	Crit. Diff	P-Value	
1, 2	-1.111	.861	.0119	S
1, 3	-1.447	1.040	.0068	S
1, 4	-.605	1.040	.2513	
2, 3	-.336	1.050	.5269	
2, 4	.506	1.050	.3417	
3, 4	.842	1.201	.1675	

To refine this comparison, two additional steps were taken. First, cases where it was suspected that students were guessing were deleted (n=3). As in the pilot study, guessing was defined as instances where three or more (out of five) false positives appeared on a student's Yes- Test. This left a group of 109 students. Second, the effect of learner's vocabulary size (created by categorizing their score on the Levels test), was evaluated using a two way analysis of variance (ANOVA).

This analysis (Table 8), showed no significant interaction between treatment effect and vocabulary size (p=.96), however, both treatment (p=.040) and vocabulary size (p=.0001) had significant main effects. That is to say, the lack of significant interaction shows that the relative performance of the treatment is independent of vocabulary size. As can be seen in the table and graph below (Table 9), learners with larger vocabulary sizes consistently learned more words in each of the three treatments than learners with smaller vocabulary sizes. As in the one way ANOVA, treatment 1 (reading) was significantly different from treatments 2 (activities) and 3 (writing).

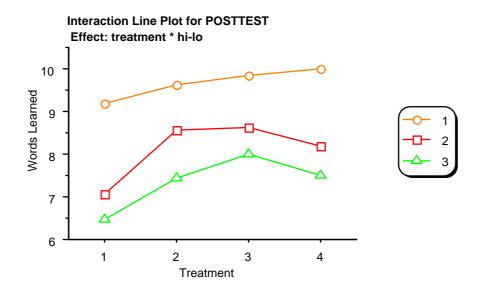Table 8: ANOVA results for comparison of the 4 learning conditions and vocabulary size

ANOVA Table for POSTTEST

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
treatment	3	25.995	8.665	2.917	.0381	8.750	.676
hi-lo	2	58.613	29.307	9.865	.0001	19.730	.989
treatment * hi-lo	6	4.173	.695	.234	.9644	1.405	.109
Residual	97	288.162	2.971				

Table 9: Fisher's results for comparison of the 4 learning conditions and vocabulary size

Fisher's PLSD for POSTTEST
Effect: hi-lo
Significance Level: 5 %

	Mean Diff.	Crit. Diff	P-Value	
1, 2	1.418	.805	.0007	S
1, 3	2.436	.899	<.0001	S
2, 3	1.018	.787	.0118	S

Interaction Line Plot for POSTTEST
Effect: treatment * hi-lo

Although it appears from the graph in Table Eight that treatment 4 has the highest mean for the high-level vocabulary group, the difference between treatment four and other groups was not found to be statistically significant.

Discussion

This study investigates the effectiveness of two teaching techniques that are commonly used in the ESL classroom to teach new vocabulary words, that is via reading-based and activity-based tasks, and compares them with a third approach, via writing-based tasks. The results were fairly clear cut.

Students learned significantly more new vocabulary words from the pushed-output technique (writing), than they did from either input or activity-based techniques, thus, the answers to research question 1 is affirmative.

A secondary finding was that regardless of the method used, it is possible in even a relatively brief period of time, for learners to make substantial gains in their understanding of new vocabulary words.

The results of this study are also of significance to SLA theory. Although other studies have argued that depth of processing tasks would necessarily require longer amounts of time to complete (for example, see Laufer & Hulstijn, 2001), this study has shown that even when time on task was controlled, the task which required the deepest level of processing (writing original sentences), significantly outperformed the other conditions.

21

Results thus support Craik and Lockhart's (1972) *Depth of Processing Theory* for the population examined.

Despite the body of literature which suggests that vocabulary size is an important aspect of reading comprehension and academic performance (Hirsch & Nation, 1992, Laufer & Sim, 1985, Marshall & Gilmour 1993, Laufer 1992a,b), this study showed no significant differential effect for lexical size as to the effectiveness of the three techniques compared. Indirect evidence for the importance of developing a large vocabulary can be seen, however, in the fact that learners in the study with larger vocabulary sizes consistently learned more words in each of the four treatments than learners with smaller vocabulary sizes. That is to say, the larger a student's vocabulary size, the more words they are likely to learn irrespective of the technique used.

Another significant finding of this study relates to the amount of words students were able to learn in the input-based technique, the least effective of the 4 techniques compared. In a study by Zahar, Cobb and Spada (2001), the authors claim that, based on both their own research, as well as a review of other L2 vocabulary reading studies, students can only be expected to learn an average of one out of 12-14 words that are encountered in typical reading tasks, with higher level learners (as measured by their score on the Levels Test) learning slightly more words than lower level learners.

In this experiment, however, students in the reading condition learned significantly more than Zahar, Cobb and Spada have hypothesized, with an average of 5/10 unknown target words correctly identified on the post treatment vocabulary quiz. The main difference, of course, is that the studies reviewed by Zahar, Cobb and Spada were focused on incidental learning, whereas in this experiment the learning was much more explicit (since the definitions of the words were given out, and were underlined in the passage). The higher rate of learning then, may be further evidence that increases in attention leads to greater depth of processing and better retention of vocabulary words.

Another interesting finding in this experiment is with regard to the superior retention of words displayed in condition 4 (writing new words in sentences without example sentences). Although Swain has argued that pushed output not only needs to convey a message, but that it needs to be conveyed precisely, coherently, and appropriately, students in condition 4 achieved higher scores on their post tests (than the other 3 conditions) irrespective of whether the sentences they wrote were coherent, appropriate, or even correctly written. This held true when students made grammatical errors, spelling errors, errors in usage, and even in cases where students left whole sentences blank. It may be that the act of attempting to write new words in original sentences required a level of processing so significantly deeper than the other conditions, that more information about the new words was retained even when the sentences were written incorrectly.

Pedagogic Recommendations

Although SLA research findings usually have classroom applications, they are not likely to be adapted by most teachers if the suggestions are time consuming ones that require a great deal of commitment, special tools (such as computers, video cameras, VCRs etc.), or extra training. Most teachers are extremely busy people with little extra time in their schedule. Fortunately, one of the most obvious implications of this study,

that assigning students the task of writing original sentences for new vocabulary words is more effective than developing controlled readings or word activity sheets for them, is actually an implication which greatly decreases the amount of preparation time needed when compared with more traditional approaches.

Another pedagogic implication of this study relates to the use of extensive reading as a means of increasing vocabulary size. As mentioned in the discussion section, students learned significantly more words in the reading condition than would be predicted by in the literature. This suggests that there may be pedagogic value in underlining, defining, or otherwise drawing attention to target words in extensive reading passages. For example, if vocabulary learning were a secondary goal of a reading class, the results of this experiment suggest that a teacher would be able to greatly increase the chance that a word would be learned simply by drawing learners' attention to it through simple techniques such as glossing, or underlining, again, tasks which don't greatly increase the amount of preparation time needed.

Limitations and Suggestions for Further Study

Several of the limitations to this study are ones common in the literature; the need for a larger n-size, the need to conduct similar experiments with different populations and proficiency levels, and the need to conduct delayed post-treatment tests (in order to look at long term retention rates for each of the three treatments).

One limitation specific to this experiment is the possibility that a "ceiling effect" may have hid significant differences between the groups and a possible differential effect for vocabulary size. In other words, because so many students learned such a high number of the 10 treatment words in each of the four treatments, the results tended to all cluster on the high end of the scale, making it difficult detect possible significant differences of treatment effects between the three groups. A subsequent experiment with a larger number of words, and more challenging tasks might reveal more significant results.

Since the current experiment only looked at minor gains in vocabulary knowledge (passive retention) over a short period of time, one suggestion for a follow-up study be to conduct a longer-term experiment to track gains in the depth of vocabulary knowledge acquired over time by each of the methods. The preliminary results of the present research would predict that pushed-output techniques should lead students to greater gains in depth of knowledge of individual words over time, but that an intensive reading-based approached could lead to the learning of a greater number of words, albeit at a shallower depth.

Conclusion

Until now, most evidence supporting Swain's (1985) *Pushed Output Hypothesis* has been qualitative in nature. This study has added quantitative knowledge to our understanding of how pushed output contributes to the acquisition of new vocabulary words in an L2. It has also provided further support for Craik and Lockhart's (1972) *Depth of Processing Theory,* by providing evidence that learning techniques which require deeper levels of processing lead to better learning, even when time on task is held constant.

In terms of pedagogy, it was found that the most effective technique for helping students to learn the meaning of new vocabulary words is also one that requires very little extra preparation time - asking students to use the new words generatively, in original sentences. For teachers who prefer input-based methods such as extensive reading, it was also found that simple techniques such as underlining, highlighting, and glossing the target words can greatly enhance the chance that those words will be learned.

Bibliography

Allen, F.A. (1983). *Techniques in Teaching Vocabulary.* New York: Oxford University Press.

Allen, P., Swain, M., Harley, B, & Cummins, J. (1990). Aspects of classroom treatment: towards a more comprehensive view of second language education, in B. Harley, P. Allen, J. Cummins, and M. Swain (eds.). *The Development of Second Language Proficiency.* Cambridge: Cambridge University Press.

Anderson, R.C. and Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Huston (ed.) *Advanced in Reading/Language Research*. Volume 2. Greenwich, CT:JAI Press.

Baddeley, A.D. (1966). Short term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18, 362-365.*

Bear, R.M. & Odbert, H.S. (1941). Insight of older pupils into their knowledge of word meanings. *School Review,* 49, 754-760.

Beck, I., McKeown, M. & Omanson, R. (1987). The effects and uses of diverse vocabulary instructional techniques. In M. MsKeown & M. Curtis (eds.) *The Nature of Vocabulary Acquisition (*pp. 147-163). Hillsdale, NJ: Erlbaum.

Beglar, D. & Hunt, A. (1999). Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing,* 16 (2), 131-162.

Broadbent, D.E. (1958). *Perception and Communication*. New York: Pergamon Press.

Broukal, M. (2001). *Practice Vocabulary.* Boston: Heinle & Heinle.

Brown, C. (1985). Requests for specific language input: Differences between older and younger adult language learners. In S. Gass & C. Madden (eds.), *Input in Second Language Acquisition* (pp. 272-284). Rowley, M.A.: Newbury House.

Brown, T. & Perry, F. (1991). A comparison of three learning strategies for ESL vocabulary acquisition. *TESOL Quarterly,* 25, 655-671.

Browne, C. (1998). Japanese high school textbooks: How readable are they? *Working Papers in Applied Linguistics*, Temple University Japan, 12:1-13.

Chall, J. (1987). Two vocabularies for reading: recognition and meaning. In M.G. McKeown & M.E. Curtis (Eds.), *The Nature of Vocabulary Acquisition.* Mahwah, N.J.: Lawrence Erlbaum Associates.

Corder, S.P. (1981). *Error Analysis and Interlanguage*. Oxford: Oxford University Press.

Craik, F.I.M. & Lockhart, R.S. (172). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Learning Behavior, 11, 671-684.*

Craik, F.I.M. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268-294.

D'Anna, A., Zechmeister, E., & Hall, J (1991). Towards a meaningful definition of vocabulary size. *Journal of Reading Behavior*, 23, 109-122.

Engels, L.K. (1969). The fallacy of word counts. *IRAL,* 6, 213-231.

Ellis, R. (1985). *Second Language Acquisition and Language Pedagogy.* Avon, England: Multilingual Matters.

Ellis, R. (1990). *Instructed Second Language Acquisition.* Oxford: Blackwell.

Elley, W.B. & Mangubai, F. (1981). *The impact of a book flood in Fiji primary schools*. Wellington: New Zealand Council for Educational Research.

Elley, W.B. & Mangubai, F. (1983). The impact of reading on second language learning. *Reading Research Quarterly*, 19, 53-67.

Elley, W.B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, 24, 174-187.

Elley, W.B. (1991). Acquiring literacy in a second language: the effect of book-based programs. *Language Learning, 41, 375-411.*

Grabe, W. & F. L. Stoller (1997). Reading and vocabulary development in a second language. In J. Coady & T. Huckin (Eds.) Second Language Vocabulary Acquisition. Cambridge: Cambridge University Press.

Gregg, K. (1984). Krashen's monitor and Occam's razor. *Applied Linguistics,* 5: 79-100.

Hafiz, F.M. & I. Tudor. (1989). Extensive reading and the development of language skills. *ELT Journal*, 34 (1), 5-13.

Hall, S.J. (1992). Using split information tasks to learn mathematics vocabulary. *Guidelines*, 14, 72-77.

Hatch, E. (1983). *Psycholinguistics: A Second Language Perspective.* Rowley, MA: Newbury House.

Hatch E., Flashner, V., & Hunt, L. (1986). The experience model and language teaching. In Day, R. (ed.), *Talking to Learn (pp. 5-22).* Rowley, MA: Newbury House.

Heath, S. (1984). *Ways With Words.* Cambridge: Cambridge University Press.

Hirsh, D. & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language,* 8(2), 689-696.

Horst, M. T. Cobb & P. Meara (1998), Beyond A Clockwork Orange: acquiring second language vocabulary through reading. *Reading in a Foreign Language,* 11, 207-223.

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. *Modern Language Journal, 80, 327-339.*

Johansson, S.,, & Hofland, K. (1989). *Frequency analysis of English vocabulary and grammar.* Oxford: Clarendon Press.

Krashen, S. (1985). *The Input Hypothesis*. London: Longman.

Krashen, S. and Terrel, T. (1983). *The Natural Approach.* New York: Pergamon Press.

Krashen, S. (1989). We acquire vocabulary and spelling via reading: Additional evidence for the input hypothesis. *Modern Language Journal*, 73(4), 440-464.

Krashen, S. (1993a). The case for free voluntary reading. *Canadian Modern Language Review, 50 (1), 72-82.*

Krashen, S. (1993b). *The Power of Reading.* Ingllewood, CA. Libraries Unlimited.

Keen, D. (1994). *Developing Vocabulary Skills*. Boston: Heinle & Heinle.

Kucera, H. (1982). The mathematics of language. In *The American Heritage Dictionary,* (pp. 37-41). Boston: Houghton Mifflin.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.),
Special language: From human thinking to thinking machines. Clevdon, England: Multilingual Matters.

Laufer, B. (1992a). How much lexis is necessary for reading comprehension? In H.B. Joint & P. Arnaud (Eds.), *Vocabulary and Applied Linguistics*. London: Macmillan.

Laufer, B. (1992b). Reading in a foreign language: How does lexical knowledge interact with the reader's general academic ability? *Journal of Research in Reading,* 15, 95-103.

Laufer, B., & Sim, D.D. (1985). Taking the easy way out: Non-use and mis-use of clues in EFL reading. *Forum*, 23(2), 7-10.

Laufer, B. & Hulsijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22 (1), 1-26.

Liu, N. A. & Nation, P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, 16, 33-42.

Long, M. (1980). Inside the 'black box': Methodological issues in classroom research on language learning. *Language Learning*, 30(1): 1-42.

Long, M. (1983a). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics,* 4(2), 126-141.

Long, M. (1983b). Does second language instruction make a difference? A review of research. *TESOL Quarterly*, 17(3), 359-382.

Long, M. (1985). Input and second language acquisition theory. In S. Gass & C. Madden (eds.), *Input in Second Language Acquisition* (pp. 377-393). Rowley, M.A.: Newbury House.

Marshal, S. and Gilmour, M. (1993). Lexical knowledge and reading comprehension in Papua New Guinea*. English for Specific Purposes*, 12, 69-81.

McCarthy, M., F. O'Dell, & E. Shaw (2001). *Basic Vocabulary in Use*. Cambridge: Cambridge University Press.

McCarthy, M., F. O'Dell, & E. Shaw (1997). *Vocabulary in Use: Upper Intermediate*. Cambridge: Cambridge University Press.

Meara, P. (1980). Vocabulary acquisition: A neglected part of language learning. In V. Vinsella (ed.), *Surveys I* (pp.100-126). New York: Cambridge University Press.

Meara, P. (1990). Some notes on EuroCentres vocabulary tests, AFinLA Yearbook, 1990, 48, 103-113.

Meara, P. (1991). Scoring a yes/no vocabulary test, unpublished paper, http://www.swan.ac.uk/cals/.

Meara, P. & Buxton, B. (1987). An alternative to multiple choice vocabulary tests, *Language Testing*, 4,142-151.

Nation, I.S.P (1990). *Teaching and Learning Vocabulary*. Heinle and Heinle. Boston: MA.

Nation, I.S.P. (1994). *New Ways in Teaching Vocabulary*. TESOL: Alexandria, VA.

Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. New York: Cambridge University Press.

Newton, J. (1995). Task-based interaction and incidental vocabulary learning: a case study. *Second Language Research,* 11 (2), 159-177.

Nobuyoshi, J. & Ellis, R. (1993). Focused communication tasks and second language acquisition. *ELT Journal*, 47, 203-210.

Nuttall, C. (1982). *Teaching Reading Skills in a Foreign Language*. London: Heinemann.

Nuttall, C. (1996). *Teaching Reading Skills in a Foreign Language (2nd Edition).* London: Heinemann.

Paribakht, T. & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady and T Huckin (Eds.) *Second Language Vocabulary Acquisition.* Cambridge: Cambridge University Press.

Paribakht, T. & Wesche, M. (1993). The relationship between reading comprehension and second language development in a comprehension-based ESL program, *TESL Canada Journal,* 11(1), 9-29.

Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal*, 19 (2), 12-25.

Read, J. (1997). Vocabulary and Testing (in Schmitt & McCarthy)

Redman, S. & E. Shaw (1999). *Vocabulary in Use: Intermediate*. Cambridge: Cambridge University Press.

Redman, S. (1997). *English Vocabulary in Use: Pre-Intermediate and Intermediate*. Cambridge: Cambridge University Press.

Robb, T. & Susser, B. (1989) Extensive reading vs. skill building in an EFL context. *Reading in a Foreign Language*, 5 (2), 239-251.

Saragi, T., Nation, P., and Meister, G. (1978). Vocabulary learning and reading. *System*, 6, 72-80.

Schmitt & M. McCarthy (Eds.) *Vocabulary: Description, acquisition and pedagogy.* Cambridge: Cambridge University Press.

Seal, B. (1997). *American Vocabulary Builder 1*. New York: Addison-Wesley.

Seal, B. (1990). *American Vocabulary Builder 2*. New York: Addison-Wesley.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics,* 10: 209-231.

Sharwood Smith, M. (1986). Comprehension vs. acquisition: two ways of processing input. *Applied Linguistics*, 7: 239-256.

Sims, V.M. (1929). The reliability and validity of four kinds of vocabulary tests. *Journal of Educational Research,* 20:91-96.

Smith, F. (1985). *Reading Without Nonsense*. New York: Teachers College Press.

Sokmen, A. (1997). Current trends in teaching second language vocabulary. In N. Schmitt and M. McCarthy (eds.), *Vocabulary: Description, Acquisition and Pedagogy.* Cambridge: Cambridge University Press.

Swain, M. (1981). Target language use in the wider environment as a factor in its acquisition. In Anderson (ed.), *New Dimensions in Second Language Acquisition Research*. Rowlet, MA: Newbury House.

Swain, M. (1985). Communicative competence: some roles for comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (eds.) *Input and Second Language Acquisition.* Rowley, MA: Newbury House.

Swain, M. (1995a). *Collaborative Dialogue: Its Contribution to Second Language Learning.* Plenary paper presented at AAAL Conference, Long Beach, CA.

Swain, M. (1995b). Three functions of output in second language learning. In G. Cook and B. Seildhofer (eds.), *Principle and Practice in Applied Linguistics.* Cambridge: Cambridge University Press.

Swain, M. & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards language learning. *Applied Linguistics*, 16(3), 371-391.

Taylor, L. (1992.). *Vocabulary in Action*. London: Prentice Hall.

Thorndike, E., & Lorge, I. (1944). *The Teachers' Word Book of 30,000 Words*. NY: Columbia Teachers College.

Watanabe, Y. (1997). Input, intake and retention: effects of increased processing on incidental learning of foreign vocabulary. *Studies in Second Language Acquisition, 19, 287-307.*

Waugh, N.C., and Norman, D. A. (1965). A primary memory. *Psychological Bulletin*, 72, 89-104.

West, M. (1953). *A General Service List of English Words.* London: Longman.

Zahar, R., Cobb, T., and Spada, N. (2001). Acquiring vocabulary through reading: effects of frequency and contextual richness. *The Canadian Modern Language Journal,* 57 (4), 540-572.

Zimmerman, C.B. (1997a). Historical trends in second language vocabulary instruction. In J. Coady and T Huckin (Eds.) *Second Language Vocabulary Acquisition.* Cambridge: Cambridge University Press.

Zimmerman, C.B. (1997b). Do reading and interactive vocabulary instruction make a difference? An empirical study. *TESOL Quarterly,* 31, 121-141.