Abstract
This paper reports on an initial validation of a new productive vocabulary test that measures depth of vocabulary using multiple contextualized examples of sentences. The participants in the study were two intact writing classes, one higher ability, the other lower ability, of L1 (n = 3) and advanced L2-EFL learners (n = 27) in first- and second-year university, comprised of 11 males and 19 females, all with previous experience living and studying in English-speaking, or other foreign-language-speaking environments. The Vocabulary Size Test (Nation & Beglar, 2007) and a new measure of vocabulary depth, the 1K-VDT, described herein, were used to measure vocabulary. Findings revealed that the higher ability group scored significantly higher than the lower ability group on both vocabulary measures. A large correlation between scores from both tests for all learners was also found. Future research is still needed; however, there was evidence that the 1K-VDT may require a threshold of vocabulary size. Implications for EFL learners and teachers are discussed.

Key words: vocabulary, breadth, depth, testing, validation

The role and importance of vocabulary in an L2 setting has been shown to be an important factor related to comprehension (Laufer, 1997). In receptive tasks, Qian (1999) reported that breadth, depth, and reading comprehension were highly correlated; and Stæhr (2009) found a positive linear relationship between breadth and listening comprehension. In production tasks, direct correlations were seen between breadth and output for writers (Laufer & Nation, 1995) and speakers (Ovtcharov, Cobb, & Halter, 2006). In settings where entire programs or faculties are offered in an L2, such as the Faculty of Liberal Arts at Sophia University, the importance of a learner’s vocabulary knowledge cannot be underestimated. In such settings, Laufer and Yano (2001) highlighted the burden that L2-speakers face - having a smaller vocabulary than L1-peers, and or, a smaller vocabulary than required to read authentic text.

In view of the importance of vocabulary, Hunt and Beglar (2005) presented a framework for developing vocabulary in EFL settings which included tasks that developed both vocabulary breadth and depth. Vocabulary breadth refers to the quantity of vocabulary items known by an individual. Knowing an item means more than its knowing its meaning (concepts, referents, associations), but also its form (spelling, pronunciation, word parts) and use (functions, collocations, constraints)
Thus, depth refers to the quality of that vocabulary knowledge. This distinction, though, between breadth and depth is not held by all researchers. Vermeer (2001), for example, argued that breadth and depth are two dimensions of the same factor, input.

Word coverage and vocabulary breadth and depth

Does size matter? Nation (2001) argued that the most frequent one-thousand words (1K), based on West’s General Service List (1953) cover more than 70% of both written and spoken text in English, including academic text, and that the most frequent two-thousand (2K) words combine to cover approximately 80% of such text. The nearly 600 words of Academic English (AWL) (Coxhead, 2000) add an additional 10% of coverage in academic texts. Taken as a whole, these three word lists of nearly 2600 words cover approximately 90% of academic text. However, researchers now believe learners need to know 98% of the vocabulary in a text, at a minimum, to be able to read comfortably (Nation & Webb, 2011). Several studies have discussed the connection between 98% word coverage and different text. Nation (2006), for example, found that between 6,000-7,000 words were necessary to provide 98% coverage in daily listening, and between 8,000-9,000 words for pleasure reading. Webb & Rodgers (2009a, 2009b) found between 5,000-9,000 words were necessary to provide 98% coverage for television, and between 6,000-7,000 words for movies. Nation & Webb (2011) argued that to read academic text comfortably in the L2, learners may need a vocabulary size as large as 20,000 words.

A number of measures are now available to estimate vocabulary breadth, including the vocabulary levels test for diagnostic purposes (VLT) (Nation, 1990), the vocabulary size test (VST) as a general measure of breadth (Nation & Beglar, 2007), and the Eurocentres vocabulary size test (EVST) (Meara and Jones, 1990). Other measures include the written vocabulary levels test (VLT) (Laufer and Nation, 1999), the word associates test (Read, 1998), a yes-no test (Meara, 1992), and CATSS, a measure of size and strength (Laufer, Elder, Hill, & Congdon, 2004). Several examples of tests to measure vocabulary depth also exist. These include the depth of vocabulary knowledge (DVK) (Read, 1993), the depth of vocabulary knowledge test (DVKT) (Qian, 2002), and the vocabulary knowledge scale (VKS) (Paribakht and Wesche, 1993). Vocabulary met by learners in academic settings is usually in contextualized situations; unknown items embedded in a lecture or reading, for example. However, the tests described above typically present items in decontextualized situations. In the following example from the VST, the word threshold is presented in a single sentence without surrounding context: They raised the threshold. It is likely, therefore, that questions are raised concerning the ecological validity of tests of this type.

Any test that does accurately measure vocabulary depth likely correlates with a test of vocabulary breadth. Indeed, the VKS has been shown to correlate
moderately with the EVST, $r = .53$, (Wesche and Paribakht, 1996). This moderate correlation may be related to input - greater input increases the number of meetings with higher frequency items (Vermeer, 2001). Nadarajan (2008) linked depth with input by showing the DVKT successfully grouped learners by predicted ability (low, intermediate, high, and native). Thus, one possible theory of second language vocabulary learning is that learners with a broader vocabulary also have a deeper vocabulary.

An important issue for learners is polysemy, or multiple meanings of a word. Ruhl (1989) argued for the monosemic bias - that English words have one fundamental meaning, and it is the surrounding words in sentences that impart more meaning than the word itself. As an example, “the thief took the jewels” is understood to mean “the thief stole the jewels” because of our knowledge of what thieves do with jewels, and not because the word take means steal, and thus, meaning may often be more than the sum of the parts of a sentence. One problem with this explanation is that dictionaries, a tool frequently used by our learners, present words as often having multiple semantic meanings (take, above for example). “[D]efinitions [in dictionaries] confuse inherent lexical meaning (semantics) with meaning provided by context (semantic and pragmatic)” (Ruhl, 1989, p. 234). It remains to be seen if most English words are monosemic - Ruhl only argued for the monosemic bias of just a handful of words. Parent (2009) and Nation and Webb (2011) also argued that we should take the monosemic bias. If, however, the dictionary is used as our model, we are obliged take the polysemic-bias, and it is likely that the most frequent words are those with the most multiple meanings. As more than 70% of text is composed of 1K words, the more a learner seeks input, the greater the opportunities to meet polysemic words.

The main goal of this paper is to begin validation for a new vocabulary depth test that presents items in contextualized, meaningful bits of language. This new test is tentatively titled the 1K-VDT (One Thousand-Word List - Vocabulary Depth Test). To begin validation of this test, scores on a measure of vocabulary breadth, the VST, will be compared with those scores from the 1K-VDT. Work, yet to be done, to fully validate this new measure, will include a Rasch analysis of items and person ability estimates.

Four hypotheses have been formulated. The first three hypotheses state that group differences for two classes, higher and lower ability, on both (1) a vocabulary breadth test and (2) a vocabulary depth test will be significantly different; and (3) that the group with higher averages on the receptive breadth test will score higher on a vocabulary depth test. The fourth hypothesis states that individual scores from both the breadth and depth tests will be moderately correlated.

Method

Participants

This study reported here is part of a larger study involving 350 students at five
different universities, both public and private, in the Kanto region. For this paper, two convenient samples, intact writing classes I taught from April - July, 2010, in the Faculty of Liberal Arts at Sophia University participated in this study. McKinley (2010) reported that students in this faculty are generally considered to have a much higher command of English than most university students in Japan. The two groups had been streamed by ability, into a lower-level (Lower) or a higher-level (Higher) class. The two classes were separated by an intermediate group that did not take part in this study. In the Lower group, there were 13 participants, 8 females and 5 males. Twelve of these participants shared Japanese as their L1, and one participant was an L1-speaker of another Asian language. All 13 participants have had experience living in English-speaking and other language-speaking countries, from several weeks to several years for one participant. The Higher group was composed of 25 members, 7 males and 18 females. Most members of this class shared Japanese as their L1, several were native speakers of English, and one member was an L1-speaker of another European language. Many members in this class had long-term experience living in English-speaking environments, or attended English-speaking schools in other language-speaking countries.

Materials

The VST

The VST test, is composed of 10 items at each one thousand-word level from 1,000- (1K) to the 14,000- (14K) word-level (Nation and Beglar, 2007; Beglar, 2010). For this test, all ten items at each of the first twelve levels, 1K through 12K, were tested, for a total of 120 items. On the VST, each item is presented in a decontextualized sentence, with four possible definitions, of which one answer is correct, and three are distractors. Below is an example item from the VST, level 6K (answer = b):

THRESHOLD: They raised the **threshold** (bold in the original).
- a.flag
- b.point or line where something changes
- c.roof inside a building
- d.cost of borrowing money

The 1K-VDT

Unlike other depth of vocabulary tests or scales, the 1K-VDT is a productive measure of vocabulary depth. In making the 1K-VDT, the items on the 1K word list were input alphabetically into a Numbers '09, version 2.0.4 file (Apple software) in one column. In the next column, a list of corresponding numbers from 1 through 967 (the actual number of items on the 1K word list) was input in ascending order, so that the item a corresponded to number 1, **able** to number 2, **about** to number 3, and so on through **youth** to 967. A random integer generator (Haahr, 2010) was then used to choose 840 numbers (and their corresponding 1K item). These 840 chosen numbers and items were then input into a new file, in the order they
were randomly chosen (item 776, 257, 568, 52, ...), the numbers were deleted, and re-numbered 1 through 840 in ascending order, and the random item generation was repeated to choose 720 items, then 600, 480, 360, 240, and finally 120 items. More than 100 items were chosen in a final list in case items in the first 100 would make poor test items. For example, it was predicted, but not tested, that months and days of the week would likely be poor test items because responses (e.g., Monday, Tuesday, February, March) would be interchangeable. Randomization of items, thus, occurred seven times, each time with a smaller number of items drawn from a shrinking pool of items. A series of six different sentences for each item was written. See Appendix for an example of one test with ten items. Sentences were modified versions of examples found in a popular English dictionary, Collins Cobuilds, Advanced Learners, (Sinclair, 2003). The first six example sentences in the dictionary for each randomly selected item were chosen, then adapted. In this way, sample items included different grammatical structures, affixes, and collocations, and possibly common expressions. If a chosen item had fewer than six different examples in the dictionary, the process of writing sample sentences was repeated starting with example one. Once each six-sentence item was written, a vocabulary profiler (Cobb, n.d.) was used to identify words outside the 1K range. Words from the 2K, AWL and off-word lists were replaced with more frequent, 1K items. As a result, vocabulary coverage on each test is composed of more than 96% of items from the 1K list. When proper nouns and names are included, coverage nears 98% or higher. To test whether the items generated were comprehensible to native or near-native English users, a group comprised of six native speakers of English and fourteen highly proficient L2-speakers of English, all members of a TESOL doctoral program, were given a test of twenty randomly selected sample items from the 1K-VDT, $m = 18.95$ (maximum 20), $SD = 1.50$. One participant, an L3-speaker of English scored a low of 14 (70%). Five of the native speakers scored a perfect score, while the sixth native speaker had left one item blank. Five of the non-native speakers also scored a perfect score.

In this study, on the 1K-VDT, learners were given a series of ten items per test, divided into two equal sections of five items each, over a series of ten weeks. Learners were able to choose either section, and were thus given the opportunity to choose a section of the test which they believed they could best answer. Below is an example 1K-VDT item (answer: *just*):

i. It’s _____ a small cut.
ii. My house is _____ around the corner.
iii. It’s _____ the right thing to do.
iv. We _____ arrived.
v. _____ do it.
vi. _____ the two of us.

**Scoring**

The original VST has a maximum score of 140 (10 items on 14 lists); however,
only the first 12 lists were used (maximum 120). No points were deducted for incorrect answers. For the 1K-VDT, learners completed a total of 50 items. No points were deducted for incorrect answers. Certain spelling errors were considered acceptable. For the item above, for example, one learner wrote jast, instead of just, and this was accepted.

Procedure
The VST was given over two days in mid July, 2010; the 1K-8K on a Thursday, and the 9K-12K on the following Monday, both during regular class hours. Day 1 of the VST took less than 20 minutes to complete, and day 2 took less than 10 minutes to complete. The 1K-VDT was given over a period of ten weeks, one test per week, beginning in mid-April. Each 1K-VDT took approximately 5-10 minutes to complete, with more advanced learners finishing more quickly. It should also be noted that results from the 1K-VDT made up 5% of the learners semester grade.

Results
In the Higher group, the class began with twenty-five students; however, five students, four females and one male did not complete either one or the other or both sections of the VST. Scores from this group will include the twenty participants, six males and fourteen females, who completed the ten 1K-VDTs and both sections of the VST. In the Lower group, three female students did not complete either one or the other or both sections of the VST. Scores from this group will include the ten participants, five females and five males, who completed the 1K-VDT and both sections of the VST. In total, thirty students, eleven males and nineteen females in two groups, completed all ten of the 1K-VDT and both sections of the VST. For all 30 students, on the VST, $m = 84.60$, $SD = 11.41$; and for the 1K-VDT, $m = 43.61$, $SD = 3.47$.

Hypothesis 1: VST average scores for the two groups will be significantly different
The mean and standard deviations were calculated for these two groups for the VST, Higher ($n = 20$), $m = 88.30$, $SD = 11.76$; Lower ($n = 10$), $m = 77.20$, $SD = 5.99$. An unpaired t-test was run to test whether these groups were different as hypothesized. The test was significant, that is participants in the Higher group scored on average higher than those in the Lower group $t (28) = 2.79$, $p < 0.001$, $r = .51$. According to Field (2009), an effect size from 0.50 is considered large, thus the resulting effect size for the VST is above the threshold for a large effect size.

Hypothesis 2: 1K-VDT average scores for the two groups will be significantly different
The mean and standard deviations were calculated for each of these groups for the 1K-VDT, Higher ($n = 20$), $m = 44.65$, $SD = 4.39$; Lower ($n = 10$), $m = 41.23$, $SD = 3.30$. An unpaired t-test was run to test whether these groups were different as hypothesized. The test was significant; that is participants in the Higher group scored on average higher than those in the Lower group $t (28) = 2.17$, $p < 0.05$, $r =$
According to Field (2009), an effect size from 0.30 - .49 is considered moderate. **Hypothesis 3**: the group with the higher score on the VST will be the same group which scores higher on the 1K-VDT.

The results show that the higher scoring group on the VST was, as predicted, the higher scoring group on the 1K-VDT. Note that the lowest performing students on both the VST and the 1K-VDT came from the higher-level group, not lower-level group of students. On the VST, two students from the higher-level group had lower scores, 67 and 70, than the lowest scorer in the Lower group, 72. Similarly, one student in the Higher group had a lower score, 32, than the lowest scorer in the Lower group, 37. See Table 1.

<table>
<thead>
<tr>
<th>Test</th>
<th>Higher (n=20)</th>
<th>Lower (n=10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VST</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(m = 88.30)</td>
<td>(m = 77.20)</td>
<td></td>
</tr>
<tr>
<td>(SD = 11.76)</td>
<td>(SD = 5.99)</td>
<td></td>
</tr>
<tr>
<td>(max = 107)</td>
<td>(max = 91)</td>
<td></td>
</tr>
<tr>
<td>(min = 67)</td>
<td>(min = 72)</td>
<td></td>
</tr>
<tr>
<td>VDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(m = 44.65)</td>
<td>(m = 41.23)</td>
<td></td>
</tr>
<tr>
<td>(SD = 4.39)</td>
<td>(SD = 3.30)</td>
<td></td>
</tr>
<tr>
<td>(max = 49)</td>
<td>(max = 47)</td>
<td></td>
</tr>
<tr>
<td>(min = 32)</td>
<td>(min = 37)</td>
<td></td>
</tr>
</tbody>
</table>

**Hypothesis 4**: Scores on both the VST and the 1K-VDT will be moderately correlated.

A correlation test was run using all participants (\(n = 30\)) from both the Higher and Lower groups. There was a significant relationship between individual scores on the VST and the 1K-VDT, \(r = .74\), \(p = <.0001\), \(R^2 = .55\). This is a strong correlation.

**Discussion**

This paper had four hypotheses: that two Higher and Lower classes of learners would score significantly different on (1) a measure of vocabulary size, and (2) a new measure of vocabulary breadth; (3) that the group with the higher score on the measure of size would also be higher on the measure of depth; and (4) that scores on these two tests for all students would be moderately correlated. Indeed, the students in the higher-able class scored higher on both the VST and the 1K-VDT, and the group differences were shown to be significant, with moderate to large effect sizes. It should not be surprising that students in the higher group scored higher on the VST as this test has been previously shown to be reliable and valid (Beglar, 2010). However, the 1K-VDT had yet to be empirically tested, and as a result, its reliability was unknown. It was predicted that the higher VST scoring group would also
score higher on the 1K-VDT. We saw evidence for this with statistically significant differences between both groups and their scores on both tests. The main goal of this research, as stated above, was to begin validation of this 1K-VDT. For this small group of participants, scores on the 1K-VDT have been shown to be strongly correlated to those of the VST. Note that this correlation $r = .74$ is much greater than moderate correlation predicted in hypothesis four, which had been based on $r = .53$ between the VKT and the EVST (Wesche and Paribakht, 1996).

This paper is part of a larger project involving 350 students studying at five different universities in the Kanto region. Each of these universities differs considerably by hensachi, Standard t-scores. At one of the universities, University E, that with the lowest Standard t-score among the five involved in this study, four classes of learners completed the 1K-8K VST (maximum 80 items), $n = 96$, $m = 23.93$, $SD = 6.19$. It should be noted that this average or 23.93 is similar to a 25.00% score that could be achieved by randomly guessing. This average is also 60.67 points below the average of the 30 learners from the Faculty of Liberal Arts at Sophia University that are described in this paper, although only the 1K-8K were given to learners at University E. On two separate occasions in all four classes at University E, two versions of the 1K-VDT were distributed to the students, and on both occasions the students appeared frustrated and anxious because they were unable to complete the test - on both occasions, the average score for all students ($n=96$) was less than 1/5; and thus this project was abandoned at University E. This, however, may be evidence of what both Vermeer (2001) and Nation (personal communication) suggested; that breadth is a reflection of input, and that more input leads to greater depth. We may now be able to answer the question posed in the title of this paper: does size matter? Learners at University E had a small-sized vocabulary as measured by the VST, have had perhaps little (comprehensible) input, and were unable to perform on the 1K-VDT. Without a sizable quantity of known vocabulary items, learners are likely to have little depth of vocabulary. Thus, size does matter.

**Conclusion**

To sum up, the main goal of this paper was to begin validation for a new measure of vocabulary depth, the 1K-VDT. To achieve this goal, this paper had four hypotheses. Evidence appears to have been found which support all four hypotheses, although the correlation found in hypothesis four was stronger than expected.

It was also discussed that a threshold level of vocabulary breadth may be necessary to measure vocabulary depth. We must caution, however, that the two tests, the VST and 1K-VDT, are dissimilar. The former is a reading test which requires learners to match a word through active recognition of form with meaning from among four choices, one correct with three distractors; the latter is a productive reading and writing test which requires learners to deduct the answer from clues presented as six sample sentences.
Knowing the important role that items from the 1K-list and other high frequency words plays in comprehension and production, it is vital that learners have a deep understanding and knowledge of these words. Fortunately for English learners in Japan, most of these 1K-words are those learned beginning in junior high school. An important goal for many English learners in Japan to which they orient themselves is successfully passing a high stakes English entrance examination at a prestigious university. Chujo (2004) estimated that learners need to have between the 3,000 and 5,000 words in their repertoire, unfortunately well beyond the vocabulary found in high school textbooks, to achieve 95% coverage on most Japanese university entrance examinations. As a result, the study of higher frequency words is quickly supplanted by word books and word lists of less frequent words, which the learners believe will help them to achieve success in university entrance exams. Language educators need to stress the importance of both high and low frequency items. Because high frequency items are most commonly found in text, learners benefit most from knowing these words deeply. To do so, learners need large amounts of comprehension and production, to have opportunities to meet the most frequent items in their many different forms, meanings, and use. At the same time, these most frequent items do not provide sufficient coverage of Japanese university entrance examinations, and less coverage of L1 daily conversation, books, newspapers, television, and film. Therefore, learners need to study less frequent items also. As we have seen here, learners with a greater vocabulary as measured by the VST also had a deeper vocabulary as measured by the 1K-VDT. For language learners to be ultimately successful, whether in entrance examinations or in language learning in general, it may not be possible to acquire a large amount of vocabulary without deeply acquiring the most frequent items. The corollary may also be true, that learners may not acquire a deep vocabulary without acquiring a large vocabulary.

References


vocabulary. Reading in a Foreign Language, 17, 23-59.
Laufer, B. (1997). The lexical plight in second language reading: words you don’t
know, words you think you know and words you can’t guess. In Coady, J. and
Hucking, T. (Eds.), Second language vocabulary acquisition: a rationale for
need both to measure vocabulary knowledge? Language Testing, 21, 202-226.
DOI: 10.1191/0265532204lt277oa.
productive ability. Language Testing, 16, 33-51.
learners understand how much they don’t understand? Reading in a Foreign
International Review, 32, 27-32.
Studies, University of Wales.
Eurocentres Learning Service.
The International Journal of Language Society and Culture, 26, 93-106.
House.
Cambridge University Press.
Nation, I.S.P. (2006). How large a vocabulary is needed for reading and
teacher, 31, 9-17.
Boston, MA: Heinle.
productions orales : mesure fiable du niveau de compéence langagiére.
Retrieved from researcharchive.vuw.ac.nz/bitstream/handle/10063/970/thesis.
pdf?...1
knowledge in reading comprehension. Canadian Modern Language Review,


Appendix. Sample vocabulary test (1K coverage = 96.80%) with answers in **bold**.

**Fill in each the blanks with one word. Do Test (A) or (B).**

**Vocabulary Test (A)**

<table>
<thead>
<tr>
<th>Ex.</th>
<th>Test (A)</th>
<th>Test (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>She held the young boy in her <strong>arms</strong>.</td>
<td>Matsuzaka has a good <strong>arm</strong>.</td>
<td></td>
</tr>
<tr>
<td>As they walked, he offered her his <strong>arm</strong>.</td>
<td>The political <strong>arm</strong> of the group met with the media.</td>
<td>Mom armed us with supplies to get the house ready.</td>
</tr>
<tr>
<td>Both sides agreed to <strong>disarm</strong>.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. **The **air** was hot this morning when I stepped outside my apartment.**
   - Bombs went off in the **air**.
   - We could have taken the train but we went by **air** instead.
   - The **air** in the room was unpleasant.
   - An interesting story **air** on TV last night.
   - We need to clear the **air** between us and find answers to our problems.

2. **She had an **opportunity** to travel but did not take it.**
   - Students need more work **opportunities**, not less.
   - Women have fought for good work **opportunities** for years.
   - This was a good photo **opportunity**.
   - There is no better **opportunity** than now to get this finished.
   - When **opportunity** comes, you had better take it.

3. **As I **passed** the store, I saw my friend.**
   - I enjoy sitting and watching people **pass**.
   - The new road will **pass** through center of the city.
   - Can you **pass** me the salt?
   - After the death of his mother, the house **passed** to him.
   - A week **passed** before they spoke to each other again.

4. **His students hope he will **accept** the book as a present.**
   - She offered to drive him to town and he **accepted**.
   - He is good for her but she will not **accept**.
   - Most stores **accept** cash.
   - Machines do not **accept** 1-yen coins.
   - He was willing to **accept** he started the trouble.

5. **The students began **talking** at the same time.**
   - My friends and I were **talking** for hours yesterday.
   - First, you should **talk** with your teacher about this problem.
   - I went to an interesting **talk** last Sunday.
   - **Talks** between the two sides are going well.
   - People are already **talking** about the November election in the US.
6 I _need_ to use the toilet.
   There is no _need_ to talk that way.
   The car _needs_ fixing.
   There is a _need_ for you to learn to think on your own.
   You do not _need_ to worry.
   There are many children in _need_ in other countries.

7 I _recognized_ many of my students playing near the station.
   The student _recognized_ what he had to do to pass the entrance test.
   Few countries _recognize_ Palestine.
   Ichiro is _recognized_ as a great player.
   I had changed so much my friends did not _recognize_ me.
   He was _recognized_ as a danger.

8 We asked the teacher many _questions_.
   The workers were _questioned_ about the way they work.
   Do you ever _question_ what the doctor tells you?
   The people called into _question_ the plan he had.
   The _question_ of helping the environment is an easy one for many people.
   Many easy _questions_ came up on the test.

9 He put the book _on_ the table.
   There were pictures _on_ the walls.
   She put _on_ a hat because it was cold.
   He lay _on_ his back.
   She put her hand _on_ his hand.
   We went out to eat but my father did not have any money _on_ him.

10 Many people equate _power_ with money.
    People used to believe that only men could have _power_.
    The police do not have the _power_ to stop people randomly on the street.
    The Democratic Party of Japan has only been in _power_ since 2009.
    Obama is known for the _power_ of his speeches.
    The Giants, whether you are a fan or not, are a great _power_ in baseball.