

Second-Language Vocabulary Growth in the First Three Semesters of College

Phil Bennett & Tim Stoeckel

Abstract

This paper describes a project aimed at assessing English as a foreign language lexical knowledge of students in their first three semesters of study at college. Four equivalent forms of an instrument to test written receptive knowledge of the words in the General Service List (GSL) and the Academic List (AWL) were developed for this purpose. A validation study with 334 participants found that the instrument had satisfactory dimensionality, the vast majority of its items displayed good technical quality, and Rasch person reliability estimates ranged from .87 to .93 for the four forms. The main study used this instrument to track the vocabulary growth of 144 students from two cohorts over each of their first three semesters of college. On average, these students entered college with knowledge of approximately 1,440 (56.0%) of the words tested, a figure which increased to 1,790 (69.6%) after three semesters. With a minimum criterion set at 80% for demonstrating satisfactory knowledge of a word list, only a minority of students entered college with satisfactory knowledge of either the first or second half of the GSL or the AWL. After three semesters, the majority of students reached this threshold for the first half of the GSL but not for the other two word bands.

Introduction

Second language (L2) vocabulary was once viewed as both relatively easy to develop and less worthy of investigation than grammatical competence or the traditional four skills of reading, writing, listening, and speaking (Richards, 1976). However, when the focus of educational theorists began to fall on language learners as well as language itself, studies of metacognition found that successful learners prized lexical knowledge and recognized its multi-faceted nature (Wenden, 1986). Corpus linguistics has over the last twenty years taken lexical analysis out of the realms of researcher-intuition and provided it with an empirical footing. This work has revealed the complexity of lexical patterning and clarified the size of the challenge that L2 learners face (Hoey, 2005; Nation, 2006; Sinclair, 1991).

Lexical knowledge has been shown to be a strong predictor of performance of general language ability. Studies by Stæhr (2008) and Milton, Wade and Hopkins (2010) have found strong correlations between written and aural receptive vocabulary knowledge and all of the main language skills. The results of both studies, summarized in Table 1, indicate that written and aural vocabularies interact with other skills in different ways but also that there is a close overall relationship between vocabulary size and L2 language ability.

Table 1

Spearman Correlations for Vocabulary Size and the Four Language Skills

Type of Vocabulary Knowledge	Skill			
	reading	listening	writing	speaking
written ^a	.83	.69	.73	
written ^b	.70	.48	.76	.35
aural ^b	.22	.67	.44	.71

^a From Stæhr, 2008. ^b From Milton et al., 2010.

Other studies have attempted to answer the question of how large a vocabulary is required to accomplish particular goals. A commonly-cited benchmark for reasonable, or gist comprehension of everyday discourse is 2,000-3,000 word families (Milton, 2009; Schmitt, 2010b), with a word family being defined as a headword plus its inflections and closely related derivations (e.g., *excite*, plus *excited*, *excites*, *exciting*, *excitement* and *excitedly*). For more demanding tasks, such as reading authentic text without difficulty, vocabularies as large as 8,000-9,000 word families may be required (Nation, 2006). These figures are supported by data from corpora, which show that the most frequently occurring 2,000 words of English cover almost 80% of the words used in typical academic text and around 90% of words in everyday conversation (Nation, 2001).

Another branch of inquiry has looked into the relationship between the percentage of words known in a text and overall comprehension (Hu & Nation, 2000; Laufer, 1992; Laufer & Ravenhorst-Kalovski, 2010; Schmitt, Jiang & Grabe, 2011). One finding common to all of these studies is that there appears to be a linear relationship between coverage, or the percentage of words known in a text, and reading comprehension; no evidence of a percentage at which comprehension markedly improves has been found, nor is there any indication of reading comprehension scores reaching asymptote, when increases in coverage no longer appear to affect comprehension. Unsurprisingly, the general conclusion is: the more vocabulary, the better the comprehension. This statement notwithstanding, two percentage figures have consistently been referred to as benchmarks for predicting learner comprehension of a given text. At present, 98% coverage is considered the point at which learners are likely to be able to read independently (Hu & Nation, 2000; Schmitt, Schmitt & Clapham, 2001). For assisted comprehension (i.e., with

teacher support and/or dictionary use), the picture is less clear, but a minimum of 95% coverage may be a reasonable estimate (Laufer & Ravenhort-Kalovski, 2010; Schmitt et al., 2011). Although the difference of only 3% between these figures may appear slight, it represents a change from one word in 50 being unknown to one word in 20. Over a text of 500 words, this would mean an increase from 10 unknown words to 25.

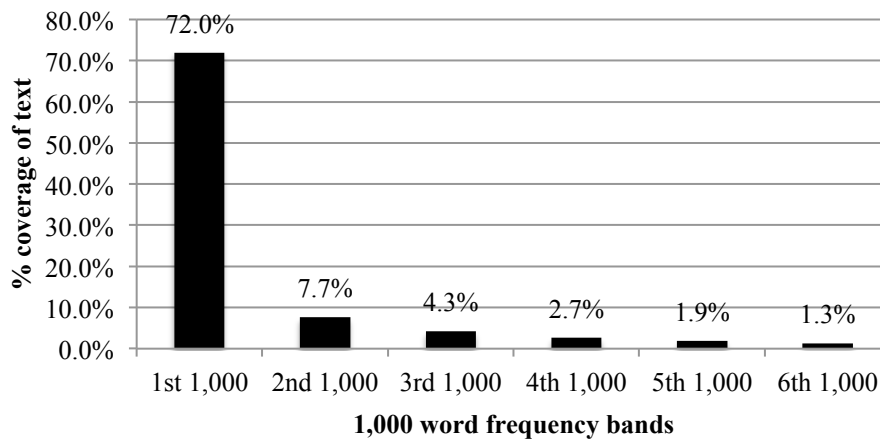
Dictionary use might be considered capable of compensating for poor lexical knowledge, but the research has been inconclusive. Whereas some studies have found no improvement in understanding of L2 texts with dictionary use (Bensoussan, Sim, & Weiss, 1984; Hulstijn, 1993; Nesi & Meara, 1991), others have found significant gains in comprehension (Knight, 1994; Shieh & Freiermuth, 2010). In these latter studies, however, learners with poor lexical knowledge understood less than learners with high vocabulary knowledge (Knight, 1994), even when given ample time to complete the reading task (Shieh & Freiermuth, 2010). This suggests that while dictionary use can be helpful, it is no panacea for limitations in vocabulary knowledge. A likely reason for this is that humans have a finite amount of processing ability which cannot be dedicated to comprehension of a text until automaticity of lexical processing has been achieved (Browne, 2008).

Frequency lists

As it is now widely recognized that vocabulary knowledge is closely tied to overall language competence, researchers have tried to identify the words most beneficial for learners to acquire. Word frequency is a useful guide in this regard as there is tremendous variation in the rate at which words occur in language. In a typical text, a relatively small proportion of different word families comprises a large percentage of all of the words in the text. Figure 1 (adapted from Nation, 2001) shows

how the text coverage provided by successively less frequent groups of words becomes progressively smaller.

Figure 1. Percentage of text coverage provided by successive 1,000-word frequency bands in the Brown Corpus (adapted from Nation, 2001).



This implies that while frequency offers a useful initial guide for determining which vocabulary to study, at some point learners will benefit from switching to a more focused approach to lexical development that takes into account their individual study purposes. Such an approach could be catered to by the General Service List (GSL) (Bauman & Culligan, 1995; West, 1953) and the Academic Word List (AWL) (Coxhead, 2000). The GSL was originally developed for use in writing simplified reading materials, but since its composition took into account frequency, range of use, and lack of specialized terms, it has become widely used as a notional core vocabulary list. The GSL consists of 2,284 word families divided into two sublists which approximate the first and second thousand words of English. The AWL was developed to meet the needs of learners studying in an academic environment who already demonstrate adequate knowledge of the GSL. As with the GSL, the AWL used frequency and range of use as criteria for inclusion. The AWL contains 570 word families that are not in the GSL but that appear frequently in texts drawn from four broad academic fields (arts, commerce, law, and science). Within these fields,

texts were sampled equally from 28 subject areas (Coxhead, 2000). Table 2 shows coverage across a range of genres provided by the GSL and AWL.

Table 2

Text Type and Coverage Provided by the GSL and AWL

Levels	Conversation	Fiction	Newspapers	Academic
1 st half of GSL	84.3%	82.3%	75.6%	73.5%
2 nd half of GSL	6.0%	5.1%	4.7%	4.6%
AWL	1.9%	1.7%	3.9%	8.5%
Total	92.2%	89.1%	84.2%	86.6%

Note. Adapted from Nation, 2001.

Knowledge of the GSL and AWL could be expected to provide coverage of 84 to 92% of the words in a typical text, depending on genre. Although these figures are still short of the 95% estimate for assisted comprehension described earlier, the addition of proper nouns and the probability that students will also know some words beyond these word levels will bring the targets closer.

Word frequency has also been identified as a predictor of whether a lexical item is likely to be known by L2 learners (Milton, 2009). Several studies have shown that if learners are tested on their knowledge of a range of words at differing frequency bands, the most well-known words will be those of highest frequency, and the least recognized will be those that appear least often in the language (Beglar, 2009; Richards & Malvern, 2007). Plotting the results of such tests produces charts similar to that shown in Figure 2.

It should be noted that this is a general trend across groups. At an individual level, it is not uncommon for learners to show a deficit at a particular frequency band. Indeed, some studies (e.g., Milton, 2009) have found up to 40% of learners with vocabularies that deviate to some degree from this trend. Likewise, all of the words from a given frequency band should not be assumed to be of equal difficulty; factors

such as morphology, cognate status, part of speech and concreteness have also been shown to affect word recognition (Daulton, 2008; de Groot, 2006; Hayashi & Murphy, 2011, Stoeckel & Bennett, in press). Moreover, Milton (2009) has shown how Zipf’s Law implies that the effects of word frequency diminish in less frequent word bands. Vocabulary tests that cover a broad range of frequency bands have shown evidence of this, with mean scores in mid- to low-frequency bands occasionally deviating from the general trend, as shown in Figure 3.

Figure 2. Frequency profile of word recognition by British learners of French. Adapted from Richards & Malvern, 2007.

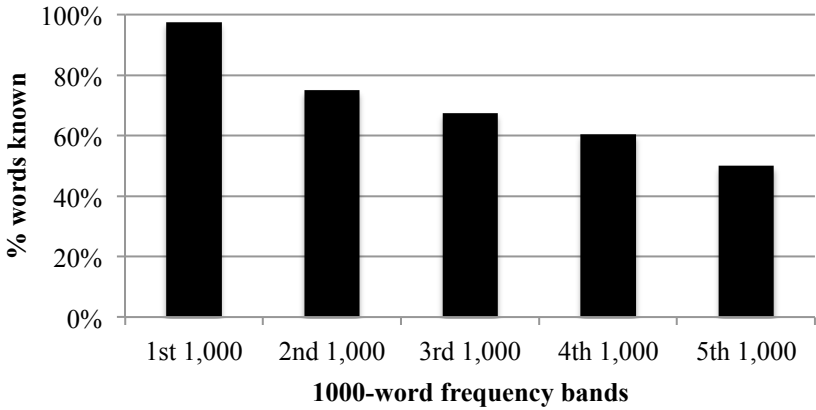
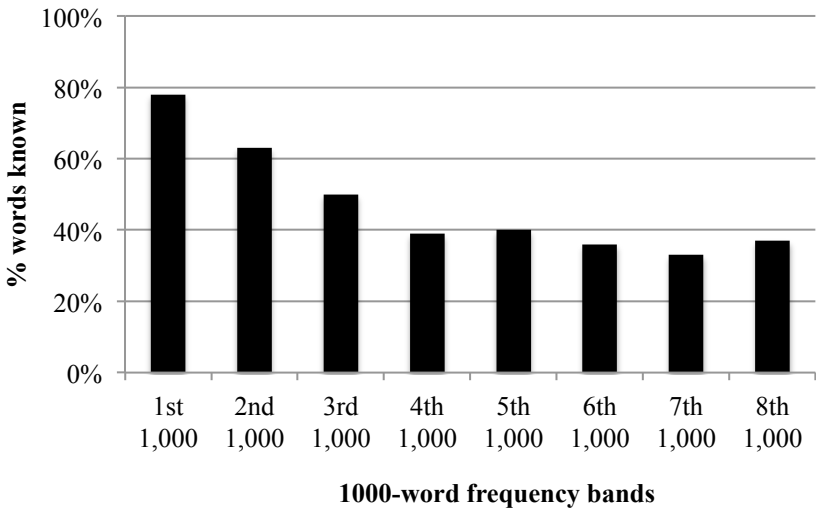


Figure 3. Frequency profile of word recognition by Japanese learners of English. Adapted from Aizawa, 2006.



Frequency can help to predict both the likelihood of a word being encountered and of it being recognized by learners. As such, it is a highly useful criterion to consider in text analysis and in establishing achievement goals for language learners. If learners were provided with specific frequency-based targets and their standing in relation to them, they would have salient goals by which they could evaluate their own progress in developing lexical knowledge. Goals that are specific and challenging yet attainable have been described by Dörnyei (2001) as important components of goal-setting theory. Goal-setting – along with planning, self-monitoring, metacognitive awareness, and use of learning strategies – is one of the components of Tseng, Dörnyei and Schmitt's (2006) construct of self-regulation in vocabulary acquisition.

Similarly, if teachers and administrators were aware of students' vocabulary profiles and growth over time, they would be better able to judge the developmental appropriateness of program materials and to assess program efficacy in relation to vocabulary goals.

Vocabulary Testing Instruments

Among the instruments widely used to provide estimates of receptive vocabulary knowledge are yes/no tests, the Vocabulary Levels Test (VLT), and the Vocabulary Size Test (VST). Yes/no tests have been used in education since at least the 1940s (Bear & Odbert, 1941). In this test design, learners are simply presented with a list of words and asked to indicate which words they know. Anderson and Freebody (1982) enhanced this design by adding a number of pseudowords to check and correct for overestimation of word knowledge. In L2 studies, Meara (1992) has made extensive use of yes/no tests, arguing that their simplicity allows for a far greater number of words to be tested than multiple-choice formats. Criticisms of

yes/no tests have centered on the use of formulas to correct for the selection of pseudowords, and the tendency for test takers from different language backgrounds to respond to the pseudowords in varying ways (Schmitt, 2010b). It should also be noted that yes/no tests require respondents only to indicate word recognition rather than to demonstrate receptive knowledge.

In the VLT, learners must match three out of six words to given definitions (see Figure 4). This test assesses vocabulary knowledge at the 2,000, 3,000, 5,000 and 10,000-word frequency bands, as well as having questions on words drawn from the AWL (Nation, 1983; Schmitt et al., 2001). Validation studies have been carried out on this instrument (Beglar & Hunt, 1999; Schmitt et al., 2001), and it has been widely used as both a diagnostic tool and in research to estimate vocabulary size (Laufer, 1998; Qian, 2002; Stæhr, 2008).

Figure 4. Sample item from the Vocabulary Levels Test. From Schmitt et al, 2001.

1	debate		
2	exposure		
3	integration	_____	plan
4	option	_____	choice
5	scheme	_____	joining something into a whole
6	stability		

The third widely-used vocabulary measurement instrument is the VST (Nation & Beglar, 2007; Beglar, 2009). This is a multiple choice test in which learners read a sentence containing the target word in a natural yet non-defining context and select the correct definition from four choices (see Figure 5). The VST assesses knowledge of 10 words from each of the first to the fourteenth 1,000-word frequency bands of the British National Corpus. As a measure of overall vocabulary size, the VST is

designed to provide an indication of how successfully students will perform with certain materials and tasks and to assess how vocabulary grows over time.

Figure 5. Sample item from the Vocabulary Size Test. From Beglar, 2009.

1. **miniature:** It is a **miniature**.
- a a very small thing of its kind
 - b an instrument for looking at very small objects
 - c a very small living creature
 - d a small line to join letters in handwriting

An important aspect of test design is that consideration should be given to the needs of test takers and the context in which they are learning. In academic settings, it is imperative that efforts are made to maximize recognition of the most frequent 2,000 words of English and to improve recognition of words on the AWL. This might suggest that the VLT would be a suitable instrument for such contexts, but there are reasons why it is less than ideal. First, the format of the VLT described above does not assess knowledge of the first 1,000 words of English. There is a separate VLT form containing picture items to assess these words, but to our knowledge, it has never been validated. Second, there are only two forms available of the most recent version of the VLT, meaning that any attempt to assess vocabulary growth longitudinally would risk a testing effect influencing the results as learners became familiar with the tested words.

For the VST, the same problem exists. Only two forms have been published, and a validation study has been conducted for only the first of these (Beglar, 2009). In addition, having only ten items to estimate knowledge of each 1,000-word frequency band may raise questions over the test's reliability if not over the entire 140-item form then certainly at each frequency band. In academic contexts with learners of low to

intermediate L2 proficiency, knowledge of total vocabulary size is of lesser importance than judging students' understanding of particular word bands. Finally, the lack of questions specifically targeting academic vocabulary means that students would not necessarily be tested on words they are likely to encounter in such contexts.

The following sections describe the development and initial validation evidence of a new test of vocabulary knowledge and then report on a study of English lexical development with students in the first three semesters at Miyazaki International College (MIC), a small English-medium liberal arts college in Kyushu, Japan.

Test Development

The primary purpose of test development was to address the shortcomings of existing instruments for repeated diagnostic assessment of L2 vocabulary knowledge in academic settings. Specifically, we aimed to develop four equivalent forms of a test of written receptive knowledge of the words on the GSL and AWL. This section briefly describes item and test form development and initial validation work. See Bennett and Stoeckel (in press) for a more complete description.

Item Development

Given their applicability to academic contexts, the GSL and AWL were selected as the word lists that would be used as the basis for the test. Eighty headwords were randomly sampled from each of the first 1,000 words of the GSL (hereafter GSL1), the second 1,000 words of the GSL (GSL2) and the AWL. Most test items were written following a blueprint similar to that of the VST. Each sampled headword was presented in bold-face type, followed by a short sentence that used the word in a natural, non-defining context. The Corpus of Contemporary American English (<http://corpus.byu.edu/coca/>) was consulted to confirm that one of the most

commonly occurring members of the target word family was used in these example sentences. This was followed by a definition of the target word and three distractors. Care was taken to use simplified language in test items so as not to introduce construct-irrelevant difficulty (Messick, 1995). That is, for test items targeting knowledge of words in the GSL, only words from the GSL1 were used, and for items targeting knowledge of words in the AWL, only words in the GSL were used. To avoid construct-irrelevant easiness (Messick, 1995), the distractors were written so as to be plausible substitutes for the target word.

A small number of test items differed in format from that of the VST in that pictures rather than text were used for the four answer choices. This approach was adopted in order to avoid having to use words in the answers which were of lower frequency than the target word itself. Two other deviations from the VST format were made after early piloting of the instrument revealed that, despite written directions to skip unknown words, many students had a high ratio of wrongly-answered to skipped items, suggesting that they were guessing and thereby inflating scores and reducing reliability (Zimmerman & Williams, 1965). To reduce guessing, a fifth choice (hereafter Choice E) was added which reads, “I don’t know this word,” together with the threat of a penalty in the test instructions. These two changes were retained when an examination of the data revealed both a reduced ratio of wrongly answered to skipped items (Bennett & Stoeckel, 2012) and an improvement in reliability estimates from .86 to .92. Sample text and picture items are provided in Figures 6 and 7.

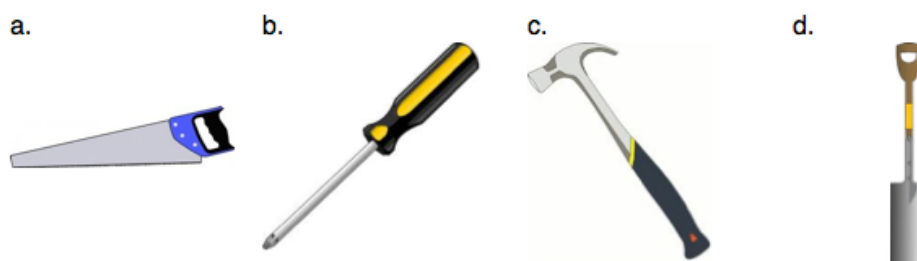
Figure 6. Example vocabulary test item with text-based answer choices.

bias: Be careful of **bias** in your writing.

- a. grammar mistakes
- b. language that is not exact
- c. unfair opinions
- d. informal language
- e. I DON'T KNOW THIS WORD.

Figure 7. Example vocabulary test item with pictures as answer choices.

hammer: I need a **hammer**.



- e. I DON'T KNOW THIS WORD.

Test Form Development

The data from piloting were also used to initially estimate item difficulties. Items were divided into four forms of 60 items each, with 20 items at each of the GSL1, GSL2 and AWL levels. The forms were balanced according to item difficulty, picture items, the parts of speech of the target words, and whether target words existed as loanwords in Japanese.

In order to assess the initial item difficulty calibrations and to judge the equivalence of test forms, some items were taken from their original forms and shared across the other forms to act as anchors (Wolfe, 2000). The end result was four 90-item test forms with 30 items at each level.

Summary of Validation study

An initial validation study was then conducted with students enrolled in two colleges in Japan ($n = 334$). The first was MIC, and the second was a medium-sized university of foreign language study in Osaka. The four 90-item test forms were spiraled in each of the participating 21 class sections. The data were analyzed using the Rasch dichotomous model with the Winsteps software package to determine construct dimensionality, the technical quality of items, the reliability of the four test forms, and the relative difficulty of the forms.

The instrument as a whole was found to be unidimensional, and all but four of the 240 test items displayed good technical quality. That is, more difficult items were likely to be answered correctly only by persons of higher ability, and easier items were typically answered incorrectly only by persons of low ability. The four items with poor technical quality were found to contain ambiguity or overly complex grammatical constructions in the wording. These four items were revised and will be monitored in future test administrations.

Rasch person reliabilities for the 90-item forms ranged from .92 to .95, and with the anchor items removed, the reliabilities for the 60-item forms ranged from .87 to .93.

Rasch analysis transforms raw scores to a logit scale, meaning that for every possible raw score, the logit values from each test form can be compared to assess equivalency. We found that for any given raw score across the four test forms, the logit values were within one standard error of each other. More specifically, Forms A and C were nearly identical in terms of difficulty, Form B was the most difficult, and Form D was the easiest. The difference in difficulty between Forms B and D was equivalent to about three points over the 60-item form.

Overall, the test forms appear to be good measures of the construct of written receptive vocabulary knowledge. The research method we have chosen allows for underperforming items to be identified, revised, and then placed back into the item pool to be re-assessed. In this way, the test forms undergo a process of continual refinement (Wise & Kingsbury, 2000). Similarly, because Rasch analysis produces quantitative estimates of difficulty for each test item, it will be possible to redistribute items among the four forms to more closely approximate the goal of equivalency. This is another step that we plan to take once we have obtained sufficient data on our revised items.

Study of Vocabulary Development at MIC

Purposes

The primary purpose of this aspect of the project was to estimate MIC students' knowledge of the words in the GSL and AWL both upon entry into MIC and as they progress through the first three semesters.

Method

Participants. Members of the MIC first-year cohort of 2011 ($n = 74$) participated in the study for two years. This group included 71 Japanese and three Korean students. Members of the 2012 cohort ($n = 70$), consisting of 67 Japanese, two Chinese, and one Korean, participated for the second year of the study.

Instrument. The instrumentation included Forms A, B, C, and D of the vocabulary test. During the study, the length of these test forms varied from 60 to 90 items as some items were shared across other test forms to (a) act as anchors for the validation study described above and (b) increase test reliability. A paper and pencil format was used except in the cases of 47 students with whom a computer-based version was piloted in December 2012.

Protocol. The test was administered six times during the 2011 and 2012 academic years, at the beginning and end of the spring semesters and at the end of the fall semesters. The 2011 cohort participated for the first five of these, and the 2012 cohort the final three. The basic protocol was for examinees to encounter the test forms sequentially such that those who used Form A for one administration would use Form B for the next, and so on. This protocol was followed except for in April 2012 when we changed from using the same test for each given administration (i.e., Form A in April, Form B in July, etc.) to spiraling all four test forms in each class section for each administration. Care was taken to record which test forms students received in April 2012, and the basic protocol has been followed since.

Measures. A formula which penalizes guessing was used to score the tests. Correct answers earned one point, skipped items or those answered with Choice E earned zero points, and wrongly answered items earned minus one-third. The result of this formula was used to calculate the percentage of items correct for each of the three sections of the test, which in turn were used to estimate the approximate number of words each student knew at the GSL1, GSL2, and AWL levels. For group means, these calculations were made twice, once with the complete data set and once removing students who missed one or more test administrations. The results were nearly identical, and those utilizing the full data set are reported below.

Results

Table 3 presents the means, standard deviations, and Rasch person-reliability estimates for each vocabulary test form in each administration. The instrument displayed satisfactory reliability, with coefficients ranging from .82 to .96 throughout the study and no lower than .89 after the addition of Choice E in December 2011.

Table 3

Measurement Properties of the Vocabulary Tests

test date	Test form (no. items)	n^a	M	SD	Rasch Person reliability
2011 April	A (60)	73	37.0	8.2	0.82
2011 July	B (90)	74	63.6	11.7	0.86
2011 December	C (90)	69	60.8	12.3	0.92
2012 April	A (90)	33	52.9	14.2	0.92
	B (90)	32	56.2	14.2	0.93
	C (90)	31	58.6	15.5	0.94
	D (90)	33	52.8	20.4	0.96
2012 July	A (72)	33	49.9	13.8	0.94
	B (72)	33	50.5	9.7	0.89
	C (72)	33	49.0	11.2	0.93
	D (72)	33	49.2	11.2	0.91
2012 December	A (72)	19	41.9	14.9	0.93
	B (72)	19	44.6	14.6	0.93
	C (72)	17	39.0	12.3	0.90
	D (72)	11	42.7	13.1	0.93

^aFor each of the three 2011 administrations, one version of the test was given to the available members of the 2011 cohort. In April and July 2012, the four versions were given at each administration to available members of both cohorts. In December 2012, the four versions were given to members of the 2012 cohort; members of the 2011 cohort were on study abroad.

Table 4 shows the means and standard deviations of the scores in each word level for both cohorts combined; the first row of data displays vocabulary knowledge upon entry into MIC, and the subsequent rows exhibit change over time. Keeping in mind that the GSL is roughly 2,000 word families and the AWL 570, the average student demonstrated knowledge of roughly 1,290 GSL and 150 AWL word families upon entry to MIC and experienced an average gain of about 280 words from these lists during the first semester, with a much smaller gain of about 70 words for the second and third semesters combined.

During the second semester, there was a loss in all three word bands, for which there is no clear explanation. Choice E was introduced at the end of the second

semester in 2011 (Year 1 December); this was a significant change in test format for the cohort of that year which could be a plausible cause of lower test scores. However, this explanation is unsatisfactory because the 2012 cohort, which did not experience a change in test format, also experienced a decrease during the second semester.

Table 4

Mean Percentage of GSL and AWL Words Known Through Three Semesters

date	<i>n</i> ^a	% of Words Known (<i>SD</i>)		
		GSL1	GSL2	AWL
Year 1 April	139	70.7 (21.4)	58.3 (23.1)	27.0 (21.3)
Year 1 July	142	79.6 (17.2)	67.2 (19.3)	44.2 (22.4)
Year 1 December	129	77.5 (17.4)	63.4 (19.5)	39.2 (21.6)
Year 2 April	63	81.5 (14.0)	67.3 (15.6)	41.1 (18.9)
Year 2 July	64	82.9 (15.5)	69.6 (19.7)	46.7 (20.0)

^aThe discrepancy in the size of *n* between years 1 and 2 is due to the fact that the 2011 cohort has participated in the study for two years, and the 2012 cohort for just one.

There was considerable individual variation in vocabulary scores, which can be seen in Figures 8-10. These figures show the percentage of students who achieved satisfactory knowledge as opposed to moderate or large gaps in knowledge of each word band over time. The criterion for satisfactory knowledge was set at 80% based on Milton (2009), who demonstrated that groups of high ability learners achieve an average score of 85 to 90%, and not 100%, even for high frequency word bands. Given that this is an average with some learners scoring lower, we feel that 80% is an appropriate standard. The criterion for having large gaps in knowledge was arbitrarily defined as knowledge of less than 60% of a word band.

Figure 8 shows that less than half of the students demonstrated satisfactory comprehension of the GSL1 upon entry into college, and roughly three quarters did after three semesters. Figure 9 reveals that only a minority of students achieved satisfactory knowledge of the GSL2 after three semesters and Figure 10 that just 5% of students reached this level for the AWL in the same time frame.

Figure 8. Percentage of students at three levels of GSL1 knowledge during the first three semesters of college.

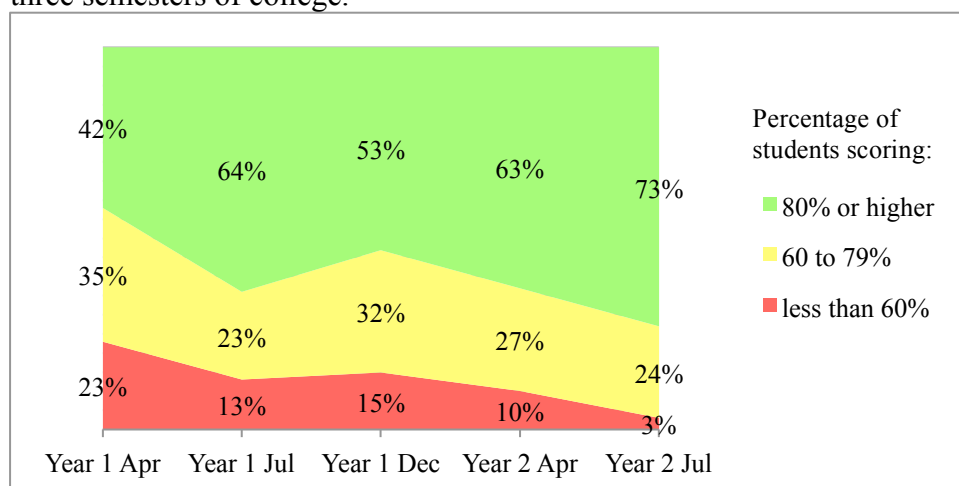


Figure 9. Percentage of students at three levels of GSL2 knowledge during the first three semesters of college.

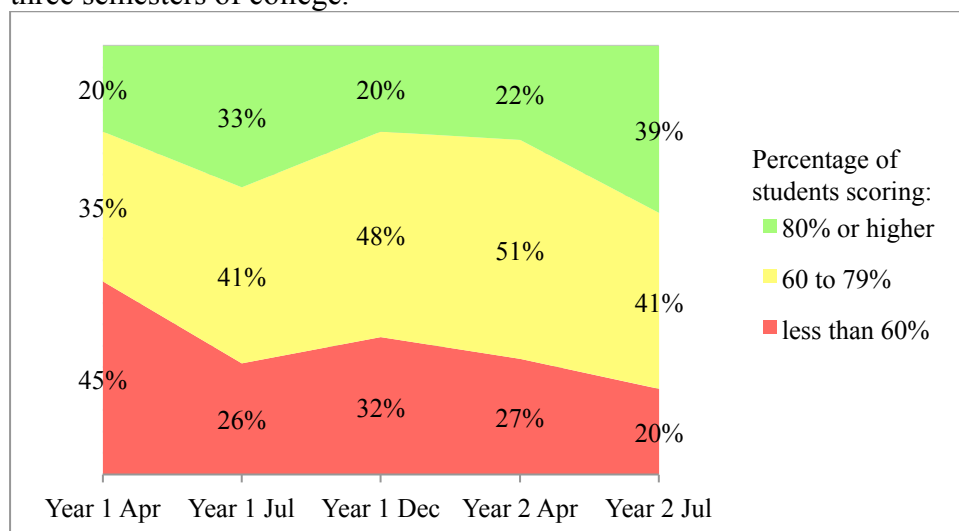
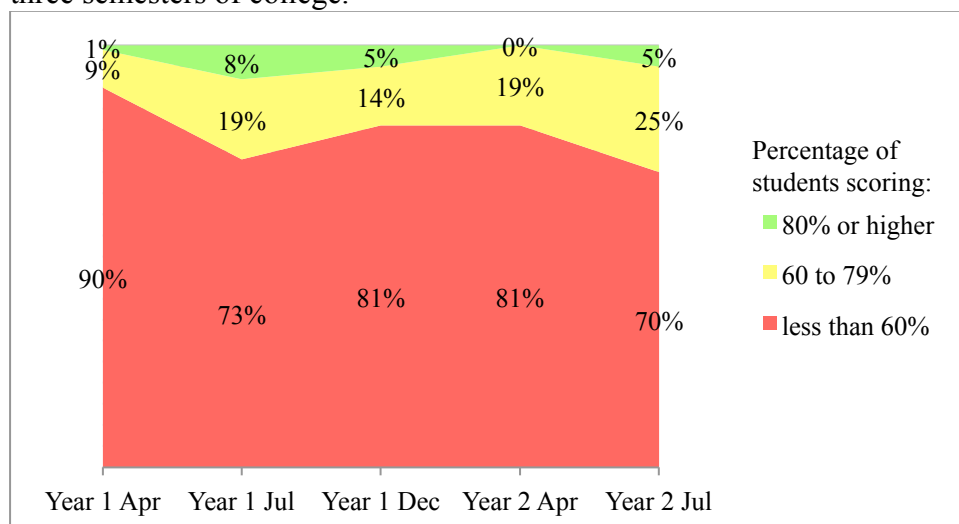


Figure 10. Percentage of students at three levels of AWL knowledge during the first three semesters of college.



Discussion

We have described a project aimed at assessing students' lexical knowledge in order to provide them with feedback for goal setting, to assess the efficacy of the curriculum in terms of promoting vocabulary development, and to help instructors make informed decisions regarding the lexical demands of materials and classroom activities. The instrument designed to accomplish these tasks has displayed satisfactory reliability and dimensionality, and the vast majority of its items have demonstrated good technical quality. Regarding vocabulary knowledge, our results indicate that few students enter MIC with mastery of the GSL, the core vocabulary necessary for accomplishing everyday tasks in English, and virtually none with mastery of the AWL. There is clear evidence of vocabulary growth, but after three semesters most students do not appear to achieve satisfactory understanding of the GSL2 or the AWL.

Implications

Keeping in mind the figures presented in Table 2 regarding the coverage provided by the GSL and AWL for various genres of text, it is likely that the majority of MIC students in the first three semesters fall short of both the 98% coverage required for unassisted comprehension or the 95% figure for assisted comprehension of authentic texts.

How might this situation be addressed? There has been considerable debate over whether lexical knowledge is better acquired through explicit instruction or implicit, exposure-based, learning (Han and Ellis, 1998; Krashen, 1989). Most researchers would now argue that a well-planned vocabulary learning component within a larger educational program would allow for both approaches, since each complements the other (Sökmen, 1997).

The argument for explicit attention to vocabulary in the classroom is that it will lead to greater noticing, and therefore uptake, among learners (Schmitt, 2010a). It has also been shown that programs which draw attention to, and have learners work on vocabulary knowledge in class lead to greater gains than those which rely solely on incidental learning (Folse, 2004; Laufer, 2005). Hunt and Beglar's (2005) framework for vocabulary development describes three forms of explicit lexical instruction. The first of these is studying decontextualized lexical items. This involves the learning of new vocabulary, consolidation activities for previously learnt words, and expansion activities to raise awareness of word families, affixation, collocational patterns and secondary or abstract meanings. The second form of explicit instruction is training in dictionary use. Nation (2008) suggests that dictionary training helps develop word knowledge by providing examples of natural usage and requiring learners to consider the appropriate senses of polysemous words. The final form of explicit instruction is training learners in inferring word meaning from context. While this is actually a compensation strategy to aid reading rather than a vocabulary learning skill per se, it reinforces knowledge of affixes and word families. It should be stated, however, that some studies have found low success rates for inferring word meaning and extremely low rates of retention for words learned in this fashion (for reviews, see Laufer, 2003, 2005). To have a reasonable chance of inferring word meaning correctly, learners may require knowledge of around 98% of the contextual words (Hunt & Beglar, 2005), and any words learnt in this fashion ought to be reinforced with other learning activities.

As for implicit instruction, the clearest advantage is that it will allow access to a far greater range of vocabulary than could ever be provided through explicit classroom instruction. Simple time constraints mean that learners must acquire much

of their vocabulary from exposure to discourse. Implicit instruction also provides access to contextual, particularly collocational, word knowledge that can be difficult to teach in the classroom. And finally, implicit learning occurs alongside other class activities, making it ideal for programs with multiple goals (Schmitt, 2010a).

Hunt and Beglar (2005) argue that while explicit instruction can lead to gains in lexical knowledge, the learning it provides is of limited value unless it can be activated in context. Extensive reading (ER) programs are designed to provide as much exposure as possible to contextualized comprehensible input. Both quantity and quality of exposure are crucial here: vocabulary learning is incremental, and without review of previously encountered words through regular reading, any gains may quickly disappear. Similarly, ER material must be at an appropriate level for learners. If the vocabulary demands are too high, it will not be possible for fluent reading to occur because learners will have little chance of understanding without using a dictionary. Likewise, material that is too easy will not provide students with frequent enough opportunities to recycle recently-learned words (Nation, 2001).

Automaticity, or speed of access, is one aspect of vocabulary knowledge that must be nurtured for fluent language use to occur (Meara, 1997), and implicit learning activities are ideally suited to promoting this. Through activities that entail repeated encounters with key vocabulary, learners are given the opportunity to comprehend and use targeted language without having to process new meanings afresh. Integrated tasks, in which students might first encounter a word in written or audio-visual material and are then required to use it in discussion or a written response about the content of the material, are one way to achieve this. Other possibilities include narrow reading, in which learners read several texts on related topics, and mixer activities, in

which learners complete the same spoken activity several times with different partners (Hunt & Beglar, 2005).

Finally, in addition to teaching activities, consideration should be given to training students to be good vocabulary learners. The concept of self-regulation has been drawn from educational psychology and applied specifically to vocabulary learning in a foreign language (Dörnyei, 2005; Tseng et al., 2006). Self-regulation describes learners' ability to maintain commitment to learning by establishing goals and incentives, to increase self-awareness of learning style preferences and tendencies to procrastinate, to manage disruptive emotional states, and to control their environment so as to take advantage of positive influences and avoid negative influences that might hinder their learning (Tseng et al., 2006). These are abilities that can be developed through instruction and guided experience. Nation (2008) and Thornbury (2002) describe the need to develop positive attitudes toward vocabulary learning through training students in strategies for understanding, recording, reviewing, and utilizing vocabulary, and by creating opportunities for them to experience success that will enhance motivation.

On a broader scale, there is clear value in establishing both short- and long-term goals for lexical development at the level of individual courses, language programs, and institutions (Laufer, 1992; Nation, 2001, Schmitt, 2008). With data provided by studies such as this, instructors can make informed decisions regarding the appropriateness of instructional materials, the need for explicit instruction of particular lexical forms, and the suitability of learning goals for individuals or classes.

Testing can also play an important role in course programming. The instrument described in this paper is intended to provide diagnostic feedback for learners and instructors over multiple semesters of study; however, given the strong

relationship between lexical knowledge and performance in the four main language skills, achievement tests, the results of which are included in course grading, should be considered both at regular intervals throughout courses and at their conclusions (Nation, 2001).

Future Research

The findings of this study raise several questions worthy of further investigation. First, it would be informative to analyze a representative sampling of the teaching and learning materials used during the first three semesters at MIC in order to ascertain the lexical burden they place on students. Specifically, these resources could be analyzed with lexical profiling software such as that available on the Lextutor website (<http://www.lex tutor.ca/vp/>) to determine the percentage of coverage provided by the GSL and AWL. This information, combined with the data we now possess on our students' vocabulary knowledge, would enable us to estimate the percentage of words our students are familiar with in the materials they encounter.

Second, and more important, it would be useful to assess students' level of comprehension of the texts currently used during the first three semesters at MIC. The 95 and 98% coverage required for assisted and unassisted comprehension (respectively) are probabilistic: learners whose lexical knowledge falls short of these figures are sometimes able to compensate by making use of world knowledge and/or reading strategies such as previewing texts, using context clues or making inferences (Laufer & Ravenhorst-Kalovski, 2010). In the MIC environment, students receive significant amounts of support in the form of not only direct vocabulary instruction and glossing of low-frequency words but also active learning techniques that engage learners with the ideas in a text in multiple ways.

A third line of inquiry would be into the rate of overall vocabulary growth of MIC students. The present study indicates that on average our students gain knowledge of approximately 350 words from the GSL and AWL during the first three semesters, or slightly less than one new word for every two hours of classroom exposure to English. Due to the lack of longitudinal studies on vocabulary development in the Japanese context, this figure cannot easily be interpreted. It should also be noted that our students learn low-frequency words in addition to those on the GSL and AWL. Our feeling is that the present rate of vocabulary growth should serve as a benchmark against which further developments in the college program can be compared.

Limitations

This study has several limitations. First, the measure of vocabulary knowledge is restricted to the written receptive aspect of the construct. Though this approach is consistent with the majority of existing studies, the degree to which students are able to understand English vocabulary in aural contexts or use it productively is unknown. Second, the instrument assessed only breadth of knowledge, that is, how well learners could match a word with its definition. Depth of word knowledge (e.g., common collocates, part of speech, polysemy) was not investigated. Third, because we assessed familiarity with only the GSL and AWL, we cannot make informed statements regarding overall vocabulary size. Fourth, the lexical knowledge of examinees with small vocabulary sizes may have been underestimated because these learners are likely to have been unable to understand some of the words in the item stems or answer choices. Previous research has demonstrated that learners score higher on tests of vocabulary when answer choices are in the L1 (Ruegg, 2007), but the degree of difference in scores would be mediated by both the size of learners'

vocabulary knowledge and the actual words used in the test items. Finally, we have evidence of differential item functioning (DIF) for some test items when comparing responses of Japanese to Korean respondents (Stoeckel & Bennett, in press). That is, even when differences in ability were accounted for, some test items were easier for one group over the other. Though we have identified the cause of DIF in some items as related to construct-relevant factors such as loanword status of English words in the Korean and Japanese languages, because equivalency of the four test forms was determined with item difficulty estimates derived from responses of only Japanese students, the equivalency of test forms for our Korean students (and also those of other nationalities) is uncertain. Because these students comprise a small percentage of our student body, it is unlikely that this impacts cohort-level estimates of growth over time, but it may result in less precision in reports of growth for these individual students.

Concluding Thoughts

These limitations notwithstanding, it is our hope that this project has had and will continue to have beneficial outcomes in the MIC context. For the past two academic years, participants have been provided with regular reports of individual vocabulary knowledge and growth together with study lists targeting the first sizeable gap in their understanding of the GSL and AWL (for an example, see the Appendix). Though obtaining estimates of vocabulary development is an important first step, there is a need to develop methods of instruction and guidance that will help students to fulfill their learning potential. With a clearer understanding of students' level of lexical comprehension, it should be possible to provide learners with more individualized support that will aid them in comprehending course materials and becoming more involved in classroom discourse. Finally, considering the relationship

between vocabulary size and performance in the traditional four skills, this project could play a role in helping the institution achieve its broader aims of producing graduates with high levels of language ability who are informed of global issues and confident in expressing their views.

Acknowledgements

We thank Benjamin Peters for his feedback on an earlier draft of this paper.

References

- Aizawa, K. (2006). Rethinking frequency markers for English-Japanese dictionaries.
In M. Murata, K. Minamide, Y. Tono, and S. Ishikawa, (Eds.), *English Lexicography in Japan* (pp. 108-119). Tokyo: Taishukan-shoten.
- Anderson, R. C., & Freebody, P. (1982). *Reading comprehension and the assessment and acquisition of word knowledge* (Technical Report No. 249). Retrieved from University of Illinois at Urbana-Champaign, Illinois Digital Environment for Access to Learning and Scholarship website:
https://www.ideals.illinois.edu/bitstream/handle/2142/17540/ctrstreadtechrepv01982i00249_opt.pdf?sequence=1
- Bauman, J., & Culligan, B. (1995). *About the General Service List*. Retrieved from <http://jbauman.com/aboutgsl.html>
- Bear, R. M., & Odbert, H. S. (1941). Insight of older pupils into their knowledge of word meanings. *The School Review*, 49(10), 754-760.
- Beglar, D. (2009) A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, 27(1), 101-118. doi:10.1177/0265532209340194
- Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 Word Level and University Word Level Vocabulary Tests. *Language Testing*, 16(2), 131-162.
- Bennett, P., & Stoeckel, T. (2012). Variations in format and willingness to skip items in a multiple-choice vocabulary test. *Vocabulary Education and Research Bulletin*, 1(2), 2-3.
- Bennett, P., and Stoeckel, T. (in press). Developing equivalent forms of a test of general and academic vocabulary. *JALT 2012 Making a Difference: Conference Proceedings*.

- Bensoussan, M., Sim, D., & Weiss, R. (1984). The Effect of Dictionary Usage on EFL Test Performance Compared with Student and Teacher Attitudes and Expectations. *Reading in a Foreign language*, 2(2), 262-76.
- Browne, C. (2008). A Research-based approach to developing Portable Technology Software for testing and teaching high frequency vocabulary. In I. Koch (Ed.) *CamTESOL Conference on English Language Teaching: Selected Papers Volume 4, 2008* (pp. 8-18). Retrieved from http://www.camtesol.org/Download/Earlier_Publications/Selected_Papers_Vol.4_2008.pdf#page=15
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238. doi:10.2307/3587951
- Daulton, F. E. (2008). *Japan's built-in lexicon of English-based loanwords*. Clevedon: Multilingual Matters.
- de Groot, A. M. B. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Language Learning*, 56(3), 463-506.
- Dörnyei, Z. (2001). *Teaching and researching motivation*. Harlow: Pearson.
- Dörnyei, Z. (2005). *The psychology of the language learner: individual differences in second language acquisition*. New Jersey: Lawrence Erlbaum Associates.
- Folse, K. (2004). *Vocabulary myths: applying second language research to classroom teaching*. Ann Arbor: University of Michigan Press.
- Han, Y. & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, 2(1), 1-23.
- Hayashi, Y. & Murphy, V. (2011). An investigation of morphological awareness in Japanese learners of English. *Language Learning Journal*, 39(1), 105-120.
- Hoey, M. (2005). *Lexical Priming*. Abingdon: Routledge.

- Hu. M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403-430.
- Hulstijn, J. H. (1993). When Do Foreign- Language Readers Look Up the Meaning of Unfamiliar Words? The Influence of Task and Learner Variables. *The Modern Language Journal*, 77(2), 139-147.
- Hunt, A. & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language*, 17(1), 23-59.
- Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal*, 78(3), 285-299.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: additional evidence for the input hypothesis. *The Modern Language Journal*, 73(4), 440-464.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud and H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: Macmillan.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics*, 19(2), 255-271.
- Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire more vocabulary by reading? *The Canadian Modern Language Review*, 59, 565-585.
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, 5, 223-250.
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading*

- in a Foreign Language*, 22(1), 15-30. Retrieved from <http://nflrc.hawaii.edu/rfl/>
- Meara, P. (1992). *EFL vocabulary tests* (1st ed.). Swansea: Centre for Applied Language Studies, University College Swansea.
- Meara, P. (1997). Towards a new approach to modelling vocabulary acquisition. In N. Schmitt and M. McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy* (pp. 237-257). Cambridge: Cambridge University Press.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist* 50(9), 741-749. doi:10.1037//0003-066X.50.9.741
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
- Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in English as a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, and M. Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 83-98). Bristol: Multilingual Matters.
- Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, 5(1) 12-25.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524759
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59-82. doi:10.3138/cmlr.63.1.59
- Nation, I. S. P. (2008) Lexical awareness in second language learning. In J. Cenoz and N. H. Hornberger (Eds.), *Encyclopedia of Language and Education*, 2nd

- Edition, Volume 6: Knowledge about Language* (pp. 167-177). New York: Springer.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Nesi, H., & Meara, P. (1991). How using dictionaries affects performance in multiple-choice EFL tests. *Reading in a Foreign Language*, 8, 631-631.
- Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective. *Language Learning*, 52(3), 513-536.
- Richards, B., & Malvern, D. (2007) Validity and threats to the validity of vocabulary measurement. In H. Daller, J. Milton, and J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 79-92). Cambridge: Cambridge University Press.
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10(1), 77-89.
- Ruegg, R. (2007). The English vocabulary level of Japanese junior high school students. In K. Bradford Watts, T. Muller, & M. Swanson (Eds.), *JALT2007 Conference Proceedings*. Tokyo: JALT. Retrieved from <http://jalt-publications.org/archive/proceedings/2007/E013.pdf>
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329-363.
- Schmitt, N. (2010a). Key issues in teaching and learning vocabulary. In R. Chacón-Beltrán, C. Abello-Contesse, and M. Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 83-98). Bristol: Multilingual Matters.

- Schmitt, N. (2010b). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan. doi:10.1057/9780230293977
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26-43.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88.
- Shieh, W., & Freiermuth, M. R. (2010). Using the DASH Method to Measure Reading Comprehension. *TESOL Quarterly*, 44(1), 110-128.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sökmen, A. J. (1997). Current trends in teaching second language vocabulary. In N. Schmitt and M. McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy* (pp. 237-257). Cambridge: Cambridge University Press.
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal* 36(2), 139-152. doi:10.1080/09571730802389975
- Stoeckel, T., & Bennett, P. (in press). Sources of differential item functioning between Korean and Japanese examinees on a second language vocabulary test. *Vocabulary Learning and Instruction*, 2(1).
- Thornbury, S. (2002). *How to teach vocabulary*. Harlow: Longman.
- Tseng, W., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: the case of self-regulation in vocabulary acquisition. *Applied Linguistics*, 27(1), 78-102.
- Wenden, A. L. (1986). What do second-language learners know about their language learning? A look at retrospective accounts. *Applied Linguistics*, 7(2), 186-205.

- West, M. (1953). *A general service list of English words*. London: Longman.
- Wise, S. L., & Kingsbury, G. G. (2000). Practical issues in developing and maintaining a computerized adaptive testing program. *Psicológica, 21*, 135-155.
- Wolfe, E. W. (2000). Equating and item banking with the Rasch model. *Journal of Applied Measurement, 1*(4), 409-434.
- Zimmerman, D. W., & Williams, R. H. (1965). Chance success due to guessing and non-independence of true scores and error scores in multiple-choice tests: Computer trials with prepared distributions. *Psychological Reports, 17*(1), 159-165.

Appendix

Example Report of Vocabulary Knowledge

Figures A1 and A2 provide an example of the reports which are given to students after each test administration.

Figure A1. Example first page of vocabulary test reports for students. Estimated percentage of words known in the GSL1, GSL2, and AWL word bands is provided for the most recent and all previous test results.

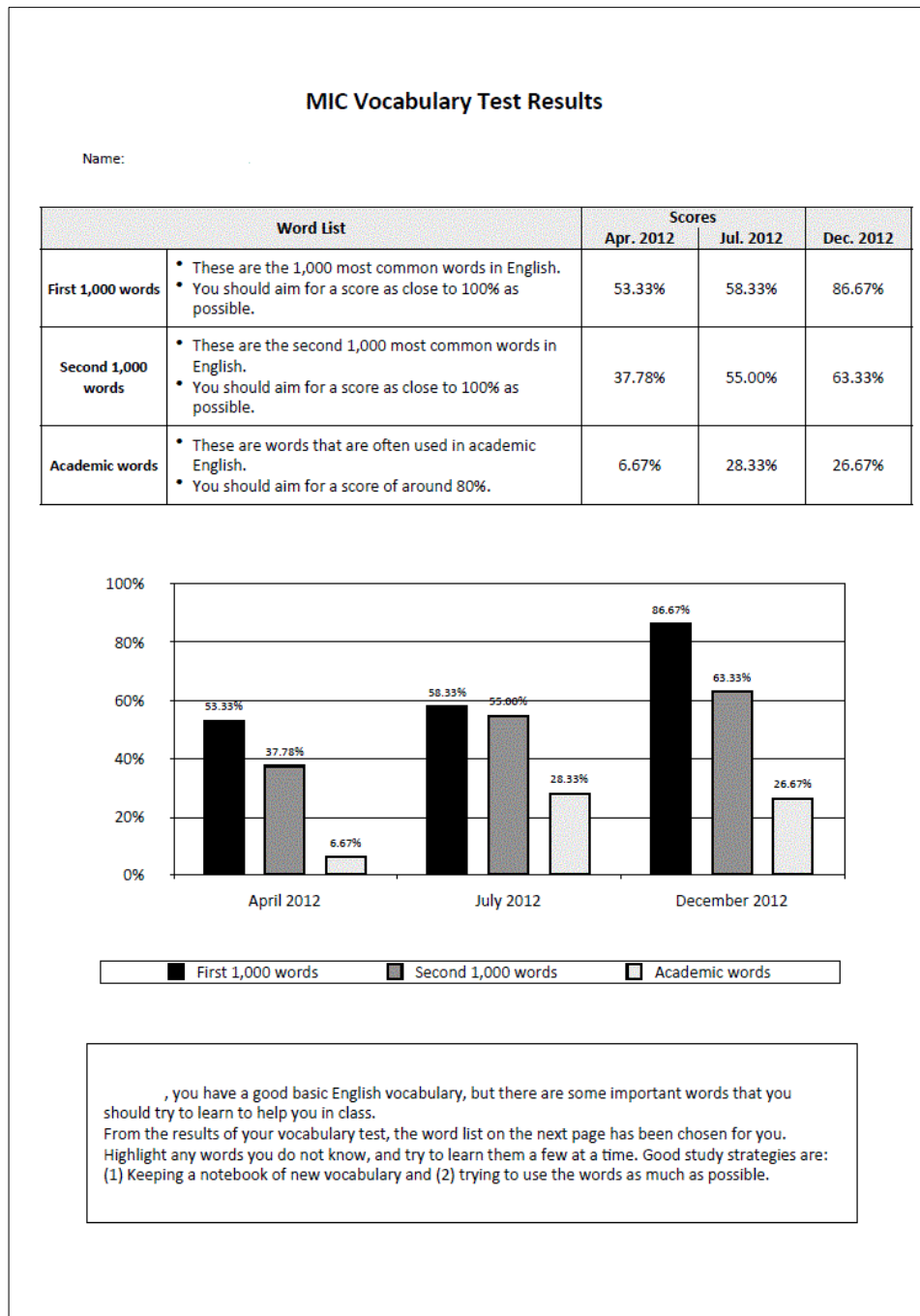


Figure A2. Example second page of vocabulary test reports for students. Page 2 provides students with a list of specific words to study based upon test results. The top of the page indicates to the student which words are included in the list (in this example it is part 2 of the second 1,000 words) and instructs students in how to use the information to address gaps in lexical knowledge.

Second 1,000 words (Part 2 of 2)											
<ul style="list-style-type: none"> • These words are very common and would be useful for you to learn. • Highlight any words you do not know and try to learn them. • Ask your teacher for advice on building vocabulary. 											
quantity	weekend	merchant	cheap	shut	connect	cheer	straighten	objection	false	prevention	astonish
fun	treasury	coal	dish	confident	upset	nut	scarce	silver	fortune	urgent	shave
chicken	egg	mud	overcome	ruin	distant	split	lunch	tent	cap	aunt	feather
forgive	efficient	extension	cat	introduction	greet	melt	slavery	saddle	thread	zero	sausage
holy	grain	recognition	sacrifice	courage	flood	swear	creep	wrap	haste	idle	lid
wooden	calculate	kiss	complain	actor	excuse	sugar	sweat	nest	bare	fever	debt
prompt	drag	crop	elect	belt	insect	bury	gay	grind	shirt	Christmas	fade
crime	opposition	sail	roar	stir	ocean	wipe	stiff	spell	bargain	regret	confess
sorry	worship	attractive	sake	package	ceremony	faint	brave	plaster	leather	jaw	classification
republic	arrest	habit	temple	punish	decrease	creature	seize	arch	rail	soap	descend
anger	discipline	relieve	self	reflection	prize	tail	convenient	swell	butter	pronounce	cape
visitor	string	wisdom	compete	breathe	harm	wealth	horizon	friendship	dot	empire	mild
pile	harbor	persuade	nurse	anywhere	insure	earnest	moderate	bath	inquire	bowl	clever
violence	cow	certainty	stuff	amuse	verse	envelope	bleed	brass	obey	defendant	soup
steel	grand	cloth	stomach	dull	pot	invention	coin	taxi	hut	daylight	whistle
wing	funny	eager	peculiar	fate	sincere	sheep	fond	wax	axe	dip	scenery
delay	insurance	deserve	repair	net	cotton	splendid	autumn	duck	transition	suspicious	humble
gentleman	reduction	sympathetic	storm	fellowship	leaf	stamp	clarify	button	collar	imaginary	fancy
pour	strict	cure	ton	fault	rub	float	omit	invent	delivery	ash	decay
medicine	translate	complicate	bundle	warmth	outline	brick	loyal	remedy	reproduce	carriage	polite
stroke	suspicion	dig	grateful	decisive	organ	rice	needle	bush	confession	educator	tribe
bite	noble	curse	crown	vessel	imitation	businessman	lessen	thunder	pan	saw	shield
lung	inquiry	weigh	boundary	pity	caution	backward	complaint	weaken	prejudice	stove	veil
lonely	journey	priest	nowhere	steam	mineral	qualification	pad	poverty	voyage	rubber	kneel
admission	hesitate	excessive	asleep	pin	disagree	artificial	steep	scrape	tobacco	rug	tray
stupid	extraordinary	quarrel	clock	bound	blade	attraction	skirt	arrow	simplicity	misery	explosive
scratch	borrow	widow	boil	companion	trick	lamp	curtain	tender	paste	awkward	log
composition	owe	modest	altogether	toe	treasure	curl	calculation	cruel	cake	rival	heap
broadcast	funeral	dine	lend	reward	immense	shower	laughter	soften	elephant	roast	suck
drum	ambition	politician	holiday	forbid	convenience	elder	solemn	mouse	ribbon	deed	ladder
resist	mixture	custom	precious	wherever	disapprove	bunch	grease	hay	harvest	preference	gap
neglect	slope	educate	wander	tower	destructive	bell	interfere	anyhow	ashamed	explosion	towel
absent	criminal	salesman	ugly	bathe	fork	steer	explode	alike	cave	theatrical	refresh
passenger	seldom	nail	reputation	lodge	noon	flavor	fasten	circular	customary	cultivate	essence
adventure	map	tap	ticket	swallow	ownership	spit	flag	juice	thief	collector	fur
beg	spin	eastern	pretend	multiply	tune	rob	resign	shelf	damp	miserable	ambitious
pipe	praise	possession	dismiss	bow	polish	cream	postpone	bake	sew	wrist	lump
beard	spare	satisfaction	delicate	kingdom	poison	interrupt	patience	hatred	rust	rabbit	annoy
bold	plow	behave	despair	garage	shame	pen	boast	cautious	separation	accustom	toy
meanwhile	telegraph	mercy	awake	permission	loyalty	weave	rope	basket	waiter	tide	heal
devil	barrel	scatter	tea	pump	cottage	orange	envy	wreck	pet	insult	shallow
rescue	airplane	width	straw	thumb	repetition	crush	rid				