Measuring the vocabulary size of native speakers of English in New Zealand secondary schools

Abstract

The primary aim of this study was to examine the vocabulary size of native speakers of English in New Zealand secondary schools. Two equivalent forms of the 20,000 version of the Vocabulary Size Test were used in this study. Two hundred and forty-three 13 to 18 year old native speakers of English at secondary school took an individually-administered version of the test. The data from this study fits with the vocabulary size estimates for younger native speakers of Biemiller and Slonim (2001). The results suggest that most native speakers at secondary school have enough general purpose vocabulary to cope with their reading at school, and any deliberate attention to vocabulary should focus on subject specific vocabulary.

Introduction

The literature on reading is full of calls for the teaching of vocabulary (for example, McKeown & Beck, 2004; Stahl, 2005) because the number of words that a learner knows will have a direct effect on reading comprehension (Biemiller, 2005). It is possible to quantify the relationship between a learner's vocabulary size and the percentage of words that the learner is likely to be familiar with in a reading text (Nation, 2006) and to relate this percentage coverage to the degree of comprehension of the text. Typically, the research shows that the more words that a learner knows, the greater the likelihood of comprehension of the text (Schmitt, Jiang and Grabe, 2011), with a 98% coverage of the vocabulary of the text being the minimum optimal level for most learners to gain adequate comprehension. A whole range of other factors can have a major effect on comprehension, but solely from a vocabulary perspective, 98% coverage is desirable. Although 98% seems high, this means that in every

100 running words of text (roughly every 10 lines), there will be two unfamiliar words for the learner to deal with. Over a long text, this number of unfamiliar words can grow to well over 1000. To read novels and newspapers with 98% coverage of the words in the text, learners would need to have a receptive vocabulary size of around 8000 to 9000 word families (Nation, 2006). A word family is a headword and its closely related inflected and derived forms (Bauer & Nation, 1993).

When suggesting a vocabulary research agenda, Pearson, Hiebert and Kamil (2007) use the example of a receptive vocabulary measure targeting reading. This measure would be of value in determining the degree to which native-speaking learners need vocabulary instruction focusing on what Beck, McKeown and Kucan (2002) call Tier 2 words (the non-technical words that mature language users use when they read and write). Research on vocabulary frequency (Schmitt & Schmitt, 2012; Nation, 2013) suggests that these can be more clearly described as the 6000 mid-frequency words of English (the 4000 to 9000 word family frequency levels). If these words are largely already known, at least receptively, by native-speakers at secondary school, a more appropriate focus to support reading may be the Tier 3 words, namely the technical words specific to domains of study.

Typically, the range of scores on a vocabulary size test is quite large, indicating that this aspect of language development does not proceed at the same pace for all learners. Research has raised concerns over the vocabulary gap that has been shown to exist between learners of different socio-economic backgrounds. While some of this research is methodologically faulty (see Nation & Webb, 2011: 197-200 for a critique of Hart and Risley, 1995), there is plenty of evidence of large individual differences in vocabulary knowledge (Farkas & Beron, 2004; Biemiller, 2005). A vocabulary size test can quantify these differences making it possible to look at the feasibility of bridging the gap over time.

What is vocabulary size?

A test of vocabulary size measures how many words a learner knows. It typically measures a learner's knowledge of the form of the word and the ability to link that form to a meaning. A receptive vocabulary size measure looks at the kind of knowledge needed for listening and reading. A receptive test measures whether learners can provide or choose a meaning when they see the form of the word. A productive measure looks at the kind of knowledge needed for speaking and writing, so it measures whether learners can provide a word form to express a meaning. In this study we are concerned with measuring receptive vocabulary knowledge, particularly the knowledge required for reading.

The reason for testing affects what is counted as a word. The word family (Bauer & Nation, 1993) is the most suitable unit for measuring reading vocabulary knowledge because if the reader knows the base form of a word or a member of the family and has some familiarity with the common word building devices of English, then with a little help from the context it is possible to work out the meaning of previously unfamiliar members of the word family (Biemiller, 2005). Here is an example of a word family based on the headword *acquaint:*

Acquaint, acquainted, acquainting, acquaints, acquaintance, acquaintances, acquaintanceship, acquaintanceships, unacquainted.

Some members are likely to be quite frequent, while others are rather uncommon but nevertheless systematically related to the headword. If the word family was not used as the unit of counting, then knowing *aquaint* and *acquainted* would be counted as knowing two different words.

Previous research on vocabulary size

In the first three quarters of the 20th century, most tests of vocabulary size were based on samples taken from dictionaries. Unfortunately, sampling from dictionaries is almost inevitably methodologically unsound (Thorndike, 1924; Lorge & Chall, 1963; Nation, 1993) leading to highly over-inflated estimates of vocabulary size (Seashore and Eckerson, 1940; Smith, 1941; Diller, 1978). These over-estimates occurred because what is counted as a word was not clearly described and consistently applied, the total number of words in the dictionary was not accurately counted, and the spaced-sampling method (the nth word on every nth page) resulted in a strong bias towards selecting high-frequency words.

At the upper end of the scale, Diller (1978) estimated that seven year olds knew over 50,000 words and junior high school students more than 100,000 words. Seashore and Eckerson (1940) suggested college students knew over 58,000 basic words (dictionary entries). In a re-calculation of Smith's (1941) results using Seashore and Eckerson's tests, adjusting for sampling errors and types of words counted, Lorge and Chall (1963) estimated that instead of first-graders knowing 16,900 basic words as Smith found, they were more likely to know 6,500 basic words.

Thorndike (1924) pointed out that using spaced sampling from a dictionary gave results that depended on the size of the dictionary used, and resulted in the strong bias mentioned above towards high frequency words because these words occupy relatively more space in a dictionary than low frequency words. Goulden, Nation and Read (1990) took account of Thorndike's critique and sampled from Webster's 3rd International Dictionary using data from Thorndike and Lorge (1944) to control for any word frequency bias, and using an explicit description of what would be counted as a basic word. This same description was used to calculate the number of basic words in the dictionary. The sample was given as a checklist test to twenty university students, the results indicating that they knew around 17,000 basic words. This suggested a learning rate of just under 1,000 word families a year.

Using a similar method, D'Anna, Zechmeister, & Hall (1991) also found that university students' vocabulary size was likely to be less than 20,000 basic words. These lower estimates of vocabulary size are more likely to be accurate than the earlier highly inflated figures.

Measures such as the Peabody Picture Vocabulary Test and the vocabulary sections of the Progressive Achievement Tests used in New Zealand schools are not measures of vocabulary size. The selection of words for the tests does not systematically represent a larger quantifiable population of words. Pearson, Hiebert and Kamil (2007) note tests like the Peabody test are norm-referenced and can be used to compare scores among students, but they are not criterion-referenced: the scores cannot be used to say how many words the students know out of a given much larger population of words.

Biemiller & Slonim (2001) tested the vocabulary size of two samples of native speakers from kindergarten to grade six (10 years old) in Canada. They selected words from Dale and O'Rourke's (1976) *Living Word Vocabulary*, which is a list of over 40,000 words. It includes test-based data on the likelihood of school-children at various grade levels knowing each word. The researchers used word-focused interviews with the younger children. For example, in Grade 2 (around 7 years old), the children were read statements such as, 'Johnny fell and broke his arm. What does *arm* mean?' (p. 501). The children were asked to respond by speaking, acting, or pointing to explain what the word meant. Older children in the study were asked to read each test item and write the meaning. Biemiller and Slonim found an average vocabulary size of 5,200 root words (word families) for six year olds which increased to 8,400 root words by nine years old (p.501). These figures equate roughly to two or three words being added to the children's receptive vocabulary each day, making an increase of around 1000 word families per year. As expected, this study showed that vocabulary size increased with age (p.505).

Biemiller (2005) updated the Biemiller and Slonim (2001) data estimating that the number of root meanings known at the end of Grade 2 was around 6,000, "increasing by 1,000 a year to 10,000 at the end of Grade 6" (aged 11 years). He noted a gap between the lowest quartile students and the higher scoring students and stresses the importance of addressing this gap, at least to make sure that lower achieving children do not fall further behind in their vocabulary growth. A goal of the present study is to see if learners, even those with smaller than average vocabulary sizes, do have enough vocabulary to cope with reading their school texts.

Factors affecting vocabulary size

Vocabulary size increases with age (Biemiller & Slonim, 2001; Farkas & Beron, 2004). Race and socioeconomic background also are significant factors affecting size, with African-Americans and learners of low socioeconomic backgrounds having smaller vocabulary sizes (Farkas & Beron, 2004; Hoff, 2003). Gender has also been investigated as a factor. Scarcella and Zimmerman (1998) looked at male and female English as a Second Language (ESL) students' scores on a test of academic lexicon, finding that males gained higher scores even when a wide range of other variables such as length of residence and age of arrival in the United States were controlled for. Using measures of lexical richness, Prados (2010) found no difference between male and female writers in English as a Foreign Language (EFL), just as Biemiller & Slonin (2001) found no difference in their data for males and females (p. 502). The present study focuses on age, but because schoolwork is likely to be a major factor as well, school year is also looked at, both alone and in relation to age. ESL learners were not included in the data analysis and we did not attempt to classify the learners according to ethnicity.

This study focuses on the vocabulary size of native speakers of English aged from 13-18, on factors affecting the vocabulary size of these learners, including age, school year, gender, and the version of the VST.

Research questions

The study addressed the following questions.

1. What are the means and ranges of vocabulary sizes of native speakers at secondary school?
2. What factors affect the vocabulary size of these students?
3. Are these sizes adequate for reading school texts?

Methodology

Participants

This research was a cross-sectional study with the participants ranging in age from 13 to 18. They came from eight schools in Aotearoa/New Zealand in Palmerston North and Wellington. The schools in the study ranged from deciles 6 to 10, with the majority in deciles 6 and 9 (see

http://www.minedu.govt.nz/Parents/AllAges/EducationInNZ/SchoolsInNewZealand/SchoolD ecileRatings.aspx for more information about decile ratings). Decile levels, with roughly 10% of all schools in New Zealand in each of ten decile levels, are used to indicate the socio-economic status of the community that a school gets its students from. The decile levels are used to allocate government funding with lower decile schools getting more funding. It should be noted that even mid-decile and high-decile schools have some learners from low

income families. Table 1 shows how the ~~227~~ 243 students were distributed across age and school year levels.

The Vocabulary Size Test (VST)

Two versions of the Vocabulary Size Test each containing 100 items sampled from 20 frequency ranked lists each of 1000 word families were used in this study. These versions (C and E) are kept in-house, but two other equivalent versions of the test (A and B) are available on Paul Nation's website (http://www.victoria.ac.nz/lals/about/staff/paul-nation; see Coxhead, Nation & Sim, in press, for more on parallel versions of the VST). Both computer-based and paper-based versions of the test were used in this study, depending on the availability of computer facilities and time of testing at each school. After completing the test, participants were given their raw test result, vocabulary size, and an information sheet about what their score meant.

The VST uses a stem plus a four choice multiple-choice format. The item stem consists of the word followed by a very simple non-defining sentence containing the word. The non-defining sentence has the roles of indicating the part of speech of the word, limiting the meaning of the word where words may have a homograph or very different senses, and slightly cueing the meaning by presenting an example of use. Previous research (Henning, 1991) indicated the value of such a context.

The test measures knowledge of the written word form, the form-meaning connection, and to a smaller degree concept knowledge (Nation, 2013: 49). The test measures largely decontextualised knowledge of the word although the tested word appears in a simple non-defining context in the test. The distractors are the same part of speech as the correct answer, and in most cases the distractors are the meanings of words from around the same 1000 word

frequency level as the correct answer. In the following example, the items in square brackets do not appear in the test, but show how the choices were made.

emir: We saw the <emir>.

a.	bird with two long curved tail feathers	[peacock]
b.	woman who cares for other people's children in eastern countries	[amah]
c.	Middle Eastern chief with power in his own land	[emir]
d.	house made from blocks of ice	[igloo]

The VST is presented in a written form, tests knowledge of a word form and its associated meaning, and uses a recognition rather than recall format. The first two characteristics clearly match with the goal of the test to measure the vocabulary knowledge required for reading. The use of choices makes the test easier than a test where the learners would have to recall rather than choose the meaning, but this more sensitive measure of knowledge can be at least partially equated to the support for meaning recall that background knowledge and that wide and narrow linguistic contexts provide when meeting a word when reading a text. The use of choices also increases the reliability of the scoring.

The test items allow learners to use partial knowledge of a word.   The test has distractors that do not share core elements of meaning with the correct answer. So, the item testing *azalea* simply requires learners to know that an azalea is a plant. The choices are written in much easier language than the tested word. For the first and second 1000 word levels, only words from the first 1000 were used in the choices. For words from the 3000 word level upwards, only the first 2000 words are used.

The first version of the VST was a one in 100 sample of 14,000 word families (Nation & Beglar, 2007) and thus contained 140 items. Beglar (2010) evaluated this 140 item

9

monolingual VST in Japan with a mixture of native speakers and non-native speakers. He found that the test distinguished between test takers with different proficiency levels and was consistent and reliable. Rasch reliability measures were around .96. The VST sampling from 14,000 word families may be adequate for testing the vocabulary size of most non-native speakers of English, especially learners of English as a foreign language whose first language is not cognate with English, as Elgort (2013) and Nguyen & Nation (2011) have found. However, with native speaker estimates being up to 17,000 word families (Goulden, Nation & Read, 1990), a larger test was needed.

The sample on which the new test is based is a 1 in 200 sample of a total collection of 20,000 word families.   Multiplying the learner's score on the test by 200 will give the learner's vocabulary size. A learner getting a score of 65 out of 100 on the VST used in the current research is estimated to have a vocabulary size of 13,000 words (65x200). There should not be a correction for guessing because it would distort the measurement of vocabulary size since each tested word represents 200 words. The interpretation of the final scores needs to take into account that test is a partially sensitive recognition test, so the vocabulary size score is likely to be a generous estimate of vocabulary size. For steps for designing a vocabulary size test based on frequency ranked lists of word families, see Nation and Webb (2011).

Read and Chapelle's (2001) framework for vocabulary testing includes test purpose, validity considerations, mediating factors, test design, and validation. From a test design perspective, the VST is a discrete, selective, relatively context-independent vocabulary test presented in a multiple-choice format. Test purpose covers inferences, uses, and intended impacts.

Inferences: At the item level, the test measures receptive knowledge of a written word form. At the test level, it provides an estimate of total vocabulary size where vocabulary knowledge is considered as including only single words (not multiword units). Vocabulary size does not

include proper nouns, transparent compounds, marginal words like *um, er, gee,* and abbreviations. It does not measure the ability to distinguish homonyms and homographs.

Impacts: If it is used as intended, it is a relatively low stakes test for learners. One consequence may be that it substantially underestimates the vocabulary size of learners who are not motivated to perform to the best of their ability, especially those who are low achievers within their education system. This could result in faulty instructional decisions being made about their vocabulary learning needs. Thus the test needs to be administered to such learners on a one-to-one basis with oral support provided where needed.

Test administration

Each participant sat the test individually, which involved an administrator sitting next to the test-taker as she or he answered the test questions on a computer or on paper. The test administrator keeps the test-taker motivated and on-task. Where necessary, the administrator helps with the pronunciation or reading of words. It would have been useful to have included an orthographic recognition measure to control for this factor more systematically than by merely providing help where it was needed. Each test took around 30 minutes. Two versions of the test were used in this study, based on results from an earlier study which suggested these versions were parallel (Coxhead, Nation & Sim, in press).

Ethics permission was granted by the Victoria University Ethics Committee. Consent was sought from parents or caregivers for participants who were under 16 years old.

Results

What are the means and ranges of vocabulary sizes of native speakers at secondary school?

There were 243 students who were native English speakers and who took a test individually. The mean test scores differ by age, as Table 1 shows below (see Table 4 for F test results).

Except for the two students aged 18, there is an increase each year by age. These increases are rather small except for the increase between ages 15 and 16 (65.46-58.93 = 6.53 or 1,368 word families). Between the ages of 13 and 17 these increases average out to 3.20 or 608 word families per year. The standard deviations tend to decrease by age and are roughly around 10 points on the test or 2,000 word families, showing that there can be quite a range of vocabulary sizes at any particular age level.

Table 1: Descriptive statistics for age level

| Age | n. | Mean | Std. Dev. | Std. Err. | 95% Confidence Interval for Mean | | Min. | Max. |
					Lower Bound	Upper Bound		
13	49	54.02	11.267	1.610	50.86	57.18	33	76
14	81	57.38	11.707	1.301	54.83	59.93	32	78
15	42	58.93	10.144	1.565	55.86	62.00	35	81
16	37	65.46	8.265	1.359	62.80	68.12	47	78
17	32	66.81	7.739	1.368	64.13	69.49	50	85
18	2	61.50	19.092	13.500	35.04	87.96	48	75
Total	243	59.48	11.289	0.724	58.06	60.90	32	85

Students aged 13 to 15 with the smallest vocabulary sizes have written receptive knowledge of at least 6,400 word families (32x200). The minimum vocabulary sizes are all 32 or higher

out of 100. The participants aged 16 to 18 with the smallest vocabulary sizes know at least 9,400 word families. Note in column 9 that some students have very large vocabulary sizes for their age, from 15,000 to 17,000 word families.

In each school year level there are students of varying ages. So, there are students aged 13 in years 9 and 10, with most students at 13 years old being in year 9 (40 students). 14 year olds are spread fairly evenly between school years 9 and 10 with a few (5) in school year 11. This spread of age levels across school years is partly the result of students' birthdays being spread across the calendar year, as Table 2 shows.

Table 2: Results by age and school year

Age	School Year	Mean	Std. Deviation	n.
	9	54.26	11.152	46
13	10	50.33	15.044	3
	Total	54.02	11.267	49
	9	56.59	8.787	32
	10	58.79	13.790	43
14	11	51.50	7.369	6
	Total	57.38	11.707	81
	10	56.65	10.540	20
15	11	60.58	10.183	19
	12	63.67	3.055	3

		Mean	s.d.	n.
	Total	58.93	10.144	42
	11	55.33	4.163	3
16	12	66.48	8.347	25
	13	66.00	7.211	9
	Total	65.46	8.265	37
	12	70.00	5.099	4
17	13	66.36	8.010	28
	Total	66.81	7.739	32
	13	61.50	19.092	2
18	Total	61.50	19.092	2

Results in Table 2, along with the vocabulary size means for school year presented in Table 3 below, indicate that there are increases in vocabulary size by school year although there is no increase in this data from year 12 to 13, suggesting that age is probably a better indicator of vocabulary size than school year. Both the F test and Welch's test confirm that the means are statistically different by school year ($F(4, 238) = 11.478$, $p < 0.0005$).

Table 3: Vocabulary size means for school year

School year	Mean	s.d.	n.	Increase
9	55.52	10.252	78	

10	57.76	12.882	66	2.24	
11	58.07	9.764	28	0.31	
12	66.66	7.707	32	8.59	
13	66.03	8.203	39	-0.63	
Total	59.48	11.289	243		

What factors affect the vocabulary size of these students?

The second research question focused on the factors of age, gender, school year, and test version and how they might affect the results. A one-way analysis of variance was used to test whether the mean result was different by the different groups. A robust ANOVA (Welch's test) was used because of the inhomogeneity of variances between groups. Games-Howell Post Hoc tests were performed where there were more than two groups. Table 4 shows 102 females and 141 males sat the test with very similar average scores of 59.73 and 59.30 out of 100.

Table 4: Factors that may affect the vocabulary size of secondary school students

Variable	Comparison	n.	Mean	s.d.	F
Gender	Female	102	59.73	11.095	$F(1, 241) = 0.085$, p = 0.771
	Male	141	59.30	11.463	
Age	13	49	54.02	11.267	$F(5, 237) = 8.913$, p = 0.000
	14	81	57.38	11.707	

	15	42	58.93	10.144	
	16	37	65.46	8.265	
	17	32	66.81	7.739	
	18	2	61.50	19.092	
School year	9	78	55.22	10.252	$F(4, 238) = 11.478$, $p < 0.0005$
	10	66	57.76	12.882	
	11	28	58.07	9.764	
	12	32	66.66	7.707	
	13	39	66.03	8.203	
Test version	C	126	58.86	11.051	$F(1, 241) = 0.789$, $p = 0.375$
	E	117	60.15	11.54	

Table 4 also shows no significant difference in mean test score by test version. The standard deviations were very similar (11.051 and 11.549). There was no significant difference between the average scores (p = 0.375). There were significant differences by age and by school year. Because the group standard deviations were different, we used Welch's test which gave us the same overall result as the F test. Table 5 shows no significant difference by school decile.

Table 5: School decile and the vocabulary size of secondary school students

Decile	N	Mean	Std. Deviation	F
6	127	58.09	11.198	$F(2, 240) = 2.745$, $p = 0.066$
8	4	67.50	2.380	
9	112	60.77	11.363	
Total	243	59.48	11.289	

We decided to perform a multivariable regression using all the factors: age, school year, gender, test version and decile. However, statistics indicated that year and age were multicollinear (VIF > 7 and Tolerance < 0.2 for both), and therefore we did a stepwise regression. The only factor to enter was age, with a t statistic of 6.430, $p < 0.0005$, and an R^2 of 0.146. This suggests that age was a more significant predictor of test scores than school year.

The third research question: Are these sizes adequate for reading school texts? is answered in the following discussion.

Discussion

The results showing the means and ranges of vocabulary size across several age and school year levels reveal that even the learners with the smallest vocabulary sizes know many thousands of words and know a large proportion of the high-frequency and mid-frequency words of English. Some learners have very large vocabulary sizes which are well outside the normal ranges (see Table 1).

The results did not show regular increases from year to year and did not nicely fit an average increase of around 1000 word families a year, perhaps because the number of repetitions words becomes less as learners move further into the low-frequency words. The data does show increases in vocabulary size by age. Measuring the increases is complicated by the spread of birthdays across the school year, the lack of a one-to-one correspondence between age and school year, and probably most importantly, a reasonably large variation in vocabulary size at any particular age. Therefore, a longitudinal study may be needed to show regular age-related increases in size.

There were no gender differences in vocabulary size, just as Biemiller and Slonim (2001) also found no differences with younger learners. Fortunately the two test versions gave similar results, allowing us to combine the results from the two measures. The two tests are from a group of six which were tested for equivalence in another study (Coxhead, Nation & Sim, in press) and the results of the present study are supported by this previous study.

To make sense of these vocabulary size figures, we have to see how many words learners need to know to read their school texts. Table 6 presents the results of an analysis of the vocabulary of a secondary school science text used in New Zealand secondary schools (Coxhead, Stevens & Tinkle, 2010). The words in the text have been classified into five groups – high-frequency words (the most frequent 3000 word families of English), mid-frequency words (the 4th to 9th 1000 words of English) (Schmitt & Schmitt, 2012), low-frequency words (the remaining English word families up to the 25th 1000), words not in the twenty-five 1000 word lists, and proper nouns and marginal words. Column 3 shows that with knowledge of 9,000 word families along with proper nouns etc, learners will be familiar with 98.10% of the running words in the science text. That means that less than 2 words per 100 might be unfamiliar to them. Many of these possibly unfamiliar words may be technical words related to science. The most frequent words not in the lists but in the book are *protista*

(8 occurrences), monera (5), thigmotropism (5), Brownian (4), C6H1206 (4), limewater (3), monerans (3).

Table 6: The vocabulary levels in a secondary school science text at Year 9 (13 years old)

Frequency level	Coverage	Cumulative coverage	Families	Cumulative families
High-frequency	84.68%	84.68%	3000	3000
Proper nouns, letters etc	5.63%	90.31%		
Mid-frequency	7.79%	98.10%	6000	9000
Low-frequency	1.55%	99.65%	15000	24000
Not in the lists	0.35%	100.00%		
Total	100.00%		24000	

A sample from a science text (Hook, 1997) with the word level of the words marked below shows how vocabulary size relates to the vocabulary load of a text. In the sample, the words in the first 3000 (the high-frequency words of the BNC in Table 5) are not marked. {4} = 4th 1000, {5} = 5th 1000 and so on. {!}= not in the lists. {4}-{9}= mid-frequency, {10}-{24}=low-frequency. Low-frequency words and words outside the lists are in **bold**.

The Five Kingdoms

Related species are put into a {8}genus. similar **{10}genera** are put into a family. Similar families are put into an order. Similar orders are put into a class. Similar classes are put into a

{12}**phylum**, and finally {4}biologists have discovered over two million species so far. Each species must have a unique species name to avoid confusion. A species is given a two - word {31}Latin name. ({31}Latin was the language of {3}Ancient {31}Rome.) The scientific name for the dog species is {19}**Canis** {!}**Familiaris**, and {19}**Canis** {13}**lupus** is the name of the wolf species.

From this perspective, 13 year old secondary school learners with vocabulary sizes averaging around 11,000 word families easily have enough vocabulary to cope with the vocabulary of the science text. They will need to learn the unknown technical words but that is a normal part of learning a new subject area.

Limitations and future research

One limitation of this study is that the test does not measure the degree of knowledge of partly known words. That is, because the test uses a single format – multiple-choice – it cannot measure whether a word that is answered correctly in the test is known well or is just partly known. Further research is needed to determine whether, with a 1 in 200 sampling rate, the VST is fine-grained enough to pick up the new vocabulary that is learned each year. A second limitation is that it is not possible to generalize the results to all New Zealand secondary schools. However, mid-decile and high-decile schools contain some lower socio-economic students, and thus the range figures indicate that although means may be lower in lower-decile schools, the lowest scores will still show knowledge of many thousand word families.

Although this study looked at receptive vocabulary knowledge for reading, there was no reading measure used in the study. There were several reasons for this including the degree of

imposition on the schools and the students, and the need to focus on vocabulary measurement as this was the first time research on actual vocabulary size has been carried out in New Zealand schools. A future study could look at the vocabulary/reading comprehension connections. It may also be worth looking at the effect of ethnicity, although this is a fraught area.

Implications

Except in a very small number of cases, teachers need not be overly concerned with the general purpose vocabulary knowledge of their native-speaking secondary school students. They do need however to give attention to making sure that students continue to increase their vocabulary sizes by doing lots of reading and tackling the subject-related vocabulary they meet in their study. Most native-speaking secondary school students are well advanced in their knowledge of the 9,000 high-frequency and mid-frequency words of English and teachers need to continue to support their vocabulary growth through their subject-matter study.

Acknowledgements

References

Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253-279.

Beck, I. L., McKeown, M. G., & Kucan, L. (2005). Choosing words to teach. . In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and Learning Vocabulary: Bringing Research into Practice* (pp. 209-222). Mahwah, N.J.: Lawrence Erlbaum Associates.

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing, 27*(1), 101-118.

Biemiller, A. (2005). Size and sequence in vocabulary development. In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and Learning Vocabulary: Bringing Research into Practice* (pp. 223-242). Mahwah, N.J.: Lawrence Erlbaum Associates.

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*(3), 498-520.

Coxhead, A., Nation, I. S. P., & Sim, D. (in press). Creating and trialling six forms of the Vocabulary Size Test. *TESOLANZ Journal.*

Coxhead, A., Stevens, L, & Tinkle, J, (2010). Why might secondary science textbooks be difficult to read? *New Zealand Studies in Applied Linguistics, 16*(2): 35-52.

Dale, E., & O'Rourke, J. (1981). *The Living Word Vocabulary*. Chicago: World Book/Childcraft International..

D'Anna, C. A., Zechmeister, E. B., & Hall, J. W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behavior: A Journal of Literacy, 23*(1), 109-122.

Diller, K. C. (1978). *The Language Teaching Controversy*. Rowley, Mass.: Newbury House.

Elgort, I. (2013). Effects of L1 definitions and cognate status of test items on the Vocabulary Size Test. *Language Testing, 30*(2), 253-272.

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: differences by class and race. *Social Science Research, 33*, 464-497.

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 11*(4), 341-363.

Hart, B., & Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore: Paul H. Brookes Publishing Co.

Henning, G. (1991). *A study of the effects of contextualization and familiarization on responses to the TOEFL vocabulary test items*. Princeton, N.J.: Educational Testing Service.

Hoff, E. (2003). The specificity of environmental influency: socioeconomic status affects early vocabulary development. *Child Development, 74*(5), 1368-1378.

Hook, G. (1997) *Year 9 Science. New Zealand Pathfinder Series compact course books*. Auckland: New House Publishers Ltd.

Lorge, I., & Chall, J. (1963). Estimating the size of vocabularies of children and adults: an analysis of methodological issues. *Journal of Experimental Education, 32*(2), 147-157.

McKeown, M., & Beck, I. (2004). Direct and rich vocabulary instruction. In J. F. Baumann & E. J. Kame'enui (Eds.), *Vocabulary Instruction: Research to Practice* (pp. 13-27). New York: Guilford Press.

Nation, I. S. P. (2013). *Learning Vocabulary in Another Language*. (2nd edition). Cambridge: Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59-82.

Nation, I. S. P. (1993). Using dictionaries to estimate vocabulary size: essential, but rarely followed, procedures. *Language Testing, 10*(1), 27-40.

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Nation, I. S. P., & Webb, S. (2011). *Researching and Analyzing Vocabulary*. Boston: Heinle Cengage Learning.

Nguyen, L. T. C., & Nation, I. S. P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal, 42*(1), 86-99.

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Modern Language Journal, 42*(2), 282-296.

Prados, M. D. (2010). Gender and L1 influence on EFL learners' lexicon. In R. M. J. Catalan (Ed.), *Gender Perspectives on Vocabulary in Foreign and Second Languages* (pp. 44-73). New York: Palgrave Macmillan.

Read, J., & Chapelle, C. (2001). A framework for second language vocabulary assessment. *Language Testing, 18*(1), 3-32.

Scarcella, R., & Zimmerman, C. (1998). Academic words and gender. *Studies in Second Language Acquisition, 20*, 27-49.

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. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*(4), 484-503.

Seashore, R. H., & Eckerson, L. D. (1940). The measurement of individual differences in general English vocabularies. *Journal of Educational Psychology, 31*, 14-38.

Smith, M. K. (1941). Measurement of the size of general English vocabulary through the elementary grades and high school. *Genetic Psychology Monographs, 24*, 311-345.

Stahl, S. (2005). Four problems with teaching word meanings (and what to do to make vocabulary and integral part of instruction). In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and Learning Vocabulary: Bringing Research into Practice* (pp. 95-114). Mahwah, N.J.: Lawrence Erlbaum Associates.

Thorndike, E. L. (1924). The vocabularies of school pupils. In J. C. Bell (Ed.), *Contributions to Education* (pp. 69-76). New York: World Book Co.

Thorndike, E. L., & Lorge, I. (1944). *The Teacher's Word Book of 30,000 Words*. New York: Teachers College Columbia University.