DE GRUYTER OPEN

LIF – Language in Focus Journal, Volume 2, 2016, DOI: 10.1515/lifijsal-2016-0008

English Vocabulary Size in Adults and the Link with Educational Attainment

Sarah Wilson, Alex Thorne, Molly Stephens, Jessica Ryan, Sarah Moore, James Milton and Georgia Brayley

Swansea University, UK

Abstract

Modern estimates of English native speaker vocabulary size have concentrated on acquisition in childhood (e.g. Biemiller & Slonim, 2001) and among undergraduates (e.g. Milton & Treffers-Daller, 2013). There seems to be an assumption that vocabulary size is pretty stable in adults and that the estimates for undergraduates are likely to be applicable to the broader population, at least until age related decline begins sometime after the age of about 60 (Burke & MacKay, 1997). The study reported in this paper examines a cross-section of adults aged from 20 to over 60, and from graduate and non-graduate populations. The results suggest that the graduate population has a marginally larger vocabulary size than the non-graduate population and it is speculated that the difference is probably too small, and the vocabulary sizes too large, to drive educational differences. The graduate population also differs from the non-graduate In the graduate population vocabulary size appears to continue to grow with age, but while vocabulary size scores vary in the non-graduate population, the differences noted are not statistically significant after the age of 30.

1. Introduction

Attempts to measure the vocabulary knowledge of native speakers of English have a considerable history and there are plenty of reasons for wanting to makes these kinds of measurement beyond simple academic interest. Anglin (1993, 2) lists just some of these. Vocabulary knowledge provides the essential building blocks of language and without vocabulary neither language production nor language comprehension is possible. There are now theories which suggest other aspects of language such as grammar may be conditional on acquiring vocabulary knowledge of the right size and quality, (e.g. Bates and Goodman 1997). Not surprisingly therefore, vocabulary knowledge is at the heart of diagnostic tests for language and more general educational attainment. For example, vocabulary size in infancy is considered a strong predictor linguistic and cognitive abilities at four years (Feldman, Dale, Campbell, Colborn, Jurs-Lasky, Rockette & Paradise, 2005) and even at eight years (Marchman & Fernald 2008). Vocabulary size is also linked to the acquisition of competence in reading (e.g. Ouellette 2006) and, in turn, to success in school (e.g. Biemiller & Boote 2006; Bornstein & Haynes 1998; Tymms, Merrell & Henderson 1997).

A characteristic of these historical measurements is the huge variation in the scale of the sizes which are produced. Writing in 1877 Professor Holden estimates his own vocabulary size at 32,000 words (cited in Kirkpatrick, 1891). Kirkpatrick himself estimates the vocabulary of native speakers of English at between 10,000 and 100,000 words dependent on education. This is not even the largest estimate and Hartmann (1946) produced estimates of 200,000 words. At the heart of this variation is an uncertainty over how such estimates should be made and the earliest researchers were well aware of these problems and that resolving them was necessary if meaningful estimates were to be made which could be compared with each other. Thus, Kirkpatrick's paper identifies a variety of issues which researchers even today will recognise. For some of these we have reached assumptions and developed methodologies

which, we think, allow better, more useful and more reliable estimates to be made. But there are other issues which have as yet to be systematically addressed and where we are still working with assumptions about the scale of the mental lexicon. It is these issues of how vocabulary knowledge behaves among adults, and particularly outside formal education, and the relationship with educational attainment that this study seeks to investigate. Kirkpatrick's paper, while it is old, turns out to be a very good starting point for explaining the issues involved.

1.1 What's a word?

A principal concern in making an estimate of the number of words someone knows, is to decide what a word is and what should, therefore, be counted. Kirkpatrick reflects that he is counting base words and all inflected and derived word forms separately. This method gives him an estimate of his own vocabulary he can understand, but he reflects (p 107) that the previous estimate by Professor Holden does not describe its method, so it is unclear whether the two estimates are comparable. Earlier estimates largely seem to have followed Kirkpatrick in treating every different inflected and derived word form as a different word and this does much to explain the size of the estimates, by for example Seashore, which can be in the hundreds of thousands.

More recent estimates have followed a different approach and have chosen to count word families rather than word types to arrive at a useful calculation. A word family is comprised of a base word and its more frequent and regularly inflected and derived forms. So, *manage, manages, managing, management* and *mismanage* might all be counted as part of a single base word which in this case would be *manage*. It is thought this is a more useful approach since it appears that words are often stored as base words in the mental lexicons of native speakers and these are retrieved when needed and rules for inflection and derivation can then

be applied. Counting words this way probably gives a better idea of the scale of learning required to gain a fully working lexicon. A learner does not need to learn every derived and inflected form individually. If one form within a word family is acquired, then the others can be extrapolated from this. Estimates of vocabulary knowledge in second language learners often use a reduced form of the word family called a lemma where just the most frequent inflections are included since, it is thought, this most usefully reflects the state of knowledge of low level learners. Confusingly all of these are often referred to as just *words* in the literature.

Estimates made using a lemma or word family as the unit of count are, of course, smaller than those like Hartmann and Kirkpatrick who counted every word type as a different word. Typically these, more modern, estimates are between 10,000 and 20,000 words (properly word families). Golden, Nation & Read (1991), using a checklist test, suggest that educated native speakers know 17,000 to 20,000 words. D'Anna, Zechmeister & Hall (1991), using a multiple choice format, suggest their undergraduates know 16,785 words by recognition and 14,076 words which they could define, although D'Anna et al suspect the format will lead to over-estimation. Milton & Treffers-Daller (2013) suggest UK undergraduates enter university with a 10,000 word definition vocabulary and gain about 500 words each year while studying. While these estimates a smaller than the earlier ones, it will be noted that Goulden et al's estimate it still about double the size of that of Milton & Treffers-Daller. However, Milton & Treffers-Daller's test format is a more demanding one than in Goulden et al or D'Anna et al, since it requires that all test words are explained or defined, and this may explain the smaller size of their estimate. This is the second point which Kirkpatrick raises.

1.2 What's knowing a word?

Kirkpatrick is aware that he is making decisions over what to count as knowing a word which will affect the scale of his estimate. Some words he suspects he has never encountered before but he can deduce their meaning, or thinks he can, from their form or their component parts. Other words he recognises but thinks he could not explain or define. Still others he can define and use. He chooses to count all of these as known words and excludes form his count only the words he thinks he has never seen before. This approach gives him as inclusive an estimate, and as large an estimate, as he can manage and he knows a more demanding knowledge criterion would reduce this. He does not check to see whether his assumptions about his knowledge are correct.

More recent approaches appear to apply rather more stringent criteria for knowing. Nation's Vocabulary Size Test (VST) (Nation & Beglar, 2007) requires that a word form must be linked to a meaning if it is to be counted as known, and the test actually gives the meaning, from a choice of four, in a multiple test format. In addition to providing a prompt for the correct answer the multiple choice format also provides scope for guesswork among the testees. Both Gyllstad et al (2015) and D'Anna et al (1991), who also make vocabulary size estimates using this format, conclude that there must be over-estimating in size counts made this way. Goulden et al (1991) are more demanding in the knowledge requirement of their test and require that a definition must be produced for a word to be counted, but their choice of test format would seem to make this requirement less demanding that it at first appears. Goulden et al's estimate appears to be based on self-reporting so there is no check on whether test takers are correct in their assumptions of what they knew and this also must inflate the scores they produce. Milton & Treffers-Daller (2013) who did undertake a check on the Goulden et al checklist method concluded that scores might be inflated by as much as 20%.

1.3 How do you construct a test?

This leads to a further concern Kirkpatrick considers. Once you decide what a word should be and what knowing should be, how do you construct a test of this knowledge? Historical research, and some more recent estimates, make a choice of words based on a dictionary sample. Kirkpatrick is very aware that a dictionary may not contain all words in the language and will mis-estimate for that reason. There are specialist lexicons for many subject areas and a general dictionary will not contain these even if it is a large one such as the versions of Webster's that both Kirkpatrick and Goulden et al use. In a very real sense the size of the dictionary conditions the size of the estimate. There are other issues associated with dictionary sampling such as the way frequent words are likely to be over-represented by this technique which seem likely to undermine the accuracy of an estimate made using dictionary sampling.

The most recent test of vocabulary knowledge, especially in estimating foreign language knowledge in English (e.g. Meara & Milton's X-Lex, 2003; Nation's Vocabulary Size Test (VST) in Nation & Beglar 2007), use frequency information rather than dictionaries as the basis of word sampling. Goulden et al (1991) base the first half of their test on, effectively, words in the first 25,000 words in the Thorndike and Lorge (1944) frequency lists resorting to a dictionary sample to attempt an estimate of knowledge beyond this frequency range. They conclude that knowledge beyond 25,000 words in English is so slight that for practical purposes good estimates can be produced using the frequent, and frequency-based, sample only. Nation's VST samples from his own BNC based frequency lists.

1.4 Variation in vocabulary knowledge and the link with educational attainment

A conspicuous feature of Kirkpatrick's estimates is the high variation in vocabulary size among speakers of English he reports, and he ascribes much of this variation to the effects of

education. He expects, but does not demonstrate, that knowledge will vary by a factor of up to 10; graduates with vocabularies between 20,000 and 100,000 words, non-graduates with approximately 10,000 words. Recent studies have not addressed this issue in adult populations directly. There often seems to be an assumption that there will not be the sort of variation Kirkpatrick expects. Goulden et al (1991), for example, begin their study with the premise that there is an English native speaker vocabulary size, a general or universal size, which might provide a useful goal for learners of English as a foreign language. Wray (2008, 58) similarly suggests that, 'a central tenet of modern linguistics has been that … variation is secondary, and there is a primary level of equality across all individuals.' Milton and Treffers-Daller (2013) note how little variation in vocabulary size there is among the undergraduate university population they investigate. But their focus on learners in university is potentially missing the variation which may be occurring in the population as a whole.

We do have some information from studies of school age learners. Biemiller & Slonim (2001), for example, compare learners from kindergarten up to Grade 6 in what they term advantaged and normative populations. They observe differences in the two populations although nothing like the scale of difference that Kirkpatrick expects. The difference in vocabulary size is biggest in the youngest age groups where the advantaged group mean is about 50% larger than that in the normative population. However, the results from the study also suggests that these differences are mitigated by time and education. Thus, the two populations become more similar as the learners grow older and progress through school. By grade 6 the difference in the two groups' mean vocabulary sizes is only about 10% and the mean vocabulary size in both groups, about 9,000 words, is sufficient for full comprehension (Nation 2006). Potentially this narrowing of the difference between the groups might be the effect of the education process where a lexically rich school environment promotes learning particularly in the less advantaged population. Or it might be a ceiling effect since it is not

clear what the maximum possible score on the test is. Nonetheless, Biemiller & Slonim note the concern that Matthew effects may be occurring (Stanovich 1986) where it suspected that learners with large vocabularies on entry to school will prosper, develop their language skills disproportionately, and do better generally in this environment than children who enter school with smaller vocabularies. It is not clear from the Biemiller and Slonim data that this is occurring. Other studies (e.g. Shaywitz et al 1995) also fail to demonstrate that Matthew effects occur given the size of vocabulary that learners rapidly develop and the similarity in vocabulary size noted in the two groups. But there are echoes in the literature of Kirkpatrick's assumption that there will be a relationship between vocabulary size and educational attainment. Milton and Treffers-Daller note, for example, that in a university level population there is a moderate positive correlation between vocabulary size and the degree class obtained by students. In second language acquisition, lexical deficiencies have long been associated with underperformance in the education system (Milton 2012).

There may be a reason why, at least in English, all speakers' vocabulary knowledge will tend to converge at about the same level and this is the effect of Zipf distributions (for an explanation of Zipf distributions see Milton, 2009). Zipf draws attention to the way that in any language there are a small number of words which are highly frequent, a very large number of words which are infrequent and, in between these, a middling number of words which are of middling frequency. The effect of this is that a comparatively small number of words will dominate coverage; the proportion of text that these words occupy. In English, for example, the most frequent 3 words alone occupy 10% of most normal text. 50% of normal text is likely to be made up of only the most frequent 100 or so words. This in turn means that it is very easy to encounter and learn these highly frequent words, but much harder to encounter and learn words which are much less frequent. You cannot learn words that you do not encounter. At the outset of language learning, therefore, vocabulary size might be

expected to increase rapidly and Biemiller and Slonim's (2001) figures suggest children at school learn perhaps 600 or 700 words per year. But the more words you know, the fewer the words you encounter that you do not know, and it might be expected that vocabulary uptake will slow with time and, at some point, tail off. This may explain why university level speakers appear to grow their lexicons at a slower rate than children, maybe 500 words per year (Milton and Treffers-Daller 2013). It is not at all clear at what point vocabulary growth on this scale will stop but Kirkpatrick articulates an assumption that it will stop, he suggests the age of 30, and that that there is probably a steady state lexicon in adulthood.

Notwithstanding this argument, it is possible to suggest some very different ways in which vocabulary size in adults might behave. It is possible, for example, to suggest that adult lexicons may continue to grow after university age. University age students demonstrate growth at a fairly substantial rate and, other than the cessation of formal educational input, it is unclear why vocabulary learning should cease abruptly shortly after this point. It is certainly possible to suggest that the habits of reading and study developed in formal education might continue and vocabulary size might continue to grow with these activities even into advanced age. It is also possible to envisage the opposite development and to suggest that vocabulary size in adults might, after a certain point, diminish due to the various kinds of mental decline associated with age. Thus, Burke & MacKay (1997) suggest there will be aged related decline in vocabulary knowledge and use after the age of about 60.

Studies of the way vocabulary size develops during the course of adult life are missing from the literature so we have very little idea which of the scenarios outlined above might be occurring. We have little idea too of the way education might interact with these processes of vocabulary development in adults. Kirkpatrick's assumptions about the nature of the adult lexicon remain untested.

2. Aims and Objectives

The intention of this paper is to investigate the issues of vocabulary size in adults which Kirkpatrick raises but which more recent studies have yet to investigate in detail. It will investigate how vocabulary size in adults might vary and might change with age, in a general population both inside and outside university,

This broad aim asks two questions therefore:

- Do you get the huge variability in scores and specifically these big differences between college educated and non-college educated speakers which Kirkpatrick notes?
- Do speakers appear to hit a level of knowledge in adulthood after which the lexicon does not significantly grow?

In order to answer these questions there are a number of more specific objectives:

1. To collect vocabulary size estimates from English native speakers in these age ranges
    - University age
    - 30-39 years
    - 40-49 years
    - 50-59 years, and
    - 60+ years
2. To collect within each age group a mixture of both
    - university graduates, and
    - non-graduates
3. To construct a frequency profile from the data and calculate alpha scores to check the estimates made appear reliable and appears valid.
4. To compare the estimates, and the range of scores, of graduates and non-graduates to see whether the differences Kirkpatrick expected are demonstrable.

5. To compare the estimates of the different age groups to see whether vocabulary size in adulthood stabilises as Kirkpatrick expected.

3. Method

3.1 The test

The test selected to provide estimates of vocabulary was Nation's Vocabulary Size Test (Nation & Beglar, 2007). This test is based on word frequency lists, based on word families and drawn from Nation's own BNC word family lists (Nation, 2006) The test makes a principled sample of words across the first 20,000 most frequent words on this list with 100 test items . It therefore makes an estimate of vocabulary size out 20,000. Other tests based on word family counts (e.g. Goulden et al 1991, D'Anna et al (1991), and Milton and Treffer-Daller 2013) estimate adult vocabulary size to be less than 20,000 and Golden et al suggest that knowledge beyond 25,00 is so slight as to be not worth measuring. The test is multiple choice in format with a test word, in a minimal context, followed by 4 possible explanations of the test word from which the participant chooses one.

azalea: This **azalea** is very pretty.

a. small tree with many flowers growing in groups

b. light material made from natural threads

c. long piece of material worn by women in India

d. sea shell shaped like a fan

Test items are listed in frequency order with the most frequent items selected occurring first. In order to test for a frequency profile and for test reliability, these items were sub-divided into four sub-tests of 25 question each; items 1 to 15, 26 to 50, 51 to 75, and 76 to 100, called VS1, VS2, VS3 and VS4.

The test is, of course, susceptible to guesswork but has the merit of being comparatively straightforward for the participants, relatively quick, and accessible to all levels that are likely to be encountered in adults.

The tests was administered in paper formal under the supervision of the authors.

3.2 The participants

The participants were an opportunistic sample of adult native speakers available to the authors of this paper: colleagues, friends and family members. 210 participants completed Nation's test and their age and graduate status as is shown in Table 1.

Table 1: Study participants

participants	**graduate**	**Non-graduate**
18-21	18	22
30-39	21	17
40-49	12	22
50-59	21	26
60+	17	32

4. Results

4.1 Test performance

The means scores in the 4 sub-tests indicate that a frequency profile is present with vocabulary knowledge focussed among the more frequently occurring words. The mean scores on the four sub-tests are given in Table 2

Table 2: Mean scores on the vocabulary sub-tests

Sub-tests	Mean score (max 25)	SD
vs1	24.33	1.03
vs2	22.27	2.23
vs3	20.31	3.73
vs4	16.67	4.82

A Friedman test confirms the frequency profile (χ^2 = 445.59, sig<.001).

Use of the four sub-test to make a calculation for Cronbach's give the result α = .785 which is taken to indicate that the test is performing reliably.

4.2 Vocabulary Size and education

A comparison of the mean vocabulary sizes for the graduate and non-graduate groups are given in Table 3.

Table 3: Mean vocabulary sizes in graduate and non-graduate groups

Groups	N	mean	SD
Graduate group	91	17404.40	1879.00
Non-graduate group	119	16121.01	2001.73

The difference between the means is statistically significant (t = 4.727 sig<.001).

4.3 Vocabulary size and age – non-graduates

For ease of explanation it is expedient to divide the results for age into the non-graduate and graduate groups. The mean vocabulary sizes obtained by the non-graduate participants in each age band are given in Table 4.

Table 4: Mean vocabulary sizes by age in non-graduates

age	N	mean	SD
university	22	14354.55	1826.71
30-39	17	15670.59	1878.75
40-49	22	17081.82	1532.70
50-59	26	15930.77	1885.47
60+	32	17068.75	1665.29

An ANOVA confirms there is an effect for age in these results(F (4,85) =9.921, sig.<.001). A Tukey test indicates what this effect is and the results of the Tukey tests are summarised in Table 5.

Table 5: Summary of Tukey tests results for vocabulary scores in non-graduate group

age	N	Subset for alpha = 0.05	
		1	2
university	22	14354.55	
30-39	17	15670.59	15670.59
40-49	22		17081.82
50-59	26		15930.77
60+	32		17068.75
Sig		.091	.058

4.4 Vocabulary size and age – graduates

The mean vocabulary sizes obtained by the graduate participants in each age band are given in Table 6.

Table 6: Mean vocabulary sizes by age in graduates

age	N	mean	SD
university	18	14944.44	1634.63
30-39	21	17152.38	1231.10
40-49	12	17866.67	1262.99
50-59	21	18352.38	1132.97
60+	17	19070.59	640.09

An ANOVA confirms there is an effect for age in these results($F(4,85) = 29.898$, sig.<.001). A Tukey test indicates what this effect is and the results of the Tukey tests are summarised in Table 7.

Table 7: Summary of Tukey tests results for vocabulary scores in graduate group

age	N	Subset for alpha = 0.05			
		1	2	3	4
university	22	14944.44			
30-39	17		17152.38		
40-49	22		17866.67	17866.67	
50-59	26			18352.38	18352.38
60+	32				19070.59
Sig		1.000	.435	.773	.430

5. Discussion

5.1 Test performance and effect for guesswork

The test looks like it is performing well and is likely to give meaningful results. This conclusion is notwithstanding concerns over the use of a test that involves a prompt and appears prone to guesswork. The results display good reliability and the expected frequency profile, the absence of which would show the test is not performing properly. The method may merely be inflating the scores rather than undermining the entire validity of the test.

The overall results look similar to the more recent estimates for vocabulary size in native speakers of English based on what are considered to be the most appropriate ways of assessing vocabulary size. This means that estimates are made of knowledge of word families. The estimates of about 16,000 or 17,000 words in this study compare with just under 17,000 for recognition vocabulary in undergraduates noted by D'Anna et al (1991) and 17,200 words for educated native speakers In Goulden et al (1990). These results are, however, larger than the Milton &Treffers-Daller scores for undergraduates of 10,000 to 11,000 words.

This difference with the Milton & Treffers-Daller scores may be due to the occurrence of guesswork and over-estimation which the test formats used. D'Anna et al (1991), who use a similar multiple-choice test format to VST reflect that there must be some over-estimation in their results from this cause although they do not attempt any adjustment. Nation (2012) likewise recognises the scores are likely to be inflated but see no need for adjustment. Part of his rationalisation maybe that he views guesswork as drawing on sub-conscious knowledge which potentially shows partial knowledge of the target words (cited in Gyllstad et al, 2015, 298). Goulden et al (1990) also use a format prone to over-estimation since there is no

systematic check the accuracy of answers in their results. In the case of results drawn from VST, however, it is possible to estimate the potential effect of guesswork on the results using information given by the number of wrong answers a testee produces. In VST's format there is a 1 in 4 chance of guessing the right answer, so knowing the number of wrong answers allows us to estimate the number of right answers potentially gained by guesswork, and to adjust scores accordingly. This calculation gives the results in Table 8

Table 8: Adjusted vocabulary sizes for graduate and non-graduate groups

Groups	N	VST means		Adjusted means	
		mean	SD	mean	SD
Graduate group	91	17404.40	1879.00	16539.19	2505.34
Non-graduate group	119	16121.01	2001.73	14829.69	2668.46

The results now look very similar to D'Anna et al's (1991) 16,000 words for recognition and 14,000 definition vocabulary size. They are closer to, but still greater than, Milton & Treffers-Daller's result (2013). However, once a recalculation for each age group is made, so learners of the same age can be compared, then the comparison with the Milton &Treffers-Daller scores looks more convincing. The age groups scores, are summarised in the Tukey test results for graduate and non-graduate groups which are given in Tables 9 and 10.

Table 9: Summary of Tukey tests results for adjusted vocabulary scores in graduate group

age	N	Subset for alpha = 0.05		
		1	2	3
university	18	13103.7039		
30-39	21		16203.1746	

40-49	12		17155.5556	
50-59	21		17803.1746	17803.1746
60+	17			18760.7843
Sig		1.000	.051	.462

Table 10: Summary of Tukey tests results for adjusted vocabulary scores in non-graduate group

age	N	Subset for alpha = 0.05	
		1	2
university	22	12475.7576	
30-39	17	14235.2941	14235.2941
40-49	22		14574.3590
50-59	26		16091.6667
60+	32		16109.0909
Sig		.089	.060

Even adjusting for guesswork, the VST format still includes a prompt for the correct answer and some differences in score are likely to be a result of the different test methods used.

All of these estimates of vocabulary size are smaller, of course than historical estimates and this is due to the change in the unit of count which is now based on a word family. These figures suggest an uptake of 1-2 words a day up to university age, a figure explainable by explicit learning. Standard figures in the literature, which suggest childhood uptake of some 3,000 words a year (for example, Graves 2006; Nagy 1988; Marzano 2004; and Duke & Carlisle 2011), really can be laid aside as grossly misleading as to the nature of the language development that is going on among these young learners.

5.2 Vocabulary Size and education in adults

Tables 3 and 8 show the graduate group in this study have a higher mean vocabulary size than the non-graduate group and the difference is statistically significant. The t score is not affected by the adjustment for guesswork. The age groups in this study are not perfectly balanced for size, but this probably does not influence the conclusion to be drawn from this observation. In line with Kirkpatrick's suspicion and Miton & Treffers-Daller's (2013) observation, vocabulary size and educational attainment interact and are related in some way. The scale of the difference between graduate and non-graduate speakers is nothing like what Kirkpatrick suggested, even allowing for the different unit of count he uses. The difference is more like 10 or 15% rather than the two-fold or ten-fold difference he describes. This kind of difference is very similar to the difference noted by Biemiller & Slonim (2001) between the advantaged and normative populations in his grade 6 students. It is tempting to think that this kind of modest difference between the vocabularies of the educational high-flyers and the more normal populations, remains similar from this age. The standard deviations are also fairly modest in scale so there is not enormous variation in the scores of the two groups, rather, there is enormous overlap between them. The two groups tend to appear more like each other than they appear different and the difference in the means is driven by a longer tail of low scores, below 10,000 words, which is present in the non-graduate population. It is not clear whether these subjects are genuine low scorers or have failed to engage meaningfully with the test or have misunderstood the task, as Nation (2007) describes.

It is not clear from the data, of course, what is driving the difference in scores between the two populations. The explanation, as in the Matthew effect, that differences in vocabulary size are the cause of differences in educational attainment is possible but the data merely shows a link and not causality. On balance this explanation does not appear likely in this

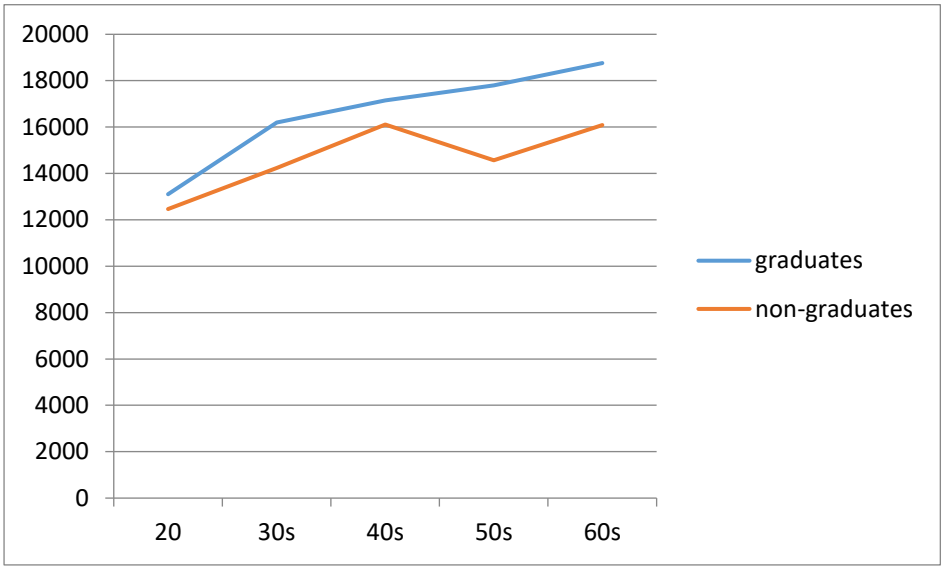
case. The scale of the difference in the vocabulary sizes are relatively small and appear likely to make a difference of substantially less than 1% in coverage, and therefore in the speakers capacity for comprehension. The vocabulary sizes of in both groups is of a scale which should provide very good coverage and comprehension. Typically the vocabulary sizes are well in excess of the 8,000 - 9,000 which Nation (2006) describes as the basic requirement for 98% coverage and full understanding. It is hard to imagine how differences of the scale described here are going to affect comprehension and therefore educational attainment. These differences are nothing like what is observed in L2 learners where learners may have vocabularies several thousand words less than Nation's 8,000 to 9,000 words and much diminished coverage as a consequence, and where vocabulary deficit really does have an impact on educational performance.

Perhaps the explanation of the relationship between vocabulary and educational attainment might look in other directions. Possibly it is differences in educational experience which drives differences in vocabulary. Potentially, an interest in education will promote reading in specialist subject areas, and extensive reading of this kind is something that is often thought to promote vocabulary growth. Possibly, too, a third factor might be driving both and even a range of other abilities, as in Spearman's G factor (Spearman, 1927).

5.3 Vocabulary size and age

The way vocabulary size varies with educational attainment continues when the relationship of vocabulary size and age are examined. The way vocabulary size interacts with age and educational attainment is summarised in Figure 11.

Figure 11: vocabulary size interaction with age and educational attainment



The data suggests that in each age group the mean for the graduate group is higher than the mean for the non-graduate group but also it suggests that vocabulary progress is also different for the graduate and non-graduate groups. The graduate group's vocabulary size continues to grow throughout adulthood. The rate of growth does not appear to be as great as that of native speakers at university. These adults appear to acquire 100-150 words a year every year, rather than the 500 words the university students in Milton & Treffers-Daller's (2013) study displayed. The way this group continues to grow vocabulary size with age may very likely help explain why Goulden et al's (1991) vocabulary size estimates are higher than those in Milton & Treffers-Daller (2013). Goulden et al's sample was comprised of older speakers, faculty members and graduate students, whilst in Milton & Treffers-Daller the sample was exclusively undergraduates. This helps challenge the idea of an ideal native speaker-like vocabulary size which non-native speakers to aim at for things like study abroad. It appears that some native speakers at least, can increase the size of their lexicons by several thousand words after leaving university.

In one sense the rate of growth seen in this group, of 100-150 words per year, that might confirm the Zipf idea that vocabulary growth in English is likely to tail off as it increases because as you grow a progressively bigger and bigger vocabulary there are fewer words left to add from normal language. But there might be a ceiling effect becoming evident here since by the time these learners approach 60 the frequency profile they display suggests there is vocabulary knowledge beyond the 20,000 word limit of the test and the absence of this limits the rate of progress they can show.

In the non-graduate group it has not been demonstrated that vocabulary size increases after about the age of 30. Our best interpretation of the results we have for the non-graduates is that, as Kirkpatrick suggested, vocabulary does not systematically increase with age but hits something like a ceiling at around 15,000 words. While vocabulary size does vary between the age groups at 30 and beyond, the differences in the means are not statistically significant.

It is not completely obvious why the graduate and non-graduate groups should differ in this way although it is possible to speculate. Maybe the habits of reading and study, which are thought to promote vocabulary growth, continue after university. Or perhaps there is a genetic predisposition for growing a larger vocabulary in some speakers (Wray 2008). Other factors by which individuals vary are genetically influenced, for example a person's height, so perhaps vocabulary learning might be influenced the same way. There is no sign of decline in native speakers' lexicons in the results collected here but populations in advance of 60 where, it is suggested, these changes may occur have not been examined in this study.

6. Conclusions

The sample size examined in this study is small in relation to the population as a whole so the results drawn from it are suggestive rather than definitive. Nonetheless, there results fill a gap in our knowledge as to progress of vocabulary knowledge after university age. Vocabulary

size does not continue to grow continuously throughout adulthood. Vocabulary growth in English tails off after about the age of 20 there appears nothing like the scale of growth that can be seen through childhood and adolescence. In some speakers it seems that the mental lexicon may hit a steady state at least in terms of size sometime after this age. This contributes to a body of evidence that shows native speaker English vocabulary size, while still large, is nothing like the scale that is often thought and which is often discussed in the literature.

The results suggest there are vocabulary differences associated with educational level both in the volumes that are known, graduates tend to know more words than non-graduates, and in the way vocabulary size appears to progress in adulthood among graduates but plateaus among the non-graduates. The differences in the size of the two groups' vocabularies is nothing like as large as anticipated by Kirkpatrick. A principal conclusion of this observation has to be that the volumes of vocabulary known by the graduate and non-graduate groups, and the likely effect on coverage and comprehension, makes vocabulary size an unlikely cause of educational differences. The validity of the thinking behind the Matthew effect is challenged by this observation. Rather than thinking in terms of the educational effect produced by vocabulary size differences, we should be thinking of finding factors which might influence both vocabulary size and educational success.

These tentative conclusions would benefit from some reworking in further studies. Ideally these studies should include a bigger and more rigorous sample, to discount any possible effect particularly of the differences in groups size encountered in this study. A different choice of test might usefully be made to avoid the guesswork which must be a feature of the responses to VST. Further research might also usefully aim at unpicking the factors that might influence both vocab knowledge and educational attainment.

References

Anglin, J. M. (1993). *Vocabulary development: A morphological analysis.* Monographs of the Society for Research in Child Development, 58(10), Serial No. 238.

Bates, E. and Goodman, J. (1997). On the inseparability of grammar and the lexicon: evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes, 12*(5/6), 507-584

Biemiller, A. & Boote, C. (2006). An effective method for building vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44–62.

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,* 498–520.

Burke, D., & MacKay, D. (1997). Memory, language and ageing . *Philosophical Transactions of the Royal Society: Biological Sciences,* 1845-1856.

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

Duke, N. & Carlisle, J. (2011). The development of comprehension. In Michael L. Kamil, P. David Pearson, Peter A. Afflerbach, & Elizabeth B. Moje (eds.), *Handbook of Reading Research,* Vol. 4. New York: Routledge, 199–228..

Feldman H. M., Dale, P., Campbell, T., Colborn, D. K., Kurs-Lasky, M., Rockette, H. &. Paradise, J. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development*. 76(4), 856–868.

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

Graves, M. F. (2006). *The Vocabulary Book.* New York: Teachers College Press. Gyllstad et al (2015

Hartmann, G. W. (1946). Further Evidence on the Unexpectede Large Size of Recognition Vocabularies among College Students. *Journal of Educational Psychology 37*, 436-439.

Kirkpatrick, E. (1891). The number of words in an ordinary vocabulary . *Science 18 (446)*, 107-108.

Marchman, V. A. & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, 11(3): 9–16.

Marzano, R. J. (2004). *Building background knowledge for academic achievement: Research on what works in schools*. Alexandria, VA: Association for Supervision and Curriculum Development.

Milton, J. (2009). *Measuring second language vocabulary acquisition* . Bristol: Multilingual Matters.

Milton, J. (2013) Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist and B. Laufer (eds.) *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis.* EUROSLA monograph 2, 57-78.

Milton, J., & Treffers-Daller, J. (2013). Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review, 4 (1)*, 151-172.

Meara, P. & Milton, J. (2003) *The Swansea Levels Test.* Newbury; Express.

Nagy, W. E. (1988). *Teaching vocabulary to improve reading comprehension.* Newark, DE: International Reading Association.

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

Nation, P. (2007) Fundamental issues in modelling and assessing vocabulary knowledge. In Daller, H., Milton, J. and Treffers-Daller, J. (Eds.) *Modelling and Assessing Vocabulary Knowledge.* Cambridge; Cambridge University Press, 33-43.

Nation, P. (2012) *The Vocabulary Size Test Instructions and Description.* Retrieved 14/06/2016 from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf.

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

Ouellette, G. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *The Journal of Educational Psychology* 98, No.3, 554–566.

Shaywitz, B., Holford, T., Holahan, J., Fletcher, J., Stuebing, K., Francis, D., & Shaywitz, S. (1995). A Matthew effect for IQ but not for reading: Results from a longitudnal study . *Reading Research Quaterly* , 894-906.

Spearman, G. (1927) *The Abilities of Man.* Macmillan; Oxford.

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy . *Reading Research Quarterly*, 360-407.

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

Tymms, P., Merrell, C. & Henderson, B. (1997). The first year at school: a quantitative investigation of the attainment and progress of pupils. *Educational Research and Evaluation*, 3, 2, 101–18.

Wray, A. (2008). Genes and the conceptualisation of language knowledge. *Genomics, Society, and Policy*, 4(1), 58-73.