

Assessing the vocabulary load of text

— Implications for timed reading —

Andrew Atkins

Abstract

This paper examines how students' vocabulary levels as determined by the Vocabulary Size Test (Nation & Beglar, 2008) can be used in conjunction with the RANGE program (Nation & Heatley, 2002) to predict the suitability of passages used in timed reading practice. The paper further discusses what could be done to facilitate reading fluency practice for these students, and finally uses previously obtained scores from timed reading to corroborate predictions.

Keywords : timed reading, vocabulary load, reading fluency, RANGE program, Vocabulary Size Test

Introduction

1. Background to the study

The main goal of timed reading is to increase reading fluency by regularly reading texts that have a similar lexical load. When the text has been read and the time taken has been noted, students then proceed to answer a small number of questions to check their understanding. The ideal situation when writing or choosing texts for timed reading is to have passages that are closely matched in terms of lexical load to the lexical knowledge of students. It seems logical to presume that as the lexical load of texts increase relative to a student's lexical knowledge there will be a decrease in reading speed and also a decrease in comprehension. In order to eliminate confounding variables, texts must be carefully chosen, as without uniformity, it would not be possible to claim that gains in reading fluency are being achieved. This paper attempts to exemplify how texts can be matched to students.

2. Vocabulary and reading

Webb & Nation (2008) provide a framework for evaluating the vocabulary load of written texts using the Vocabulary Size Test (VST) (Nation & Beglar, 2007) in conjunction with the RANGE program (Nation & Heatley, 2002). As well as using this framework, data previously collected for a timed reading research project will be used to corroborate the predicted effects of vocabulary.

Vocabulary has been found to be the best gauge of whether a text will be understood (Laufer & Sim, 1985), but it is naïve to assume that this is the only variable that will affect comprehension. A number of researchers have attempted to design instruments that test vocabulary levels. However, knowing a word is not a straightforward concept. Nation (2001) states that words fit into many interlocking systems and levels and that there are many things to know about any word, therefore measuring lexical knowledge in a holistic fashion is very complex. These levels and systems include word families, productive knowledge, receptive knowledge, written domain, and spoken domain (Nation, 2001).

3. Timed reading

The literature on timed reading is relatively sparse in contrast to that concerned with vocabulary. Tanaka and Stapleton (2007) using a combined approach found that studying shorter reading passages in class, i.e. timed reading, when combined with Extensive Reading (ER) texts outside of class produced statistically significant increases in reading speed and comprehension. In recent studies using this type of activity, Chung & Nation (2006) and Crawford (2008) found that timed reading was an effective means of developing reading rates. Both of these studies however failed to examine comprehension to a satisfactory degree. Looking only at the time taken to read a passage seems to be rather illogical when surely anyone can finish reading a passage in a short time without really comprehending the meaning. Nation and Malarcher (2007) advocate timed reading, but fail to provide much guidance as to how texts should be matched to student abilities. Chung & Nation (2006) offer that "...such a course needs to be within a controlled vocabulary so that learners do not face any lexical difficulties which may interrupt their reading."

Nation (2001) suggests a vocabulary coverage of 98% is ideal for students to learn words, while Laufer (1989) and Liu & Nation (1985) in previous studies proffered 95% as the necessary level of coverage for learning to occur. These figures seem conservative when timed reading is what is being considered rather than reading to learn vocabulary. Reading fluency gain is the variable of interest, and it seems that a slightly less conservative figure of around 90% coverage would be enough. However Hunt & Beglar (1998) suggest 80% comprehension is ideal and Crawford (2008) provides a rough guideline that 70% understanding is the target students should aim at when reading. Surely both suggestions are too liberal to be effective.

Timed reading is in the receptive domain, and receptive knowledge is usually concerned with recognising words although some argue that the productive-receptive dichotomy is more of a scale or cline because meaning has to be produced to comprehend an item (Webb, 2009). It seems to be widely accepted though, that receptive learning is more likely to produce gains in receptive

knowledge than productive learning and vice versa (Griffin & Harley, 1996; Mondria & Wiersma, 2004; Waring, 1997).

4. Assessing vocabulary

Nation and Beglar's (2007) Vocabulary Size Test (VST) seems to provide the information required for the analysis and is easy to administer in a short period of time. The VST provides data for each 1000-word level of vocabulary, and not just the total number of words known. This is important for assessing the percentage of words known by learners.

Knowing approximately the words that students are likely to know at a receptive level provided by the VST and comparing this with word profiles of texts from the RANGE program helps teachers to identify words that are likely to be unknown. The VST makes some assumptions that remain to be tested, in that the test uses the unit of word families and assumes that if students know a headword then they will have receptive knowledge of the other members of the word family. This assumption seems to hold for L1 situations (see Bertram, Laine, & Virrkala, 2000 for a discussion), but it remains to be seen if it is applicable for L2 situations. The validity and reliability figures for the test are also unknown.

Method

1. Participants

Ninety-seven first year university students (34 women and 63 men, mean age 18.7 years) in five intact classes of varying ability volunteered to participate in the study. All participants were in their first semester of tertiary education and taking a non-elective Oral Communication class meeting for ninety minutes twice a week. The participants were from a number of different faculties. Classes are streamed in to five levels at the beginning of the first semester, with level one being the least proficient, and level five being the most proficient. Two of the classes used in this study were from level 2 ($n = 16$, $n = 14$). Two of the classes were from level four ($n = 23$, $n = 22$). The final class in the study was from level five, the most able level ($n = 22$).

2. Materials

The first five levels of the Vocabulary Size Test were administered to students at the start of a regular class in the tenth week of their first semester. Only the first five levels of the fourteen-level test were used as these were felt to be the most relevant for the purpose of this investigation. The test consists of ten test items for each one thousand-word level; therefore a fifty-item test was used

for this study. Timed reading, like Extensive Reading is designed to promote fluency and as a result the lexical load is intentionally low and unlikely to exceed the most common five thousand words in English.

A secondary analysis of data obtained from a research project on reading fluency gains was also carried out as it was felt that this would corroborate the results of the VST data. Participants reading scores for the first five readings of Reading for Speed and Fluency Level 1 (Nation and Malarcher, 2007) were used. All forty of the readings in the text are exactly 300 words long and supposed to be written within a controlled vocabulary load. The texts in this case were all about animals and therefore didn't require any special schematic knowledge to comprehend them. The texts were administered in five consecutive classes at the beginning of the spring semester, 2009. Two of the classes had not completed all five of the readings.

3. Design and Procedure

The first issue to address in the process of finding suitable texts for students to read is to assess their lexical knowledge. This is done to discover the *coverage* for the texts. Coverage is a statistic, represented as a percentage, which shows the quantity of the lexis that is likely to be understood in a text. The total number of word families a student knows is of secondary interest as it is possible to know more infrequent items than frequent items. What must be assessed is the students' vocabulary knowledge at different levels in order to find gaps in knowledge that need to be pre-taught. The VST (Nation and Beglar, 2007) attempts to determine how many of the words a student knows at each one thousand-word level of the British National Corpus 14 (BNC 14) lists. If a student gets 9 out of ten items correct at one level they are thought to have knowledge of 900 out of 1000 word families at that level (Webb & Nation, 2008). The lists are based on frequency data from a large spoken corpus rather than a written corpus as this was thought to be most closely related to the abilities of L2 learners.

Results and Discussion

1. The Vocabulary Size Test

The first five level version of the VST was administered to five intact, streamed groups of students when they met within a two-day period. The classes were all at one private university and the students were freshmen from various faculties. The streaming was decided by scores participants obtained on a placement test developed in the university. One of the groups tested was from the highest level, level five. Two of the groups were from level four, and the remaining two groups were

Table 1
Mean Raw Scores on First 5 Levels of the Vocabulary Size Test

Class Number (level)	<i>n</i>	Vocabulary Size Test					Combined Total
		1000	2000	3000	4000	5000	
A (5)	22	8.50	6.36	5.95	5.41	4.23	30.45
B (4)	22	8.55	5.86	5.77	5.36	3.09	28.64
C (4)	23	8.87	5.78	5.26	5.04	3.30	28.26
D (2)	14	8.07	4.36	4.14	3.93	2.29	22.79
E (2)	16	7.63	3.25	4.06	4.38	1.75	21.06

from level two. The mean scores obtained by each group on the VST are shown in Table 1.

Mastery of any level of vocabulary is said to be 9 or 10 out of a possible 10 items. It can be seen that none of the means for any of these groups have reached this level.

It can be seen that some of the groups, despite six or more years of formal English education, have what Hunt & Beglar (2005) call an impoverished vocabulary, especially after the first 1000-word level. In Class E the word families known at the second 1000-word level seem to exemplify the *impoverished* phenomenon. Even for the members of the highest ability class, class A, they are still a long way from mastery for all but the first 1,000-words. Interestingly, class C has the highest mean for the 1000-word level, although not officially the most proficient class. If we are generous and round to the nearest whole number, the three most proficient classes can be considered to have mastered the first 1000-word level, but only that. When we look at the second 1000-word level and beyond, if we use only this data, we should have little confidence that students will know any lexis in these levels. If 95% or 98% coverage is the ideal as Laufer (1989) and Nation (2001) respectively suggest, then the only means of achieving this is to use texts that do not stray from the first 1000-word level or perhaps even the first 500-word level. Texts that contain words at the 2000-word level and above would surely prove to be much more demanding for students in all classes.

2. Vocabulary profiling

The next step in the Webb & Nation (2008) framework is to use the RANGE program to analyse texts to find the vocabulary profile. For the purposes of this investigation the texts were the first five readings taken from *Reading for Speed and Fluency 1* (Nation & Malarcher, 2007). The readings were being used in a larger study and it was of great interest to see whether the analysis of the texts would shed light on the reading success achieved for the passages. The vocabulary profiles can be seen in table 2, below.

Reading 1 appears to be one of the least demanding texts as 98% coverage is reached at the 2000-

Table 2
Vocabulary Profiles for the First 5 Texts in Reading for Speed and Fluency 1

Reading	Percentages of Items BNC levels 1 to 5					Combined Total %
	1000	2000	3000	4000	5000	
1	92.54	6.72	0.00	0.74	0.00	100.00
2	86.33	8.63	1.44	3.60	0.00	100.00
3	80.92	12.21	3.82	0.76	0.00	97.71
4	76.07	8.55	4.27	1.71	1.71	92.31
5	90.76	8.40	0.84	0.00	0.00	100.00

word level, and 90% coverage is reached the 1000-word level. Reading 2 appears more demanding due to 98% coverage being reached at the 4000-word level, although 90% coverage is reached at the 2000-word level. Reading 3 reaches 98% coverage after the 5000-word level, although 90% is reached at the 2000 word level. Reading 4 looks to be the most challenging of the five readings. 98% coverage is not attained in the first 5000 words, with 90% coverage occurring at the 4000-word level. Reading 5 is much less challenging than the previous one. 90% coverage occurs in the first 1000 words, and 98% coverage is achieved in the 2000 word level.

There are some confounding variables in this study because these readings were in almost all cases the participants' first experiences of timed reading and as a result they were unfamiliar with the format of the test. If this were not the case, and participants were seasoned timed readers, then it would be reasonable, assuming only mastery of the 1000 word level, to assume that the order of difficulty for the readings (easiest to most difficult) would be Reading 5, or Reading 1, then Reading 2, Reading 3, and Reading 4. If we take the arbitrary level of 90% coverage as a further gauge of likely difficulty, then the order remains almost unchanged, at Reading 1, Reading 5, Reading 2, Reading 3, and finally Reading 4.

3. Timed reading coefficients

To assess difficulty of the passages by an alternative means and to corroborate the predictions of the Webb & Nation (2008) framework, the mean reading coefficient obtained for each passage by each group will be examined. Obtained prior to this study, the reading coefficient is an alternative measure of comprehension used by Carew, Exton, Buckley, McGaley, and Gibson (2005) and can be seen as a combined indicator of speed and comprehension, in other words a measure of reading skill. The coefficient is the time in seconds taken to read the 300-word passage divided by the score on a five item multiple-choice test. For example, if a student read a passage in 60 seconds and scored 3 out of 5 on the test they would be given a coefficient of 20 ($60 \div 3=20$). Reading speed

Table 3
Mean Class Reading Coefficients for the First 5 Texts in Reading for Speed and Fluency 1

Class	Readings				
	1	2	3	4	5
A (5)	30.68	29.27	30.48	37.66	27.86
B (4)	37.12	29.47	37.30	36.57	30.6
C (4)	36.80	27.35	29.05	33.42	23.63
D (2)	51.83	47.38	–	–	–
E (2)	56.18	62.64	–	74.77	55.49

and comprehension have been shown by Utsu (2004, 2005) to improve over time, and although in this case the time period involved is only 15 days, the coefficients show there is no clear linear improvement in the coefficients. Although the questions that follow each passage are designed to be of a comparable difficulty, the actual comparative difficulty remains untested. Table 3 shows the mean class coefficients obtained for each reading passage.

Table 3 provides some data that is consistent with the predictions of the Webb & Nation (2007) framework, although there appears to be some data that is not explainable by vocabulary load alone.

The data provided by the most able group (class A) is the most consonant with predictions. Reading 4 has the highest coefficient (37.66), and was therefore read least fluently by the students, whereas Reading 5 has the lowest value (27.86), which means it was read most fluently and suggests it was the easiest. There is little difference between the coefficients for readings 1, 2, and 3, although as alluded to previously Reading 1 was the students' first attempt at timed reading and this may have inflated the coefficients.

For Class B, the first of the level 4 classes, there is little difference in the times for Reading 1, Reading 3, and Reading 4. As suggested before the results for Reading 1 may be suspect, so we can propose at least that Reading 3 and Reading 4 were most challenging, and Reading 2 and Reading 5 were least difficult. This is similar to what was predicted.

The second of the level four classes (class C) has outperformed the most able class in all but Reading 1, and if we exclude that reading, the order of difficulty is the same as class A and almost the same as class B.

The remaining two classes do not have complete data, but for class E the data appears to follow the same pattern as the groups discussed above, with Reading 4 the most difficult and Reading 5 the least challenging.

With all of this information at hand, in hindsight it appears that more should have been done

to simplify Reading 4 and Reading 3 to bring them in line with the other readings they were to be compared to. There are a number of options that could have been considered that seem appropriate. Webb & Nation (2008) suggest five possible means to rectify potential coverage issues, however the use of dictionaries for timed reading tasks would take too much time, as would glosses of the more difficult words. The fastest of the remaining three options is to simply eliminate the text, and with the short time taken to analyse a text by the RANGE program, this will in many cases be the most efficient use of a teacher's time. Another way of increasing coverage is to pre-teach important vocabulary, and this method is probably what will be done before Reading 4 and Reading 3 are given to students in the next cohort. The remaining option for teachers is to modify a text by changing the more difficult items identified by the RANGE program for items that would be more easily understood. This would perhaps be a little too time consuming in some cases for it to be a practical solution for teachers.

With the lower level learners who were examined in this study, the 1000-word levels are not as precise as would have been desirable. Perhaps 500-word levels would have given a more precise view of the knowledge of the students. This would mean altering the word lists used in the RANGE program, which appears to be relatively simple to do. It would also mean adapting the VST to make it more precise, and this would be a more challenging task.

Conclusions

The framework proposed by Webb & Nation (2008) appears to offer a relatively fast and efficient way to predict problematic lexical items in any written text and as a result gives the teacher time to take action before the text is presented to students. In the context of timed reading, a teacher can either pre-teach the vocabulary that is unlikely to be known in a text, simplify the text, or discard the text in favour of a more suitable one.

There are still a number of other variables that are not looked at by using only RANGE to examine a text and that have an effect on text comprehension. Schema is a very slippery variable to measure, and perhaps judgement based on experience is the only way to gauge if students will possess the necessary world knowledge to understand a passage. Previously acquired specialised knowledge is also another factor that will affect reading fluency and cannot realistically be predicted.

The data provided by the timed reading coefficients seems to generally support the predictions of the framework. The framework will certainly be utilised in the future and it will probably be most useful with lower level classes. Pre-teaching vocabulary seems to be a promising option for texts

that are slightly too challenging, whereas texts that appear to be too challenging will be abandoned in favour of more suitable readings.

VST scores combined with data obtained from the RANGE program appear to offer an effective means of evaluating texts for timed reading as well as any other written text that is going to be used in the classroom. It adds to an experienced teacher's instinct for what will *work* and will help a less experienced teacher make up for a lack of instinct.

References

- Bertram, R., Laine, M. & Virkkala, M.M. (2000). The role of derivational morphology in vocabulary acquisition: get by with a little help from my morpheme friends, *Scandinavian Journal of Psychology*, 4, 2-15.
- Carew, D., Exton, C., Buckley, J., McGaley, M., & Gibson, J.P. (2005) Preliminary Study to Empirically Investigate the Comprehensibility of Requirements Specifications. In *Psychology of Programming Interest Group 17th Annual Workshop (PPIG 2005)*, pp 182-202, University of Sussex, Brighton, UK.
- Chung, M., & Nation, I.S.P. (2006) The effect of a speed reading course. *English Teaching* 61, 4: 181-204.
- Crawford, M. J. (2008). Increasing reading rate with timed reading. *The Language Teacher*, 32(2), 3-7.
- Griffin, G. F., & Harley, T. A. (1996). List learning of second language vocabulary. *Applied Psycholinguistics*, 17, 443-460.
- Hunt, A. & Beglar, D (1998) Current research and practice in teaching vocabulary. The Language Teacher Online. Retrieved February 2, 2009 from: <<http://www.jalt-publications.org/tlt/files/98/jan/hunt.html>>
- Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language*, 17(1), 23-59.
- Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special Language: From Humans Thinking To Thinking Machines*, (pp 316-323). Clevedon: Multilingual Matters.
- Laufer, B., & Sim, D. D. (1985). Taking the easy way out: non-use and misuse of clues in EFL reading. *English Teaching Forum*, 23(2), 7-10, 20.
- Liu, N., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, 16(1), 33-42.
- Mondria, J., & Wiersma, B. (2004). Receptive, Productive, and Receptive +Productive L2 Vocabulary Learning: What difference Does It Make?, In *Vocabulary In a Second Language*, edited by Bogaards, P. & Laufer, B., pp. 79 – 100. Amsterdam: John Benjamins.
- Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Nation, I.S.P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts [software]. Downloadable from <<http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>>
- Nation, P. & Malarcher, C. (2007). *Reading for Speed and Fluency, Book 1*, Seoul: Compass Publishing.
- Tanaka, H., & Stapleton, P. (2007). Increasing reading input in Japanese high school EFL classrooms: An empirical study exploring the efficacy of extensive reading. *Reading Matrix*, 7(1), 115-131. Retrieved February 2, 2009 from <http://www.readingmatrix.com/articles/tanaka_stapleton/article.pdf>
- Utsu, M. (2004). Timed Readings no riyou to sono kouka[Timed Readings and its effects on students]. *Bulletin of Yonezawa Women's College of Yamagata Prefecture*, 39, 31-37.

- Utsu, M. (2005). Timed Readings no riyou to sono kouka2 [Timed Readings and its effects on students (Part II)]. *Bulletin of Yonezawa Women's College of Yamagata Prefecture*, 40, 27-34.
- Waring, R. (1997). A study of receptive and productive learning from word cards. *Studies in Foreign Languages and Literature (Notre Dame Seishin University, Okayama)*, 21(1), 94-114.
- Webb, S. (2009, June 20 & 21). Presentation on Vocabulary, given at Temple University Japan, Osaka Campus.
- Webb, S. & Nation, I.S.P. (2008). Evaluating the vocabulary load of written text. *TESOLANZ Journal*, 16, 1-10.

テキストの語彙負荷の査定

——速読に対する効果——

アンドリュー アトキンズ

要 旨

本論では、Timed Reading（速読）の練習に使用するパッセージの適否を予測する目的で、Vocabulary Size Test (Nation & Beglar, 2008) により判定した学習者の語彙レベルを、RANGE プログラム (Nation & Heatley, 2002) と併せて使用する方法について検討している。本論ではさらに、同じ学習者に対し、リーディングの流暢さを高めるためにできることは何かについても述べ、最後に、Timed Reading により得られたスコアを使用して、予測された適否の確認を行っている。

キーワード：速読, 語彙負荷, リーディングの流暢さ, RANGE プログラム, 語彙サイズテスト