Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension

Batia Laufer
University of Haifa

Geke C. Ravenhorst-Kalovski
College of Yezreel Valley and University of Haifa
Israel

Abstract

We explore the relationship between second language (L2) learners' vocabulary size, lexical text coverage that their vocabulary provides and their reading comprehension. We also conceptualize "adequate reading comprehension" and look for the lexical threshold for such reading in terms of coverage and vocabulary size. Vocabulary size was measured by the Levels Test, lexical coverage by the newest version of Vocabulary Profile and reading comprehension by a standardized national test. Results show that small increments of vocabulary knowledge contribute to reading comprehension even though they hardly improve text coverage. We suggest two thresholds: an optimal one, which is the knowledge of 8,000 word families yielding the coverage of 98% (including proper nouns) and a minimal one, which is 4,000–5,000 word families resulting in the coverage of 95% (including proper nouns).

Keywords: reading in L2, vocabulary, lexical threshold, adequate reading comprehension, vocabulary size, lexical coverage

Since Alderson's (1984) seminal article "Reading in a Foreign Language: A Reading Problem, or a Language Problem?" research evidence has accumulated to suggest that both general reading skills and proficiency in the L2 are important for reading in foreign languages (Grabe, 2004; Perkins, Brutten, & Pohlmann 1989; Pichette, Segalowitz, & Connors 2003). Most researchers agree that general reading skills can operate most efficiently when the reader possesses a critical mass of L2 knowledge referred to as the threshold of L2 knowledge (Bernhardt & Kamil, 1995; Carrell, 1991; Clarke, 1980; Cziko, 1978; Lee, 1997). The general reading skills include inferring the meaning of unknown words from context, awareness of text and argument structure, distinguishing between main and peripheral information, between explicit and implicit material, etc. L2 knowledge includes lexical and grammatical knowledge. We are not aware of any research that tried to define the syntactic threshold of reading. However, studies have been conducted to investigate the relationship between L2 vocabulary knowledge and success in reading comprehension, and, subsequently, to find the vocabulary threshold, that is the minimal vocabulary that is necessary for "adequate" reading comprehension. Information on lexical

threshold is important for second language education, particularly for courses with reading as their main focus, since such information may help teachers and course designers in setting vocabulary goals and designing lexical syllabi. Thus, for example, if the lexical threshold is found to be 7,000 word families, then by the end of a course in academic reading, students should try to reach this vocabulary size if they intend to engage in reading authentic academic material.

Regarding the relationship between vocabulary and reading, most researchers agree that vocabulary is a good predictor of reading, if not the best (Bernhardt & Kamil, 1995; Laufer, 1992; Nation 2001, 2006; Qian, 1999, 2002; Ulijn & Strother, 1990). As for the lexical threshold of reading, opinions vary. Before we survey literature on lexical threshold, we will clarify three key terms whose understanding is essential in any discussion of reading and vocabulary. These are: lexical coverage, sight vocabulary and "adequate" comprehension. If we say that a group of words, for example, the Academic Word List (AWL), which consists of 570 different words (Coxhead, 2002) provides coverage of 10% of an academic text, it means that 10% of an academic text (10% of all word tokens) consists of the AWL words. From the reader's perspective, this means that the knowledge of the AWL will assure the comprehension of 10% of the vocabulary in an academic text. We can also refer to coverage as the percentage of words that a reader understands. If, for example, readers have reached 95% text coverage, this means that they understand 95% of the running tokens of the text. The term "sight vocabulary" is used to refer to words whose meaning is so familiar to a person that they can be understood out of context. Therefore, when encountered in a text, these words are recognized and decoded quickly and without any cognitive effort. For example, if readers encounter the word "hypothesis" in a text and the word is in their sight vocabulary, they do not need to rely on the surrounding context to comprehend its meaning. Hence, a large sight vocabulary contributes to reading fluency and frees cognitive effort for higher level reading processes that is engaging with comprehending the text content and its implications (Mezynski, 1983; Pulido, 2007; Segalowitz, 2007). Lexical text coverage and the reader's sight vocabulary size are, therefore, two related factors of lexical threshold. The larger the sight vocabulary, the higher is the coverage of a text. Therefore any statement about the text coverage that enables comprehension will inevitably bear on how much vocabulary the learner should acquire to read "reasonably well." But how well is "reasonably well"?

The term "adequate" or "reasonable" comprehension has no clear definition since it may refer to different levels of comprehension in different contexts, and yet statements about lexical thresholds depend on what is considered adequate. Different university disciplines may require different levels of reading proficiency on an identical university entrance test due to different quantities of reading material in English as L2. Moreover, the same discipline may require a higher reading standard for a higher academic degree. For example, Business Administration students in Israel are tested on reading in English by an Israeli test when they start their Bachelor's degree and by the Graduate Management Admission Test (GMAT) as an entrance to the Master's program. Obviously, adequate reading proficiency, and consequently lexical threshold, is different in each of the above cases. Since lexical threshold is the focus of the paper, we will now survey studies that investigated it, focusing on the interaction between coverage, learners' sight vocabulary size and reading comprehension that was considered adequate in each study.

The first attempt to relate reading comprehension to lexical coverage was made by Laufer (1989). The coverage was calculated by learners' self report, underlining the unknown words in the text, adjusted for "bluffing." The latter was checked through a translation test which included most of the infrequent words of the text and subsequent comparison with the underlining in order to disclose discrepancies between self report and translation. The lexical coverage was the total number of words in the text minus the real number of unknown words, converted into percentage. Reading comprehension was measured by a reading test and an "adequate" comprehension was set at a score of 55%, which at the time of the study happened to be a passing score of the English for Academic Purposes course which the participants were enrolled in. The results showed that at 95% coverage there were significantly more participants with a score of 55 and above than with a score below 55. This was not true for other levels of coverage even though at every coverage level, some learners received passing reading grades. This means that the threshold as reflected in lexical coverage is of a probabilistic nature. Adequate comprehension may happen below it, but the chance is low. In this early study "adequate" comprehension was set at a rather low score of 55. However, most educators, including ourselves, would probably not be satisfied with such a low score.

Hu and Nation (2000) also investigated the relationship between lexical coverage and reading comprehension. They created four coverage groups (80%, 90%, 95%, 100%) by replacing some text words with non-words in the below 100% groups. The other words of the text belonged to the 2,000 most frequent vocabulary. They used two comprehension tests and defined "adequate" comprehension as the score that most learners in the 100% coverage group received; 12 correct answers out of 14 on a multiple choice test, (i.e., around 85.7% and a score of 70 out of 124 on a written recall test, [i.e., 56.5%]). If we average out the two scores, we receive 71%. However, a written recall test requires learners to demonstrate their writing ability as well. This may explain the relatively low "adequate" test score. Maybe a more valid comparison between the measures of comprehension in Laufer and Hu and Nation would involve comparing their multiple choice tests only. In this case, the difference between what was considered "adequate" in the two studies would be much larger, 55% as opposed to 85.7%. Hu and Nation found that nobody could read adequately at 80% of coverage, some learners could at 90% and 95% coverage, but they were in the minority. The conclusion of the study is that 98% is the lexical coverage for adequate comprehension. However, the two different coverage suggestions above, of Laufer and Nation, relate to two different reading scores considered to represent "adequate" comprehension. Hence, both suggestions could be correct depending on what level of comprehension is expected. Nation (2001) reporting on the two studies says: "The probabilistic threshold is 98%. With this coverage almost all learners have a chance of gaining adequate comprehension. If, instead of adequate comprehension, a standard of minimally acceptable comprehension is applied (as Laufer did in her study), then 95% coverage is likely to be the probabilistic threshold" (p. 147).

The second factor related to the threshold issue has to do with finding out the vocabulary level, in terms of the size of sight vocabulary, which learners need to reach in order to read adequately. This can be approached in two ways: by examining the coverage that words of different frequency levels provide to texts in representative corpora, or by testing students on text comprehension and relating different reading scores to learners' vocabulary size. The most comprehensive and up-to-date study taking the first approach is that of Nation (2006), in which

he trialed 14 frequency lists developed on the basis of the British National Corpus, and checked what coverage they provided of a variety of spoken and written texts. The summary of the data shows that in written texts, the first thousand most frequent word families will provide a coverage of 78% to 81%, the second thousand an additional 8% to 9%, the third thousand 3% to 5%, the fourth and fifth thousand 3%, the sixth to ninth thousand 2%, and the tenth to fourteenth thousand, less than 1%. Proper nouns cover 2% to 4% of written texts. All the other words, which do not appear in the lists, can account for 1% to 3% of the texts (ibid: 79, table 14). The lower word coverage figures provided by the most frequent words are probably characteristic of most difficult texts, while the highest figures of easier texts. If we average out these figures, we can see that readers with a knowledge of 3,000 word families and the proper nouns in a text can reach a coverage of 95% (79.5 + 8.5 + 4 + 3). To reach 98% coverage, (i.e., an additional 3%), they will need to know words from the 4th and 5th frequency bands as well. Nation says that if we take 98% as the ideal coverage, then 8,000–9,000 word-family vocabulary is needed for dealing with written texts. This is a safe estimate based on the lower coverage figures stated above. In a text with the higher coverage figures, it may be possible to reach the 98% with a knowledge of 5,000 words and proper nouns as shown above. We do not suggest that teachers or learners rely on 5,000-word knowledge for 98% coverage. The more vocabulary learners know, the safer they are in reaching the appropriate coverage. But we cannot rule out the possibility that the average vocabulary coverage figures—3,000 for 95%, and 5,000 for 98%—and the knowledge of proper nouns may often enable learners to read as required in Laufer (1989) and Hu and Nation (2000), respectively.

The second approach to finding out the threshold vocabulary level, by testing learners on text comprehension and vocabulary size, was taken by Laufer (1992). In this study, 92 learners took two standardized reading tests and a vocabulary test, either the Vocabulary Levels Test (Nation, 1983) or the Eurocentres vocabulary tests (Meara & Jones, 1989). On the basis of the vocabulary tests, they were divided into vocabulary size groups and comprehension scores were examined for each vocabulary level group. "Adequate" comprehension was set at a score of 56% on the reading test. The minimal vocabulary level at which there were more readers than non-readers was found to be 3,000 word families. A linear regression analysis showed that a 3,000 vocabulary level would predict a reading score of 56%, a 4,000 level would result in an additional 7 points, (i.e., 63%, and a 5,000 level would yield a reading score of 70% [assuming the relationship between the two variables was linear]). The study concludes with practical implications for syllabus designers to set vocabulary goals on the basis of the comprehension level expected of learners. Thus, here again, we can see how the notion of vocabulary threshold is contingent upon what is considered "reasonable" or "adequate" comprehension.

If we look carefully at all the studies mentioned so far, they seem to converge surprisingly well in their results, even though, on the surface, they seem to suggest different thresholds. Laufer (1989) found that at 95% coverage most participants could receive a score of 55% on the reading test. In 1992, she found that a vocabulary level of 3,000 word families could assure this reading score. However, in the same study, she also found that to receive a score of 70%, learners would need to know 5,000 word families. Hu and Nation (2000) suggest that 98% of coverage is required for "adequate" comprehension which is set at 71%, being the average of the two comprehension tests. The corpus data in Nation (2006) show that it is possible to reach 98% coverage with 5,000 word families and proper nouns, and 95% coverage with 3,000 word

families and proper nouns. It is often quoted that Laufer suggests a 3,000-word knowledge and 95% coverage as the threshold, while Nation suggests 8,000- to 9,000-word knowledge and 98% coverage. This oversimplification does not take into account two factors. First, the estimate of 8,000 to 9,000 word families is the highest estimate for 98% coverage, while in fact this coverage can be achieved by a lower range of vocabulary. Second, a 95% coverage and 3,000-word knowledge will suffice only for minimal comprehension around 55%, which is not claimed to be universally adequate. A higher threshold vocabulary, 10,000 words, was suggested by Hazenberg and Hulstijn (1996) for a comprehension score of 70%. Yet it is hard to compare these results to the other studies because this study was carried out with the Dutch, not English vocabulary. Besides, the list of the threshold vocabulary was created on the basis of a dictionary. Usually dictionaries list lexical items, not word families. The families figure would be considerably lower.

The studies surveyed above relate reading comprehension scores to learners' lexical coverage (Laufer, 1989; Hu & Nation, 2000), or reading scores to learners' vocabulary level (Laufer, 1992), or corpus data on word frequency to lexical coverage (Nation, 2006). The connection between the three factors: coverage, vocabulary knowledge and reading comprehension is via extrapolation. If learners with 95% coverage received the minimal reading score of 55%, and learners with the knowledge of 3,000 word families received the same score (in another study), then 3,000 word families would probably assure 95% coverage. If corpus analysis showed that 8,000 word families cover 98% of a text, and if learners read adequately when they understood 98% of a text, then adequate reading comprehension would require the knowledge of 8,000 word families.

In the present study, we combined data on the lexical coverage of several academic texts, learners' vocabulary level and reading comprehension scores of academic English. Since the texts that we analyzed for coverage were of similar nature and practically identical difficulty to the texts learners were examined on, we could find out how the reading scores on the tests were associated with coverage and with learners' vocabulary knowledge. Hence, the study contains elements from Laufer (1989, 1992), Hu and Nation (2000) and Nation (2006). Moreover, this study was carried out with a much larger sample (745 students) than any of the previous studies. The aim of the study was to find out how the two factors of lexical threshold, text coverage and vocabulary size of the learners, were related to reading comprehension. Of particular importance to us was this relationship at several reading comprehension levels which could be considered "adequate" in different educational circumstances.

Method

Participants

A total of 745 students took part in this study. Most of them (735) were students in an academic college in Israel studying in different departments and taking a course in English for Academic Purposes, a course that aims at improving students' reading comprehension in English. Prior to college, they studied English for eight years in high school. Many of them did not start college studies immediately after school, but after several years of other activities. 495 participants were

native speakers of Hebrew, 167 of Arabic, and 73 of Russian. Before entering the college, they took an English Psychometric Exam, which will be described in the *Materials* section. Suffice it to say here that this is a standardized test of reading comprehension and its maximum score is 150. The learners' level of English was not uniform. Their test scores ranged from 75 to 133. The college placed each student in one of five course levels on the basis of the score they obtained in the test. The lowest level (level one) included learners who received 75–84, level two included students with a score of 85–94, level three included 95–103, level four included 104–115, and level five included scores of 116–133. Ten additional participants were learners whose score on the test ranged between 134 and 146. This high score gained them an exemption from the course in English. The wide range of reading scores enabled us to examine the effect of coverage and vocabulary size on a variety of levels of reading comprehension.

Materials and Measures

The three variables under investigation: reading comprehension, lexical coverage and vocabulary size were measured, respectively, by a standardized test of English reading, a vocabulary profiler of texts, and the Vocabulary Levels Test.

Learners' Reading Comprehension

Reading comprehension in English was tested by the English part of the Psychometric University Entrance Test. This test is designed, administered and marked by experts in testing who work at the National Institute for Testing and Evaluation (NITE) in Israel. The psychometric test has been in use for over 20 years and has been validated with over a hundred thousand test takers. It consists of three parts: logical thinking, verbal intelligence in L1, and reading comprehension in English, and is taken under time pressure. The questions of the entire test have a multiple choice format. As the test is supposed to predict academic success of the candidate and English is the language of academic texts, the English part of the test includes about 60 questions that tap the learner's comprehension of academic English. Some questions focus on the understanding of words, some on the understanding of sentence structure, some on the understanding of global textual information, both explicit and implicit. (A passage from a practice test is in the Appendix. Information about the psychometric test can be found on http://www.nite.org.il). All the candidates receive a total score on the Psychometric Entrance Test, and also three separate scores on each part of the test. Our data on learners' reading comprehension consist of the scores on the English part of the psychometric test. The maximum score on this test is 150. The average national score over the years has been 104 and the standard deviation 24. Since different learners were examined on the psychometric test at different points of time, depending on their time of registration, they took different versions of the test. Yet the different versions are claimed to be of a very similar difficulty. Slight differences in difficulty may occur in some versions due to questions, or text, including slight lexical differences. These differences are factored into the scores that the respective test takers receive. Hence, the learners' reading scores reflect comprehension of similar texts regardless of the test version they took.

Learners' Vocabulary Size

Learners' vocabulary size was measured by the revised version of Nation's (1983) Vocabulary

Levels Test (Schmitt, Schmitt & Clapham, 2001). The test contains items from the 2,000, 3,000, 5,000, and 10,000 most frequent words, 3,000, 5,000, 10,000 and academic vocabulary. Each 'word' represents a word family, (i.e., the word, its inflections and common derivations. Each frequency level includes 30 items except the academic vocabulary section, which includes 36 items). Each correct answer receives one point, an incorrect answer or no answer receives 0 points. The participants were not given the 10,000 level as it was considered far too difficult in view of the background they had in English. They did the academic vocabulary section, but we did not include it in our data. The academic vocabulary list includes words from the second to fifth frequency levels. Hence, it cannot be considered a separate level from the other levels.

The Vocabulary Levels Test is not a precise measure of vocabulary size but a tool for examining a learner's knowledge of items from particular levels. Therefore our estimates of size are approximate. They were calculated on the basis of the 2,000; 3,000; and 5,000 parts of the test. Since the scores of the different levels are implicationally scaled (Read, 1988), we filled in the missing 4,000 level by averaging the scores received on the 3rd 1,000 and 5th 1,000. The score at each frequency level represents an approximate knowledge of 1,000 words, except the first 2,000; where the score represents knowledge of 2,000 words. If, for example a learner received 28 on the second 1,000, 22 on the third, and 8 on the fifth, his score would be 28+28+22+15+8=101. (The figure 28 appears twice as it represents 2,000 words, (i.e. two frequency levels), while the other scores represent 1,000 words each. The figure 15 is the average of 22 and 8.) Since each frequency level has 30 items, the maximum score, which represents knowledge of 5,000 words, would be $30 \times 5 = 150$. The score in our example would represent $101 \times 5,000/150 = 3,366$ word families.[1] Since this is only an approximation, we assigned this learner to the K3 level group in our sample. The learners were divided into vocabulary levels as follows: anyone who received a score representing between 500 and 1,500 words was placed at K1 level, those with a score representing 1,500–2,500 words were placed at K2 level, those with 2,500–3,500 words were at K3 level, those with 3,500–4,500 words were at K4 level, and learners with vocabulary above 4,500 words were placed at K5 level. This division meant that at each vocabulary size level, the average vocabulary size of the learners was around the respective thousand. The test was administered to the students at the beginning of the academic year in their respective classes.

The additional students who were exempt from studying English were tested individually. These students received the same test as the rest (up to K5 level). However, because of their success in the psychometric test, we expected these students to have larger vocabularies. Since in the Levels Test the K5 section is followed by K10, which makes it difficult to fill in the K6–K9 gaps, we preferred to test them on the 6K, 7K, and 8K sections of a new Vocabulary Size Test (Nation & Beglar, 2007) which was validated by Beglar (2009). In this test, every frequency level includes 10 items and each item represents knowledge of 100 words. Their final score was calculated by adding up the results of the Levels Test (K5) and the Size Test (6K, 7K, 8K).[2]

Lexical Coverage of Texts

The original tests that our participants took are not available yet to the general public. The older test versions, however, which are of similar difficulty to the later versions, have been released and are often used by prospective students for practice towards the Psychometric Entrance Test.

These are three tests that comprise 19,037 words altogether. Each test includes 6 texts with comprehension questions and additional questions involving the comprehension of selected words and structures. These were the tests that we analyzed by lexical profile available at Paul Nation's website (http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx) and at Tom Cobb's site (http://lextutor.ca).

We used a new version of the vocabulary profiler which matches a text to 20 vocabulary frequency lists constructed on the basis of the British National Corpus. The program provides an output which shows what percentage of the text is covered by each word frequency list. Words that are not in the 20,000 most frequent vocabulary appear on the output as "off list" words. A special function, added recently by Tom Cobb on "Lextutor" allows us to analyze a text in such a way that all its proper nouns which are personal and geographical names (e.g. Richard, Moscow) are recategorized to appear in the first thousand most frequent words. Such analysis rests on the assumption that these nouns do not belong to the lexicon of a particular language, and if the reader is not familiar with them, the resulting comprehension problems cannot be attributed to lexical ignorance. However, proper nouns which are regular words, (e.g. Statue of Liberty, Applied Linguistics) appear in their respective frequency lists. We analyzed each test twice: once with the above new function and once without it. In the former case, most of the names appeared in the K1 list. In the latter case, they were listed among the "off list" words. Then we calculated the difference in the number of the tokens of proper names between the two "off list" lists and converted it into percentage out of the total number of tokens. This gave us the percentage of proper names in the text.

Results

First, we will present the analysis of the English Psychometric Tests in terms of the percentage of coverage of each BNC (British National Corpus) frequency list. In Table 1, we show the coverage of 10 lists. The proper nouns have not been recategorized by the special function to be included in the K1 list, but are distributed among all lists. Thus the proper nouns which are personal and geographical names are included in the "off list" words.

Table 1. *Coverage of the English psychometric tests by BNC frequency lists* (Proper names in the "off list")

Frequency level	Coverage % Test 1	Coverage % Test 2	Coverage % Test 3	Average cumulative coverage
K1	80.15	75.91	79.58	78.58
K2	9.39	10.04	7.92	87.67
K3	2.54	3.11	3.24	90.56
K4	2.21	2.58	2.35	92.81
K5	0.74	1.09	1.27	94
K6	0.80	1.13	0.66	94.8
K7	0.32	0.48	0.77	95.4
K8	0.46	1.27	0.90	96.3
K9	0.11	0.44	0.14	96.53
K10–K20	1.10	1.00	0.88	97.5
Off list	2.19	2.97	2.32	~100

Table 2 presents the percentage of proper names calculated following their recategorization as the first 1,000 words. If we assume that the proper names are familiar to the learner, then the 95% coverage can be achieved with a knowledge of 4,000 words, which cover almost 93% and the proper nouns which cover an additional 2.1%. 98% coverage can be reached by knowledge of 7,000–8,000 and the proper nouns.

Table 2. *Coverage by proper nouns*

	Test 1	Test 2	Test 3	Average
Proper names	2	2.48	1.8	2.1

Since the aim of the paper was to find out the relationship between vocabulary size, coverage and reading scores, we will now present the combined data on the coverage data from Table 1 with the data on learners' vocabulary size and the reading score. As mentioned earlier, in the section on measuring vocabulary size, we divided the learners by intervals of 1,000 words. In Table 3, we replace the BNC lists with learners' vocabulary size. If, for example, 5,000 words cover 94 % of a text, then learners with a knowledge of 5,000 words can understand a similar percentage of this text. As mentioned in the section on measuring reading comprehension, the raw scores of reading are out of 150. The percentile scores in the table were retrieved from the adjustment tables of NITE which show how the raw scores are converted into standard scores and percentiles on the basis of national results. Because of the small number of the top students, they were put together even though their reading scores ranged from 134 to 146 and vocabulary size from 6,000 to 8,000.

Table 3. *Vocabulary size, lexical coverage and reading comprehension* (Maximum reading comprehension score = 150)

Approximate vocabulary size	Lexical coverage	Percentile on the psychometric test	Reading score: Mean (*SD*)	No. of students
1,000	78.58	50%	83 (6)	109
2,000	87.67	53%	90 (7.8)	199
3,000	90.56	66%	102 (8.9)	204
4,000	92.81	72%	111 (9.4)	200
5,000	94	83%	122 (8.3)	23
6,000	94.8			
7,000	95.43	91%–99%	138 (4)	10
8,000	96.3			

We also performed a linear regression that aimed at showing how the vocabulary score can predict the reading score. (This analysis did not include the top 10 students.) The analysis showed that the intercept was 69.88, slope –0.01, multiple R .64 and the regression formula as follows: Reading score=69.98+0.01× vocabulary size. First, the R square shows that the amount of variance in reading that can be ascribed to vocabulary is 64%. In other words, the correlation between the two is .8. Second, on average, an increase in 1,000 words raises the reading score by 10 points. This would be precise if the relationship between two variables were linear. But Table 3 reveals a slightly different picture regarding the relationship between vocabulary and reading. Moreover, the table shows how all three variables (coverage, vocabulary and reading) are related to one another. The difference between the knowledge of 1,000 and 2,000 words results in the highest difference in coverage (9.09%), but it yields the lowest difference in the reading score (7 points). Additional 1,000 words increase the coverage by less than 3%, but they increase the

reading score by 12.3 points. Each additional 1,000 words increase the coverage by less and less. Yet the increase in the reading score does not become smaller. The reading score increases by 10 points for each additional 1,000 between 3K and 5K knowledge. Beyond 5K an additional 1,000 words increase the reading score by ~17 points (the 5 learners with 6K vocabulary had an average reading score of 139). Participants with 7K and 8K vocabulary did not score higher, but this is not very important as all of the ten were able to read independently. As mentioned earlier, the coverage figures in Table 3 do not include the proper nouns, which constituted 2.1% of the texts. Therefore the precise coverage of the texts at each vocabulary size level should be increased by 2.1%. For example, learners with a 5K word knowledge can understand 94+2.1=96.1% of the text. Figure 1 shows graphically the different effects that an increase in vocabulary has on text coverage and on reading comprehension.
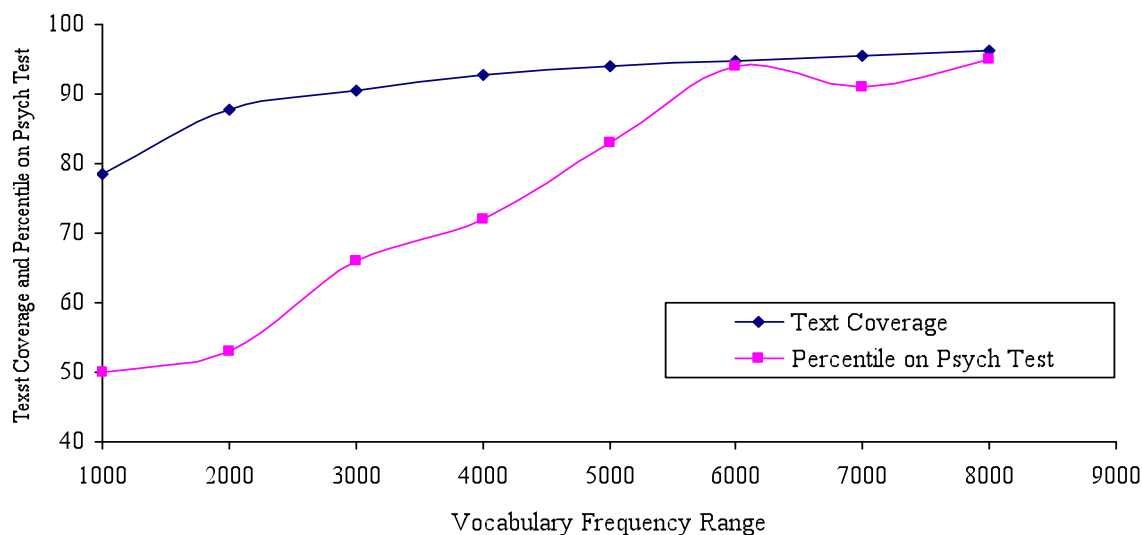

Figure 1. Text coverage and reading scores in relation to vocabulary frequency range.

Discussion

The aim of the study was to explore the relationship between text coverage, vocabulary size of the learners, and reading comprehension, particularly "adequate" reading comprehension. The vocabulary-coverage relationship found here is in accordance with corpus-based studies (e.g. Nation, 2006), which show that the less frequent the vocabulary, the smaller the portion of text coverage. However, the relationship between coverage, vocabulary and reading implies that even a small increase in lexical coverage (1.19% from 4K to 5K) may be just as beneficial to reading as a larger increase in coverage (2.25% from 3K to 4K). Interestingly, a small improvement in coverage (0.8%) from 5K to 6K, or 1.3% from 5K to 7K was associated with the best improvement in the reading score (17%).[3] There could be two possible explanations of this result. Some low frequency words may at times belong to the key words of an academic text and may therefore be crucial for comprehension. Another reason may be related to the superior automaticity of decoding that the learners with larger vocabularies possess. Laufer and Nation (2001) explored the relationship between vocabulary size and speed of decoding word meaning and found that speed on a particular word frequency level increased only when learners' vocabulary size progressed far beyond that level. This means that the participants with a large

vocabulary read more fluently the frequent words in the text, which may have given them an overall advantage over the learners with a smaller vocabulary, who had not yet attained a similar level of fluency.

The finding that the knowledge of additional infrequent vocabulary contributes to reading comprehension even though it hardly improves text coverage raises a question regarding the selection of vocabulary for teaching. Corpus-based studies claim that infrequent vocabulary does not provide a good return for the learning effort because it adds very little text coverage. However, if in spite of the small addition of coverage, there is a substantial improvement in reading comprehension, then the benefit of teaching infrequent vocabulary may be greater than previously thought.

We will now address the thorny question: what lexical coverage and vocabulary size can ensure "adequate" reading comprehension? For this purpose, we need to consider the reading scores in Table 3 in their educational context. As mentioned in the section on Participants, a score of 134 earns students exemption from studying English as a Foreign Language. The exemption rests on the assumption that learners who received a score of 134 and above (about 9% of the learners nationally) can read academic material independently (with or without the aid of a dictionary). Informal reports of these students and of teachers in various disciplines where reading in English is required suggest that these students can indeed do so. Hence, this score represents adequate comprehension if "adequate" is identical to "independent." We can see that for such comprehension, the lexical threshold is between 6K and 8K word knowledge and about 98% of text coverage.

Students who receive a reading score of 116 and above (about 23% of learners) on the psychometric test have to take English for one semester, after which they are expected to read independently. In this semester course, they read authentic academic texts and devote more time to reading strategies than to language. If "adequate" means reading with some guidance and help which is needed at the 116–133 reading level, then the lexical threshold is the knowledge between 4K and 5K words and the coverage of about 93.4 % without proper nouns, or 95.5% (93.4 % + 2.1% of proper nouns).

Learners who receive a lower score than 116 have to take two semesters of English (if the score is 104 to 115), or three semesters (if the score is between 95 and 103). Learners with a lower score than this are not accepted as regular students and have to take pre-academic courses. The college average is 102, which means that the average student in college is required to complete 3 semesters of English upon embarking on higher education. If "adequate" means the average level of a regular student entering a college or a university, who is expected to reach the "independent reading" level after 3 semesters, then the lexical threshold is 3,000 words, 90.56% coverage without proper nouns, or 92.66% with proper nouns. We cannot recommend the third option as adequate reading since according to informal reports, by teachers of other disciplines and students, many of the students in that group continue having difficulties with reading even after they have completed the required courses. Instead, we suggest two lexical thresholds: one optimal and one minimal. The optimal threshold, reflected in the score of 134 and above (91st percentile), which is achieved by 9% of learners nation-wise, predicts functional independence in reading. The students in our study who reached this level knew 6,000–8,000 words. As we had

only 10 such participants, we prefer to take a safe higher estimate of lexical threshold and suggest that it is the knowledge of 8,000 word families which results in the coverage of 98% including proper nouns. The minimal threshold, reflected in a score of 116 (77th percentile), which is achieved by 23% of learners nation-wise, enables learners to read with some guidance, and leads to independent reading after 56 academic hours (one semester). This threshold is the knowledge of 4K–5K words and the coverage of about 95% of text including proper nouns.

These results are similar to Hu and Nation (2000) and Nation (2006). Hu and Nation suggest 98% as the probabilistic coverage at which most learners can read, and 95% as the coverage at which minimally acceptable comprehension can occur. According to Table 14 in Nation (2006), the 95% coverage can be achieved by 5,000 word families with proper nouns: 78% (1K) + 8% (2K) + 3% (3K) + 3% (4K–5K) + 3% proper nouns. The 98% coverage is achieved by an additional 2% provided by 6K–9K words and an additional 1% of proper nouns or alternatively by 1% of words of higher frequency (the above calculation for 95% is based on Nation's lower figures of coverage. The first thousand words can also cover 81%).

Concluding remarks

We do not claim that reasonable reading comprehension cannot occur if learners have not reached the lexical threshold, or that the threshold will automatically yield good reading comprehension. In our data there were learners who did not fit the general pattern of "better vocabulary leading to better reading." The general reading skills of these students may have affected the reading score more than their vocabulary knowledge. However, the R square value (.64) in our regression analysis (or, correlation of .8) suggesting that 64% of variance in the reading score is accounted for by vocabulary corroborates once more the earlier claims that vocabulary may be the major factor in reading comprehension. As for the relationship between vocabulary size and coverage, there are texts, e.g. graded readers where the coverage of 95% and 98% can be reached with a smaller vocabulary than suggested here. Conversely, in some texts with a large proportion of technical and jargon vocabulary, the above coverage may require the knowledge of more low frequency words than suggested in the paper. However, when people read in the area of their expertise, they are usually more familiar with the jargon than with general vocabulary (Cohen et al., 1979). Therefore, when researching reading for general and academic purposes, it is useful to look at academic argumentative prose of general nature.

In this study, we revisited the issue of lexical threshold. We used more rigorous research tools than in earlier studies: typical reading materials, a very large sample of learners with a wide range of language proficiency, the most updated version of the Vocabulary Profile, and a comprehension test that was validated with thousands of test takers. We also attempted to conceptualize and quantify the notion of "adequate" reading comprehension and to relate it simultaneously to the two aspects of threshold, vocabulary size and lexical coverage of a text. Despite these innovations, our data confirmed some earlier results regarding the percentage of text lexis and regarding the vocabulary knowledge required for reading comprehension. We hope that these results can provide useful insights for researchers, syllabus designers, material writers and language teachers.

Notes

1. One reviewer suggested that "adding up the scores in this way may be misleading in regard to coverage as it equates words from later lists with words from earlier lists. For example, having scores of 30+30+30+0+0= 90 is likely to provide much better coverage than scores of 20+20+20+20+20= 100 because the words in the earlier lists recur much more often than the words in the later lists." This is certainly true, but better knowledge of less frequent words has been found with learners of Romance L1s due to the large number of cognates of Roman origin. In the case of learners with Semitic L1s, the scores on the test were implicationally scaled, similarly to findings by Read (1988). Therefore, in our study, there was no danger of misrepresentation of coverage.

2. The testing of 735 students who did not receive exemption from English took place before the new Vocabulary Size test was validated. Hence, we used the revised version of the Levels Test. We tested the exempted students after the test had been validated. Hence we used the 6K-8K sections from this test. We realize that there may have been a small degree of overlap between items on the 5,000 word list and some items on the 6K-8K sections of the vocabulary size test because the two are based on different words lists.

3. Among the 10 learners, a higher vocabulary size did not necessarily mean a higher reading score. Thus one learner with 6K knowledge received a reading score of 141, another one with 6K 134. One learner with 8K received 138, another one 146.

References

Alderson, J. C. (1984). Reading in a foreign language: A reading problem or a language problem? In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 1–27). London: Longman.

Beglar, D. (2009). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, *27*, 101–118.

Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics*, *16*, 15–34.

Carrell, P. L. (1991). Second language reading: Reading ability or language proficiency? *Applied Linguistics*, *12*, 159–179.

Cohen, A. D., Glasman, B., Rosenbaum-Cohen, P. R., Ferrara, J., & Fine, J. (1979). Reading English for specialized purposes: Discourse analysis and the use of student informants. *TESOL Quarterly*, *13,* 551–564.

Clarke, M. A. (1980). The "short-circuit" hypothesis of ESL reading—or when language competence interferes with reading performance. *Modern Language Journal*, *64*, 203–209.

Coxhead, A. (2002). A new Academic Word List. *TESOL Quarterly*, *34*, 213–238.

Cziko, G. A. (1978). Differences in first- and second-language reading: The use of syntactic, semantic and discourse constraints. *The Canadian Modern Language Review*, *34*, 73–189.

Grabe, W. (2004). Research on teaching reading. *Annual Review of Applied Linguistics*, *24*, 44–

69.

Hazenberg, S., & Hulstijn J. H. (1996). Defining a minimal receptive second-language
vocabulary for non-native university students: An empirical investigation. *Applied
Linguistics*, *17*, 145–163.

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension.
Reading in a Foreign Language, *13*, 403−430.

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. (1992). How much lexis is necessary for reading comprehension? In H. Bejoint & P.
Arnaud (Eds.), *Vocabulary and applied linguistics* (pp.126–132). Basingstoke & London:
Macmillan.

Laufer, B., & Nation, I. S. P. (2001). Passive vocabulary size and speed of meaning recognition:
Are they related? In S. Foster-Cohen & A. Nizegorodcew (Eds.), *EUROSLA Yearbook* 1
(pp.7–28). Amsterdam, Benjamins.

Lee, J. L., & Lemonnier-Schallert, D. (1997). The relative contribution of L2 language
proficiency and L1 reading ability to L2 reading performance. A test of the threshold
hypothesis in an EFL context. *TESOL Quarterly*, *31*, 713–739.

Meara, P. M., & Jones G. (1989). *Eurocentres Vocabulary Test 10 KA*. Zurich: Eurocentres.

Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary
training on reading comprehension. *Review of Educational Research*, *53*, 253–279.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, *5* (RELC Supplement), 12–
24.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge
University Press.

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

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

Perkins, K., Brutten, S. R., & Pohlmann, J. T. (1989). First and second language reading
comprehension. *RELC Journal*, *10*(2), 1–9.

Pichette, F., Segalowitz, N., & Connors, K. (2003). Impact of maintaining L1 reading skills on
L2 reading skill development in adults: Evidence from speakers of Serbo-Croatian learning
French. *The Modern Language Journal*, *87*, 391–403.

Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical
inferencing and retention through reading. *Applied Linguistics, 26*, 66-86.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading
comprehension. *The Canadian Modern Language Review*, *56*, 282–308.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic
reading performance: An assessment perspective. *Language Learning*, *52*, 513–536.

Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC
Journal*, *19*(2), 12–25.

Schmitt, N., Schmitt, D. & Clapham, C. (2001). Developing and exploring the behavior of two
new versions of the Vocabulary Levels Test. *Language Testing*, *18*, 55–88.

Segalowitz, N. (2007). Access fluidity, attention control and the acquisition of fluency in a
second language. *TESOL Quarterly*, *41*, 181−186.

Ulijn, J. M., & Strother J. B., (1990). The effect of syntactic simplification on reading EST texts

as L1 and L2. *Journal of Research in Reading*, *13*, 38–54.

Appendix A

A Text From the Psychometric Reading Comprehension Test

Since early times, people have built upwards, towards the stars, whenever and wherever possible. The most well-known examples of this from ancient times are the Tower of Babel and the Pharos of Alexandria, a famous lighthouse. In medieval times, the towers of churches were also built to great heights. In each of these examples, the nature of the materials used in construction imposed certain limitations on the height of the building. The brickwork, or masonry, had to be of a certain minimum thickness to support whatever was built on top of it. In general, the taller a building, the heavier it was. Therefore, however ingenious the architect's design, the building's height was limited by the strength and width of the walls of its lower stories. This remained a problem for many centuries.

It was not until the second half of the 19th century—when iron, and then steel, replaced brick as building materials—that the situation changed radically. At the same time, another obvious barrier to the construction of tall buildings was removed with the development of the safe lift, or elevator. No longer was it possible to dismiss the idea of building skyscrapers on the grounds that people would object to walking up a dozen or more flights of stairs to get to their home or office. Elisha Otis, the inventor of the elevator, is often credited with being the man who, more than anyone, made the skyscraper a viable proposition.

However, it was the work of Gustave Eiffel, an engineer and architect, which was to give the greatest boost to the actual construction of skyscrapers. He was one of Europe's most famous engineers, and had built the Eiffel Tower in Paris using iron and steel. In the early 1880s, Eiffel was given the task of building internal supports for the 45-meter-high Statue of Liberty that was to be erected at the entrance to New York Harbor. The chief material he used was steel, marking the first time that its use had been specified in the plans for any structure in New York other than a bridge. Eiffel's work on the Statue of Liberty and the Eiffel Tower convinced the Americans that steel, not brick, was the material of the future, and so the era of the skyscraper began.

Questions
23. The main purpose of the text is to describe –
(1) developments that led to the construction of tall buildings
(2) the advantages and disadvantages of the modern skyscraper
(3) the work of the people responsible for inventing the skyscraper
(4) some of the tallest buildings in the world, in the past and today

24. The Tower of Babel and the Pharos of Alexandria are presented in the text as examples of buildings –
(1) constructed from special materials
(2) that people admired long ago
(3) that were as high as medieval church towers

(4) from ancient times that were extremely tall

25. The purpose of the second paragraph is to –
(1) discuss the methods used to build tall buildings in early times
(2) describe the materials architects once used to support a building
(3) explain why, in the past, the height of a building was limited
(4) show that masonry was once used to solve architectural problems

26. According to the text, Elisha Otis was responsible for –
(1) making the building of skyscrapers a realistic possibility
(2) suggesting that steel would eventually replace brickwork
(3) inventing the idea of the skyscraper
(4) introducing new materials into the construction industry

27. It can be inferred from the last paragraph that before Eiffel's work on the Statue of Liberty, steel had been –
(1) a more popular building material in the United States than in Europe
(2) used in New York in the building of bridges but not other structures
(3) considered a material that could only be used in building tall structures
(4) used in the construction of most of the skyscrapers in New York

About the Authors

Batia Laufer is Professor of Applied Linguistics at the University of Haifa, Israel. She researches L2 vocabulary acquisition, lexicography, cross-linguistic influence, reading, and testing. Her most known work is on vocabulary threshold for reading comprehension, vocabulary development, task effect on learning, lexical quantitative assessment, and factors of word difficulty. E-mail: batialau@research.haifa.ac.il

Geke Ravenhorst-Kalovski is currently working on her PhD at the University of Haifa and teaching English for Academic Purposes at the Yezreel Valley College. She is interested in the relationships between reading comprehension, vocabulary size and form focused instruction of vocabulary. E-mail: gekek@yvc.ac.il