



The Percentage of Words Known in a Text and Reading Comprehension

NORBERT SCHMITT

*University of Nottingham
Nottingham, United Kingdom
Email: norbert.schmitt@
nottingham.ac.uk*

XIANGYING JIANG

*West Virginia University
Morgantown, WV
Email: xiangying.jiang@
mail.wvu.edu*

WILLIAM GRABE

*Northern Arizona University
Flagstaff, AZ
Email: william.grabe@nau.edu*

This study focused on the relationship between percentage of vocabulary known in a text and level of comprehension of the same text. Earlier studies have estimated the percentage of vocabulary necessary for second language learners to understand written texts as being between 95% (Laufer, 1989) and 98% (Hu & Nation, 2000). In this study, 661 participants from 8 countries completed a vocabulary measure based on words drawn from 2 texts, read the texts, and then completed a reading comprehension test for each text. The results revealed a relatively linear relationship between the percentage of vocabulary known and the degree of reading comprehension. There was no indication of a vocabulary “threshold,” where comprehension increased dramatically at a particular percentage of vocabulary knowledge. Results suggest that the 98% estimate is a more reasonable coverage target for readers of academic texts.

IN A RECENT ARTICLE, NATION (2006) CONCLUDED that much more vocabulary is required to read authentic texts than has been previously thought. Whereas earlier research suggested that around 3,000 word families provided the lexical resources to read authentic materials independently (Laufer, 1992), Nation argues that in fact 8,000–9,000 word families are necessary. The key factor in these widely varying estimates is the percentage of vocabulary in a text that one needs to comprehend it. An earlier study (Laufer, 1989) came to the conclusion that around 95% coverage was sufficient for this purpose. However, Hu and Nation (2000) reported that their participants needed to know 98%–99% of the words in texts before adequate comprehension was possible. Nation used the updated percentage figure of 98% in his analysis, which led to the 8,000–9,000 vocabulary figure.

As reading is a crucial aid in learning a second language (L2), it is necessary to ensure that learners have sufficient vocabulary to read well (Grabe, 2009; Hudson, 2007; Koda, 2005). How-

ever, there is a very large difference between learning 3,000 and 9,000 word families, and this has massive implications for teaching methodology. When the instructional implications of vocabulary size hinge so directly on the percentage of coverage figure, it is important to better establish the relationship between vocabulary coverage and reading comprehension. Common sense dictates that more vocabulary is better, and there is probably no single coverage figure, for example 98%, over which good comprehension occurs and short of which one understands little. Indeed, both Laufer (1989, 1992) and Hu and Nation (2000) found increasing comprehension with increasing vocabulary coverage. This suggests that there is a coverage/comprehension “curve,” indicating that more coverage is generally better, but it may or may not be linear. This study builds on Laufer’s and Hu and Nation’s earlier studies and uses an enhanced research methodology to describe this curve between a relatively low vocabulary coverage of 90% to knowledge of 100% of the words in a text.

BACKGROUND

Reading is widely recognized as one of the most important skills for academic success, both in

first language (L1) and L2 environments (Johns, 1981; National Commission on Excellence in Education, 1983; Rosenfeld, Leung, & Oltman, 2001; Sherwood, 1977; Snow, Burns, & Griffin, 1998). In many cases, L2 reading represents the primary way that students can learn on their own beyond the classroom. Research has identified multiple component skills and knowledge resources as important contributors to reading abilities (Bowey, 2005; Grabe, 2004; Koda, 2005; Nassaji, 2003; Perfetti, Landi, & Oakhill, 2005). However, one of the primary factors consistently shown to affect reading is knowledge of the words in the text. In general, research is increasingly demonstrating what practitioners have always known: that it takes a lot of vocabulary to use a language well (for more on this, see Nation, 2006; Schmitt, 2008). This is particularly true for reading. Vocabulary knowledge and reading performance typically correlate strongly: .50–.75 (Laufer, 1992); .78–.82 (Qian, 1999); .73–.77 (Qian, 2002). Early research estimated that it took 3,000 word families (Laufer, 1992) or 5,000 individual words (Hirsh & Nation, 1992) to read texts. Similarly, Laufer (1989) came up with an estimate of 5,000 words. More recent estimates are considerably higher, in the range of 8,000–9,000 word families (Nation, 2006).

These higher figures are daunting, but even so, they probably underestimate the lexis required. Each word family includes several individual word forms, including the root form (e.g., *inform*), its inflections (*informed*, *informing*, *informs*), and regular derivations (*information*, *informative*). Nation's (2006) British National Corpus lists show that the most frequent 1,000 word families average about six members (types per family), decreasing to about three members per family at the 9,000 frequency level. According to his calculations, a vocabulary of 8,000 word families (enabling wide reading) entails knowing 34,660 individual word forms, although some of these family members are low-frequency items. The upshot is that students must learn a large number of individual word forms to be able to read a variety of texts in English, especially when one considers that the figures above do not take into account the multitude of phrasal lexical items that have been shown to be extremely widespread in language use (e.g., Grabe, 2009; Schmitt, 2004; Wray, 2002).

Unfortunately, most students do not learn this much vocabulary. Laufer (2000) reviewed a number of vocabulary studies from eight different countries and found that the vocabulary size of high school/university English-as-a-second-

language English-as-a-foreign-language (EFL) learners ranged from 1,000–4,000.¹ Whereas a 3,000–5,000 word family reading target may seem attainable for learners, with hard work, the 8,000–9,000 target might appear so unachievable that teachers and learners may well conclude it is not worth attempting. Thus, the lexical size target is a key pedagogical issue, and one might ask why the various estimates are so different.

The answer to that rests in the relationship between vocabulary knowledge and reading comprehension. In a text, readers inevitably come across words they do not know, which affects their comprehension. This is especially true of L2 learners with smaller vocabularies. Thus, the essential question is how much unknown vocabulary learners can tolerate and still understand a text. Or we can look at the issue from the converse perspective: What percentage of lexical items in a text do learners need to know in order to successfully derive meaning from it?

Laufer (1989) explored how much vocabulary is necessary to achieve a score of 55% on a reading comprehension test. This percentage was the lowest passing mark in the Haifa University system, even though earlier research suggested that 65%–70% was the minimum to comprehend the English on the Cambridge First Certificate in English examination (Laufer & Sim, 1985). She asked learners to underline words they did not know in a text, and adjusted this figure on the basis of results of a translation test. From this she calculated the percentage of vocabulary in the text each learner knew. She found that 95% was the point which best distinguished between learners who achieved 55% on the reading comprehension test versus those who did not. Using the 95% figure, Laufer referred to Ostyn and Godin's research (1985) and concluded that approximately 5,000 words would supply this vocabulary coverage. Although this was a good first attempt to specify the vocabulary requirements for reading, it has a number of limitations (see Nation, 2001, pp. 144–148 for a complete critique). Ostyn and Godin's frequency counts are of Dutch, and it is not clear that they can be applied directly to English. Their methodology also mixed academic texts and newspaper clippings, but different genres can have different frequency profiles (see Nation, 2001, Table 1.7). Perhaps most importantly, the comprehension criterion of 55% seems to be very modest, and most language users would probably hope for better understanding than this. Nevertheless, the 95% coverage figure and the related 3,000–5,000 vocabulary size figure were widely cited.

A decade later, Hu and Nation (2000) compared reading comprehension of fiction texts at 80%, 90%, 95%, and 100% vocabulary coverages. Sixty-six students studying on a pre-university course were divided into four groups of 16–17 participants. Each group read a 673-word story, at one of the aforementioned vocabulary coverage levels. They then completed multiple-choice (MC) and cued written recall (WR) comprehension tests. No learner achieved adequate comprehension at 80% vocabulary coverage, only a few did at 90%, and most did not even achieve adequate comprehension at 95%. This suggests that the minimum amount of vocabulary coverage to make reading comprehensible is definitely above 80% (1 unknown word in 5), and the low number of successful learners at the 90% coverage level (3/16, 19%) indicates that anything below 90% is an extreme handicap. Ninety-five percent coverage allowed 35%–41% of the participants to read with adequate comprehension, but this was still a minority, and so Hu and Nation concluded that it takes 98%–99% coverage to allow unassisted reading for pleasure.² However, these results were based on only four coverage points, and involved a relatively small number of participants per coverage level.

Although informative, both of these studies have limitations, and so the vocabulary coverage figures they suggest (95%/98%–99%) must be seen as tentative. This means that the related vocabulary size requirements based upon these coverage percentages are also tentative. Whereas size estimates differ according to the assumed vocabulary coverage requirement, it is impossible to specify vocabulary learning size targets until the vocabulary coverage–reading comprehension relationship is better understood.

The two studies above have been widely cited as describing vocabulary coverage “thresholds” beyond which adequate comprehension can take place. This is unfortunate, as both studies in fact found that greater vocabulary coverage generally led to better comprehension, and the reported figures were merely the points at which adequate comprehension was most likely to occur, rather than being thresholds. Laufer (1989) found that the group of learners who scored 95% and above on the coverage measure had a significantly higher number of “readers” (with scores of 55% or higher on the reading comprehension test) than “non-readers” (< 55%). Thus, the 95% figure in her study was defined probabilistically in terms of group success, rather than being a point of comparison between coverage and comprehension. Similarly, Hu and Nation (2000) defined ad-

equated comprehension (establishing 12 correct answers out of 14 on an MC test and 70 out of 124 on a WR test) and then determined whether learners at the four coverage points reached these criteria. They concluded that 98% coverage was the level where this was likely to happen, although their study also clearly showed increasing comprehension with increasing vocabulary: 80% coverage = 6.06 MC and 24.60 WR; 90% = 9.50, 51.31; 95% = 10.18, 61.00; 100% = 12.24, 77.17. In a study that compared learners’ vocabulary size with their reading comprehension (Laufer, 1992), similar results were obtained, with larger lexical sizes leading to better comprehension.

The best interpretation of these three studies is probably that knowledge of more vocabulary leads to greater comprehension, but that the percentage of vocabulary coverage required depends on how much comprehension of the text is necessary. Even with very low percentages of known vocabulary (perhaps even 50% or less), learners are still likely to pick up some information from a text, such as the topic. Conversely, if complete comprehension of all the details is necessary, then clearly a learner will need to know most, if not all, of the words. However, this in itself may not guarantee full comprehension.

This interpretation suggests that the relationship between vocabulary coverage and reading comprehension exists as some sort of curve. Unfortunately, previous research approaches have only provided limited insight into the shape of that curve. Laufer’s group comparisons tell us little about the nature of the curve, and although her 1992 regression analyses hint at linearity, we do not know how much of the comprehension variance the regression model accounts for. Hu and Nation (2000) sampled at only four coverage points, but were able to build a regression model from this, accounting for 48.62% (MC) and 62.18% (WR) of the reading comprehension variance. Again, while this suggests some sort of linear relationship, it is far from conclusive.

It may be that vocabulary coverage and reading comprehension have a straightforward linear relationship, as the previously mentioned regression analyses hint at (Figure 1). However, it is also possible that there is a vocabulary coverage percentage where comprehension noticeably improves, forming a type of threshold not discovered by previous research methodologies. This possibility is illustrated in Figure 2. It may also be that considerable comprehension takes place at low coverage levels and reaches asymptote at the higher coverage figures, forming a type of S-curve (Figure 3).

FIGURE 1
Linear Relationship

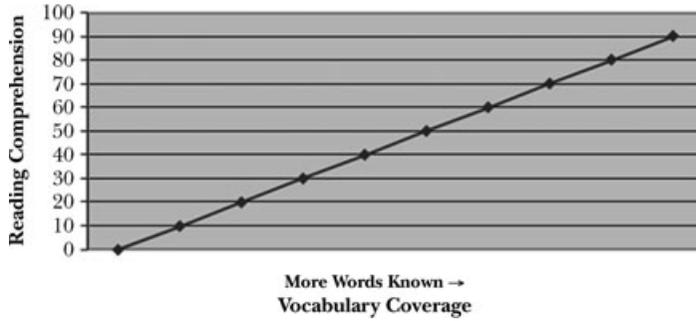


FIGURE 2
Vocabulary Threshold

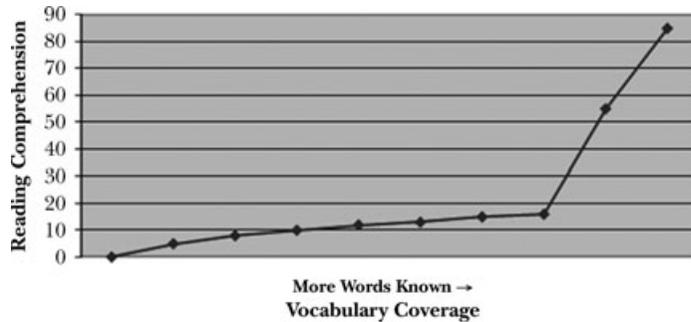
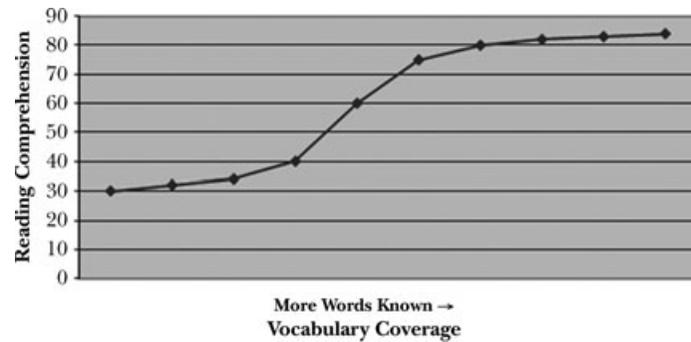


FIGURE 3
S-Curve



Of course, vocabulary is not the only factor affecting comprehension. In fact, a large number of variables have been shown to have an effect, including those involving language proficiency (e.g., grammatical knowledge and awareness of discourse structure), the text itself, (e.g., text length, text difficulty, and topic), and those concerning the reader (interest in topic, motivation, amount of exposure to print, purposes for

reading, L1 reading ability, and inferencing ability) (Droop & Verhoeven, 2003; Grabe, 2009; van Gelderen et al., 2004; van Gelderen, Schoonen, Stoel, de Glopper, & Hulstijn 2007).

Vocabulary knowledge has been shown to underlie a number of these abilities, especially language proficiency and inferencing ability. However, many component abilities make independent contributions to reading

comprehension. Background knowledge is one such factor shown to have a large effect on reading abilities. Readers with much greater knowledge of a topic, greater expertise in an academic domain, or relevant social and cultural knowledge understand a text better than readers who do not have these resources (Hudson, 2007; Long, Johns, & Morris, 2006). At a general level, it is obvious that background knowledge contributes to reading comprehension, in that readers' inferencing abilities require prior knowledge to be utilized effectively, and it would be difficult to generate a mental model of the text meaning without inferencing (Zwann & Rapp, 2006). Moreover, inferencing skills, in the context of reading academic texts, are an important predictor of reading abilities (Oakhill & Cain, 2007; Perfetti et al., 2005).

At the same time, the role of background knowledge is not always easy to specify. In many contexts, background knowledge does not distinguish better readers from weaker ones (e.g., Bernhardt, 1991), nor does it always distinguish among readers from different academic domains (Clapham, 1996). In fact, the many complicating variables that interact with background knowledge (such as those mentioned earlier) make it difficult to predict the influence of background knowledge on reading. It nonetheless remains important to investigate contexts in which the role of background knowledge can be examined to determine its possible impact on reading abilities. In the present study, two texts that differed in assumed degree of familiarity served as the stimulus materials for the tests we created. For one text, the participants presumably had a great deal of previous knowledge; for the second, they seemingly had relatively little background knowledge. This combination allowed us to explore the effects of background knowledge on comprehension in relation to vocabulary coverage.

This study addressed the following research questions:

1. What is the relationship between percentage of vocabulary coverage and percentage of reading comprehension?
2. How much reading comprehension is possible at low percentages of vocabulary coverage?
3. How does the degree of background knowledge about the text topic affect the vocabulary coverage–reading comprehension relationship?

To answer these questions, the study will use an innovative research design incorporating a very large and varied participant population, two relatively long authentic reading passages, extended, sophisticated measures of vocabulary knowledge

and reading comprehension, and a direct comparison of these measures. It should thus provide a much firmer basis for any conclusions concerning the coverage–comprehension relationship than any previous study.

METHODOLOGY

Selection of the Reading Passages

Two texts were selected for this study because they were relatively equal in difficulty, appropriate for students with more advanced academic and reading skills, of sufficient length to provide less frequent vocabulary targets and enough viable comprehension questions for our research purposes. The text titled “What’s wrong with our weather” concerned climate change and global warming, a topic for which most people would have considerable previous knowledge (hereafter “Climate”). The other text, titled “Circuit training,” was about a scientific study of exercise and mental acuity carried out with laboratory mice, about which we felt most people would have little, if any, prior knowledge (hereafter “Mice”). The texts could be considered academic in nature. The Climate passage appeared in an EFL reading textbook in China (Zhai, Zheng, & Zhang, 1999), and the Mice text was from *The Economist* (September 24, 2005). The texts varied somewhat in length (Climate at 757 words, Mice at 582 words), but were of similar difficulty on the basis of the Flesch–Kincaid Grade Level (Climate at Grade Level 9.8, Mice at Grade Level 9.7). It is useful to note that these were somewhat longer text lengths than has been the norm in much past reading research, in order to mimic the more extended reading of the real world. The passages were not modified, and so remained fully authentic. See the Appendix for the complete testing instrument, including the reading texts.

Development of the Vocabulary Test

To plot the coverage–comprehension curve, it was necessary to obtain valid estimates of both the percentage of vocabulary in a text that learners know, as well as how much they could comprehend. This study contains extended tests of both in order to obtain the most accurate measures possible.

Laufer (1989) measured her learners' knowledge of words in a text by asking them to underline the ones they felt they did not know and then adjusting these according to the results of a vocabulary translation test. However, the act of

underlining unknown words in a text is different from simply reading a text, and so it is unclear how this affected the reading process. For this reason, we felt that a discrete vocabulary test would be better, as it would have less chance of altering the natural reading process. Hu and Nation (2000) inserted plausible nonwords in their texts to achieve desired percentages of unknown words. The use of nonwords is an accepted practice in lexical studies (Schmitt, 2010) and has the advantage of eliminating the need for a pretest (because learners cannot have previous knowledge of nonwords). However, we wished to retain completely unmodified, authentic readings, and so did not adopt this approach.

Instead, we opted for an extended vocabulary checklist test containing a very high percentage of the words in the two readings. The readings were submitted to a Lextutor frequency analysis (www.lextutor.ca), and vocabulary lists were made for both texts, also determining which words occurred in both texts and which were unique to each text. We assumed that high intermediate through advanced EFL readers would know almost all of the first 1,000 most frequent words in English. Therefore, we only sampled lightly at the first 500 and second 500 frequency bands (10 words from each band), to confirm our assumption was valid. It proved to be correct, as the vast majority of participants knew all of the first 500 target words (96%) and second 500 target words (86%). Almost all remaining learners missed only 1 or 2 words in these bands. At the 1,000–2,000 band and the >2,000 (all remaining words) band, we sampled much more extensively, including more than 50% of these words from the texts in our checklist test. This is a very high sampling rate, and should provide a very good estimate of the percentage of vocabulary in the two readings that the participants knew. We selected all of the 1,000–2,000 and >2,000 words occurring in both texts to include on the test. The words that occurred in only one text were ranked in frequency order, and then every other word was selected for the test. The process resulted in the selection of 120 target words. In effect, participants were measured directly on their knowledge of a very high percentage of the words actually appearing in the two texts that they read, removing the large inferences typically required in other vocabulary measurement studies with lower sampling rates.

While testing a large percentage of words from the two texts should help ensure a valid estimate of how many words learners knew in those texts, it presents a challenge in terms of practicality. A

traditional MC test format with a contextualized sentence would take far too much time, considering that the learners needed to read two texts and finish two extended comprehension tests, as well. The same is true of “depth of knowledge” tests, like the Word Associates Test (Qian, 2002; Read, 2000), which provide a better indication of how well words are known, but at the expense of greater time necessary to complete the items. Furthermore, we administered the test to a wide variety of sites and nationalities, which made the use of L1 translation infeasible. We also needed an item format that mimicked the use of vocabulary when reading, that is, recognizing a written word form, and knowing its meaning.

The solution to this requirement for a quick, form–meaning test was a checklist (yes/no) item format. In this type of test, learners are given a list of words and merely need to indicate (“check”) the ones they know. Thus, the test measures receptive knowledge of the form–meaning link (Schmitt, 2008). This format has been used in numerous studies (e.g., Anderson & Freebody, 1983; Goulden, Nation, & Read, 1990; Meara & Jones, 1988; Milton & Meara, 1995; Nagy, Herman, & Anderson, 1985) and has proven to be a viable format. The main potential problem is overestimation of knowledge, that is, learners checking words that they, in fact, do not know. To guard against this, plausible nonwords³ can be inserted into the test to see if examinees are checking them as known. If they are, then their vocabulary score can be adjusted downward through the use of adjustment formulas (e.g., Meara, 1992; Meara & Buxton, 1987). However, it is still unclear how well the adjustment formulas work (Huibregtse, Admiraal, & Meara, 2002; Mochida & Harrington, 2006), so we decided to simply delete participants from the data set who chose too many nonwords, which made their vocabulary scores unreliable. Thirty nonwords selected from Meara’s (1992) checklist tests were randomly inserted among the 120 target words, resulting in a vocabulary test of 150 items. The words/nonwords were put into 15 groups of 10 items, roughly in frequency order, that is, the more frequent words occurred in the earlier groups. The first group had 3 nonwords to highlight the need to be careful, and thereafter each group had 1–3 nonwords placed in random positions within each group.

Development of the Reading Comprehension Tests

The research questions we explore posed a number of challenges for the reading texts chosen

for this study. We needed enough question items to spread student scores along the continuum of the participants' vocabulary knowledge for these texts. We were also constrained in not being able to ask about specific vocabulary items, which is normally a good way to increase the number of items on a reading test. Because the participants were tested on their vocabulary knowledge on the vocabulary measure, and because the two measures needed to be as independent of each other as possible, the reading tests could not use vocabulary items to increase the number of items. We also needed to limit items that involved recognizing details but would emphasize specific vocabulary items.

To develop enough reliable (and valid) comprehension items, we created a two-part reading test for each text. The first part included 14 MC items with an emphasis on items that required some inferencing skills in using information from the text. We chose the MC format because it is a standard in the field of reading research and has been used in a large number of studies. Two major books concerning the validation of English language tests (Clapham, 1996; Weir & Milanovic, 2003) provide extensive validity arguments for reading tests that incorporate MC items. In addition, the developers of the new TOEFL iBT have published a book validating the developmental process and the performance effectiveness of a major international test drawing largely on MC items for the reading part of the test (Chapelle, Enright, & Jamieson, 2007).

Numerous testing experts (e.g., Alderson, 2000; Hughes, 2003) recommend that reading assessment should incorporate multiple tasks, and so we included an innovative graphic organizer format that incorporated the major discourse structures of the text as the second part of our reading test battery. The graphic organizer (GO) completion task (sometimes termed an "information transfer task") is more complex and requires more cognitive processing than basic comprehension measures. In addition to locating information and achieving basic comprehension, the GO completion task requires readers to recognize the organizational pattern of the text and see clear, logical relationships among already-filled-in information and the information sought through the blanks. It goes beyond what an MC task can do and serves as a good complement to it. Our GOs included 16 blank spaces that participants had to fill in to complete the task. After reading the passage, students first responded to the MC questions and then filled in

the blanks in the GOs. The GOs were partially completed.

The reading test underwent multiple rounds of piloting. After we created GOs for each text, we asked a number of students to work with them to see how well they could fill in the blanks. Revisions were made on the basis of the feedback. We next pilot-tested the complete battery (both vocabulary and reading) with 120 university students in China.

The results from the piloting process led to revisions to the reading test. The original version of the test had 15 MC items and 20 GO items for each reading passage. After analysis, a number of initial items were dropped from the test based on their item difficulty and item discrimination indices. Several MC items were revised. The revised version of the MC test was again piloted with 52 Intensive English Program (IEP) students. The piloting process resulted in the 30-item reading test for each text (14 MC items, 16 GO items) that were then used in the study.

The items comprising the reading test were scored as either correct or incorrect. One point was given for each correct answer. Different from the objective nature of the MC test, the GO items allowed for variations in acceptable answers. Detailed scoring rubrics were developed for the GOs of each passage prior to scoring. Two raters scored approximately 20% of the GOs for the Climate passage, with an interrater reliability of .99 (Cronbach's alpha). Therefore, one rater scored the rest of the instruments. The reliability estimates (based on the K-R 21⁴ formula) for the reading tests are as follows: .82 for the entire reading test, .79 for the Climate reading test, .65 for the Mice reading test, .59 for the MC items, and .81 for the GO items. Due to the potential underestimation of the K-R 21 formula, the actual reliability coefficients could be higher.

The MC and GO results were analyzed separately and compared. Although the GO scores were generally slightly higher, the two test formats produced very similar coverage-comprehension curves, and so the results were combined into a single comprehension measure.

Participants

Individuals willing to recruit and supervise test administrations were recruited from 12 locations in eight countries (in order from greatest to least participation: Turkey, China, Egypt, Spain, Israel, Great Britain, Japan, and Sweden) over the second half of 2007. We received a total of 980 test

samples. After two rounds of elimination (see more information in the next section), we arrived at 661 valid samples to be included in the analysis. Test participants ranged from intermediate to very advanced, including pre-university IEP, freshman, sophomore, junior, senior, and graduate students with 12 different L1s. The age of the participants ranged from 16 to 33 years old, with an average of 19.82 ($SD = 2.53$). They had studied English for an average of 10 years ($SD = 2.74$). Among the 661 participants, 241 were males and 420 females, 212 were English majors, and 449 were from other disciplines. More information concerning the nonnative (NNS) participants appears in Table 1. We also administered the instrument to 40 native-speaking (NS) students at two universities in order to develop baseline data. The NS participants included 22 freshmen, 13 sophomores, 2 juniors, and 3 seniors (age $M = 19.68$; 33 F, 7 M) enrolled in language and linguistics courses.

Procedure

After a series of pilots, the test battery was administered to the participants in their various countries. Participants first completed a biodata page. They then took the vocabulary test and were instructed not to return to it once finished. The next step was to read the Climate text and answer the related comprehension items. Finally, they did the same for the Mice text and items. The participants were free to return to the texts as much as they wished when answering the comprehension items.

To ensure that our tests were providing accurate estimates of both vocabulary coverage and reading comprehension, the results were first screened according to the following criteria. For the vocabulary test, we needed to know that the participants were not overestimating their vocabulary knowledge, and so all instruments in which the participants checked 4 or more nonwords were deleted from the data set. This meant that we accepted only vocabulary tests in which 3 or fewer nonwords were checked, which translates to a maximum of 10% error (3 nonwords maximum/30 total nonwords).

To confirm the participants were not overestimating their vocabulary knowledge at this criterion, we also analyzed the data set when only a maximum of 1 nonword was accepted. This eliminated another 178 participants, but we found that this did not change the results based on the 3-nonword criterion. We compared the comprehension means of the 3-nonword and 1-nonword data sets, with t -tests showing no significant difference (all tests $p > .45$) at any of the 11 coverage points on either the Climate or the Mice texts (Table 3 shows the 3-nonword results). We therefore decided to retain the 3-nonword criterion.

Participants were asked to complete the GO sections of the reading comprehension tests, but inevitably the weaker students had difficulty doing this. As some students did not finish all the GO items, we needed to distinguish between those who made an honest effort and those who did not. We operationalized this as instruments in which at least five GO items were attempted for each reading. If a student attempted to answer at least

TABLE 1
Participant Information

First Language	Number of Participants	Country of Institutions	Number of Participants	Academic Levels	Number of Participants
Turkish	292	Turkey	294	IEP	135
Chinese	180	Egypt	104	Freshman	270
Arabic	101	Israel	31	Sophomore	142
Spanish	33	Britain	49	Junior	41
Hebrew	26	Sweden	5	Senior	50
Japanese	7	China	142	Graduate	23
Russian	5	Spain	33		
French	5	Japan	3		
Swedish	5				
German	4				
Vietnamese	2				
Korean	1				
Total	661		661		661

Note. IEP = Intensive English Program.

TABLE 2
Vocabulary Coverage vs. Reading Comprehension (Combined Readings)

	90% ^a	91%	92%	93%	94%	95%	96%	97%	98%	99%	100%
N (NNS)	21	18	39	33	83	146	176	196	200	186	187
Mean ^b	15.15	15.28	16.26	15.45	17.73	18.16	18.71	19.21	20.49	21.31	22.58
SD	5.25	3.88	4.52	5.21	5.30	4.39	4.76	5.07	4.71	4.74	4.03
Median ^b	15	15	17	16	18	19	19	19	21	22	23
% ^c	50.5	50.9	54.2	51.5	59.1	60.5	62.4	64.0	68.3	71.0	75.3
Minimum	6	9	4	3	6	7	7	5	6	5	7
Maximum	26	24	25	24	28	29	30	29	29	30	30
N (NS)						1	1		12	27	39
Mean ^b						18.00	18.00		21.08	21.63	23.46
SD									3.58	3.25	2.85
Median ^b						18	18		23	22	23
% ^c						60.0	60.0		70.3	72.1	78.2
Minimum						18	18		16	15	16
Maximum						18	18		27	28	30

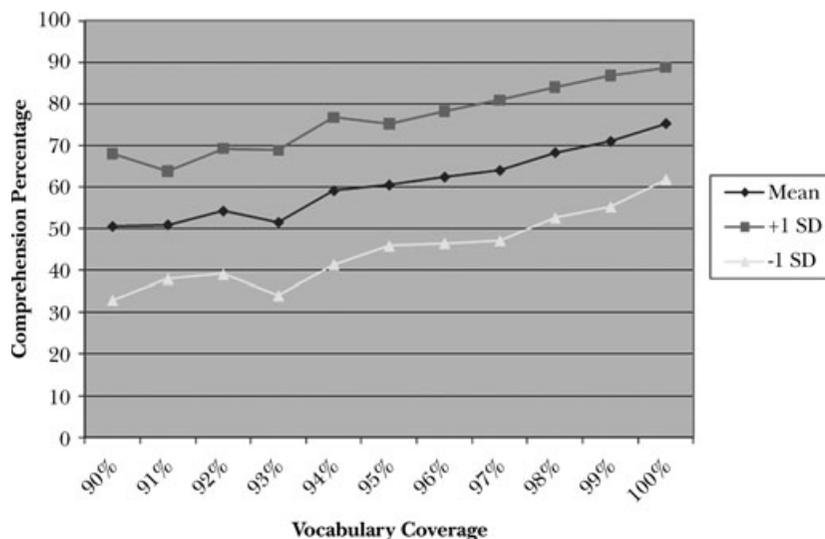
Note.

^aVocabulary coverage.

^bCombined graphic organizer and multiple-choice comprehension tests (Max = 30).

^cPercentage of possible comprehension (Mean ÷ 30).

FIGURE 4
Vocabulary Coverage vs. Reading Comprehension (Combined Readings)



five GO items in a section, we assumed that the student made an effort to provide answers for this task, and so any items left blank indicated a lack of knowledge rather than a lack of effort. Any GO section where five items were not attempted was deleted from the analysis.

Ultimately, 661 out of 980 instruments (67%) passed both of these criteria. While many data

were lost, we feel this is an acceptable compromise in order to be fully confident in the validity of the vocabulary and comprehension measures. It did eliminate the possibility of exploring the coverage-comprehension curve below 90% vocabulary coverage, but previous research (Hu & Nation, 2000) has shown that limited comprehension occurs below this level in

any case. Hu and Nation's research suggests that coverage percentages above 90% are necessary to make reading viable, and so we are satisfied to work within the 90%–100% coverage band, where we had sufficient participants at all coverage points.

The data from the 661 instruments were entered into a spreadsheet, which automatically calculated, for each participant, the total percentage of vocabulary coverage for each text. The calculations assumed that all function words were known, giving a starting vocabulary coverage of 39.89% (302 function words out of 757 total words) for the Climate text, and 46.05% for the Mice text (268/582). If the learners knew all of the first 1,000 content words (the case for the vast majority of participants), then this brought their vocabulary coverage up to about 80%. Therefore, the critical feature that differentiated our participants (and their vocabulary coverage) was knowledge of vocabulary beyond the first-1,000 frequency level. During data entry, the reading comprehension scores from GO and MC tests were also entered into the spreadsheet.

RESULTS AND DISCUSSION

Relationship Between Vocabulary Coverage and Reading Comprehension

The main purpose of this research is to describe the vocabulary coverage–reading comprehension relationship. We split the participants' results into 1% coverage bands (i.e., 90% coverage, 91% coverage, 92% coverage, etc., rounding to the nearest full percentage point), and then determined their scores on the combined GO (16 points) and MC (14 points) parts of the comprehension tests for each text (30 points maximum total). For example, we grouped all participants who had 95% coverage on either the Climate or the Mice text and then calculated the mean total comprehension scores for that text. The two texts were analyzed separately, as learners usually had different coverage percentages on the different texts. The results are illustrated in Table 2 and Figure 4.

A number of important conclusions can be derived from these results, which offer the most comprehensive description of the vocabulary coverage–reading comprehension relationship to date, at least within the 90%–100% coverage range. The answer to the first research question is that, although producing a Spearman correlation of only .407 ($p < .001$),⁵ the relationship between percentage of vocabulary coverage and percentage of reading comprehension

appears to be essentially linear, at least within the vocabulary coverage range of 90% to 100%. To state this another way, the more vocabulary coverage, the greater the comprehension. The total increase in comprehension went from about 50% comprehension at 90% vocabulary coverage to about 75% comprehension at 100% vocabulary coverage. There was no obvious point at which comprehension dramatically accelerated; rather, comprehension gradually increased with increasing vocabulary coverage. Except for a slight dip at 93%, the results point to a remarkably consistent linear relationship between growing vocabulary knowledge and growing reading comprehension. This argues against any threshold level, after which learners have a much better chance of understanding a text. Our study used a direct comparison of coverage and comprehension, and it seems that each increase in vocabulary coverage between the 90% and 100% levels offers relatively uniform gains in comprehension. This finding is in line with earlier studies by Laufer (1989) and Hu and Nation (2000), which used more indirect approaches to the coverage–comprehension relationship.

The results suggest that the degree of vocabulary coverage required depends on the degree of comprehension required. For our advanced participants and our measurement instruments, if 60% comprehension is considered adequate, then 95% coverage would suffice. If 70% comprehension is necessary, then 98%–99% coverage was required. If the goal is 75% comprehension, then the data suggest that our learners needed to know all of the words in the text. (Of course, these percentages depend on the nature of the reading comprehension tasks being administered. These percentages will change with easier or more difficult texts and tasks, but the results nonetheless highlight a strong linear relationship between the two.)

The answer to the second research question is that our learners were able to comprehend a considerable amount of information in a text, even with relatively low levels of vocabulary coverage. Hu and Nation (2000) found that even at 80% coverage, their learners were able to score about 20% on the written recall test, and 43% on the MC comprehension test. At 90% coverage, this improved to 41% WR and 68% MC. In our study, the learners could answer 50% of the comprehension items correctly at 90% vocabulary coverage. This suggests that although comprehension may not be easy when there is more than 1 unknown word in 10, learners can still achieve substantial comprehension.

Conversely, even learners who knew 100% of the words in the texts could not understand the texts completely, obtaining a mean score of 22.58 out of 30 possible (75.3%). Overall, learners who knew most of the words in a text (i.e., 98%–100%) scored between 68.3% and 75.3% on average on the combined comprehension tests. In this study, achieving maximum scores on the vocabulary measure was not sufficient to achieve maximum scores on the comprehension measure, and there are at least two reasons for this. First, it is well known that multiple component reading skills influence comprehension aside from vocabulary knowledge, including word recognition efficiency; text reading fluency; morphological, syntactic, and discourse knowledge; inferencing and comprehension monitoring skills; reading strategy uses with difficult texts; motivation to read; and working memory (Grabe, 2009; Koda, 2005, 2007). Second, the checklist test measured only the form–meaning link of the words in the texts, and so even if words were judged as known, this does not necessarily mean that participants possessed the kind of deeper lexical knowledge that would presumably enhance the chances of comprehension.

It is important to note that our results were for relatively advanced learners; indeed, even the native speakers produced very similar results. However, it is not clear whether lower proficiency learners would exhibit the same coverage–comprehension curve. If less advanced learners (presumably with smaller vocabulary sizes) tried to read similar authentic materials, the comprehension figures would surely be lower, and not knowing the most frequent 2,000 words of English could well push the coverage figure down so low that little meaningful comprehension would be possible (although even then a certain amount of comprehension would probably accrue). However, this line of reasoning is probably not very productive, as good pedagogy dictates that when vocabulary knowledge is limited, reading material must be selected that matches these limited lexical resources. It does not make much sense having students read texts for which they do not know 10% or more of the words. Even learners with very small vocabulary sizes can successfully read in an L2 if the reading level is appropriate (e.g., low-level graded readers). We might speculate that our study design completed with beginning learners and appropriately graded reading passages would produce much the same type of coverage–comprehension curve, but this is a point for future research.

One might wonder whether all words in the texts should have equal weight in our vocabulary coverage calculations. After all, some words may be absolutely crucial for comprehension, and a reader could have unexpected difficulties because a few of these critical words were unknown. However, this seems unlikely because if the words are truly central to text understanding, they are typically repeated multiple times in the text. Readers are highly likely, through such multiple exposures, to figure out the meaning of the few unknown words well enough to maintain reasonable comprehension. Therefore, we feel that the possibility of a few key words substantially impeding comprehension is remote, and believe our “equal weighting” procedure is both reasonable and supportable. In any case, it would be extremely difficult to devise a methodology for determining each word’s meaning “value” in a principled, reliable, and parsimonious manner.

The trend in the vocabulary coverage–reading comprehension relationship was clear in the data, as indicated by both mean and median scores, but it obscures a great deal of variation. The standard deviations for the means range between 3.88 and 5.30. This translates into the percentage of coverage ranging about 15 percentage points above and 15 percentage points below the mean level of comprehension at every vocabulary coverage point. This is illustrated in Figure 4, where about two thirds of the participants scored between the +1 *SD* and –1 *SD* lines on the graph. Clearly, there was a wide range of comprehension at each coverage percentage. This is also indicated by the minimum and maximum scores in Table 2. The minimum scores fluctuated between 3 and 9, with seemingly no relation to vocabulary coverage. For these learners, even high vocabulary coverage could not ensure high comprehension scores. With the maximum scores, we see the maximum 30 points at high coverage percentages, as we would expect. But we also see quite high maximum scores (24–28) even in the 90%–94% coverage band. This variation helps to explain why, although the mean comprehension scores rise in a relatively linear way with increased vocabulary coverage, the correlation figure was not as strong as one might expect. These results support the conclusion above that although vocabulary knowledge is an essential requirement of good comprehension, it interacts with other reading skills in facilitating comprehension. One might also conclude that with so much variation, it would be difficult to predict any individual’s comprehension from only his or her vocabulary coverage.

We also administered our instrument to 40 NS university students, and it makes sense to use their data as a baseline in interpreting the NNS results. They knew all or nearly all of the words on the vocabulary test: About half scored full points on the test, and 83% of the NSs scored between 99% and 100%. As expected from their presumably greater language proficiency, the raw comprehension scores of the NSs were slightly higher than the NNSs, and the NSs had lower standard deviations. However, it is somewhat surprising how small the NS margin is in this respect. At 98% vocabulary coverage, the NSs achieved a mean of 70.3% comprehension, compared to the NNS mean of 68.3%. The figures are similarly close at 99% coverage (NS, 72.1%; NNS, 71.0%) and 100% coverage (NS, 78.2%; NNS, 75.3%).⁶ *T*-tests showed that none of these differences were statistically reliable (all *p* > .05). Overall, we find that the NS margin is nonsignificant, and ranges from only 1 to 3 percentage points. This small differential pales in comparison to the variation in comprehension caused by differences in vocabulary coverage. For example, among the NSs, the difference in comprehension was almost 8 percentage points (70.3% → 78.2%), between 98% and 100% coverage, whereas for the NNSs it was 7 percentage points (68.3% → 75.3%). This far outstrips the NS margin of 1 to 3 percentage points, and it appears that vocabulary coverage may be an even stronger

factor in reading comprehension than being an NS/NNS speaker.

The NS data revealed the same linearity demonstrated in the NNS data, albeit within a very truncated coverage range (98%–100%). The mean scores rose, as did the maximum scores, although the minimum scores meandered within a tight range in a similar manner to the NNSs. Overall, the NS results were clearly parallel to, but slightly higher than, the NNS results, which provides additional evidence for the validity of the NNS findings discussed earlier in the section.

Influence of Background Knowledge on the Vocabulary Coverage–Reading Comprehension Relationship

Research has shown that background knowledge is an important factor in reading. Consequently, we were interested in how it affects the vocabulary coverage–reading comprehension relationship. To explore this, we compared the results from the *Climate* (high background knowledge) and *Mice* (low background knowledge) texts (see Table 3 & Figure 5).

One would assume that if a learner has considerable background knowledge about a reading topic, it should help them and increase their comprehension at any particular vocabulary coverage percentage. We indeed find this for the higher

TABLE 3
Vocabulary Coverage vs. Reading Comprehension (*Climate* vs. *Mice* Texts)

	90% ^a	91%	92%	93%	94%	95%	96%	97%	98%	99%	100%
C ^b – <i>N</i>	3	4	5	6	23	36	56	86	128	132	174
M ^c – <i>N</i>	18	14	34	27	60	110	120	110	72	54	13
C – Mean ^d	15.67	14.75	17.20	14.83	20.04	18.31	20.59	21.06	21.34	22.14	22.87
C – <i>SD</i>	9.50	1.89	3.70	7.68	6.43	5.68	5.50	5.38	4.93	4.51	3.81
M – Mean ^d	15.06	15.43	16.12	15.59	16.85	18.12	17.83	17.76	18.99	19.28	18.69
M – <i>SD</i>	4.66	4.33	4.66	4.68	4.55	3.91	4.11	4.32	3.89	4.70	4.92
C – Median ^d	16	15.5	18	16	22	18.5	20	22	22	23	24
M – Median ^d	14.5	15	17	16	17.5	19	18	17	19	20	19
C – % ^e	52.2	49.2	57.3	49.4	66.8	61.0	68.6	70.2	71.1	73.8	76.2
M – % ^e	50.2	51.4	53.7	52.0	56.2	60.4	59.4	59.2	63.3	64.3	62.3
C – Minimum	6	12	13	3	9	7	9	5	6	5	9
M – Minimum	8	9	4	6	6	9	7	6	7	5	7
C – Maximum	25	16	22	24	28	27	30	29	29	30	30
M – Maximum	26	24	25	24	26	29	26	26	25	28	25

Note.

^aVocabulary coverage.

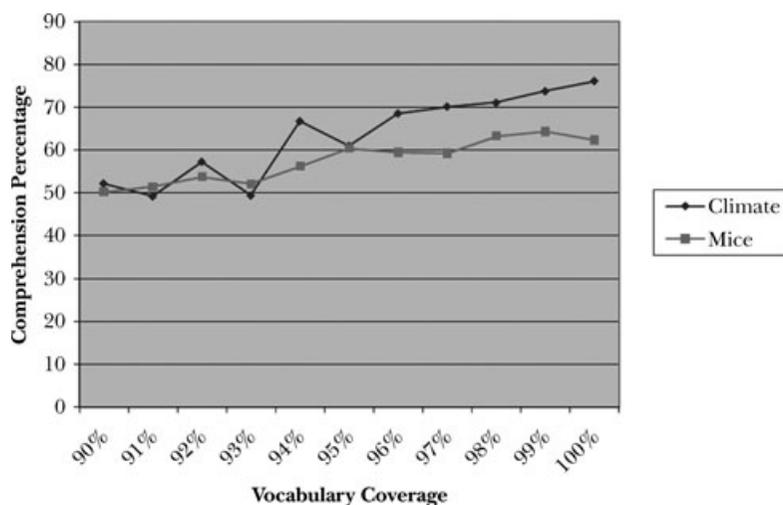
^bC = *Climate* results.

^cM = *Mice* results.

^dCombined graphic organizer and multiple-choice comprehension tests (Max = 30).

^ePercentage of possible comprehension (Mean ÷ 30).

FIGURE 5
Vocabulary Coverage vs. Reading Comprehension (Climate vs. Mice Texts)



vocabulary coverages (*t*-tests, $p < .05$ for 94% and 96%–100% coverages; $p > .60$ for 90%–93% and 95% coverages). From about the 94% vocabulary coverage (although not 95%), the learners were able to answer a greater percentage of the comprehension items correctly for the Climate text than for the Mice text. The advantage ranged from 7.8 to 13.9 percentage points (disregarding the 95% coverage level, where the means differed by only .6). In addition, for the Climate text, the maximum scores are at or nearly at full points (30) for the higher vocabulary coverage percentages, whereas for the Mice reading, they fall short by several points. Together, this indicates that background knowledge does facilitate comprehension in addition to vocabulary knowledge.

It might be argued that some of this difference could be attributable to the different texts and their respective tests. While it is impossible to totally discount this, the texts were chosen to be similar to each other in difficulty and in terms of writing style, and as a result they produced very similar Flesch–Kincaid readability figures (Climate = Grade Level 9.8; Mice = Grade Level 9.7). Similarly, the comprehension tests were comparable, following the same format for both readings, with the same number of GO and MC items. Although some of the difference in comprehension may be attributable to differences in text and tests, we feel that the majority of the effect is due to the difference that was designed into the study: background knowledge.

At the 90%–93% vocabulary coverage levels, neither text has a clear advantage for comprehension. This may be because there is no comprehension difference at these vocabulary coverage levels, or it may simply be that we did not have enough participants at these levels to produce stable enough results to indicate a trend.

The number of participants at each vocabulary coverage level is also informative. For the Climate text, the vast majority of participants knew 97% or more of the words in the text, and there were relatively few who knew less than 95%. For the Mice text, there were noticeably more participants knowing 90%–96% of the words, but many fewer knowing 98%–100% of the words. Thus, overall, participants knew more of the words in the Climate reading than the Mice reading. These figures suggest that higher background knowledge tends to go together with higher vocabulary knowledge. This is not surprising, as degree of exposure (Brown, Waring, & Donkaewbua, 2008) and the need for specific vocabulary to discuss particular topics (Hulstijn & Laufer, 2001) are important factors facilitating the acquisition of vocabulary. The more one engages with a topic, the more likely it is that vocabulary related to that topic will be learned. Thus, the amount of background knowledge should concurrently increase with the vocabulary related to a topic. This highlights the importance of maximizing exposure in L2 learning because it has numerous beneficial effects. Vocabulary knowledge and background knowledge were addressed in this study, but

maximizing exposure should facilitate a better understanding of numerous other linguistic elements, as well, such as discourse structure, schemata, and grammatical patterning.

Although the Climate text generally showed an advantage in comprehension over the Mice text, the basic trend indicated between vocabulary coverage and reading comprehension is similar for the two texts, which made it sensible to combine their data, and the combined results reported in the previous section are a good representation of the participants' reading behavior.

IMPLICATIONS

This study has expanded on previous studies in terms of the scope of the research design (direct comparison of coverage and comprehension, comprehensiveness of the vocabulary and reading measurements, and number and diversity of participants). It also offers a reconceptualization of the coverage-comprehension relationship, suggesting an ongoing linear relationship rather than discrete coverage "thresholds" (which were never going to be completely explanatory because some comprehension will always occur below a threshold, no matter where it is placed). Therefore, we feel that our study offers a much more comprehensive description of the coverage-comprehension relationship than has been previously available.

Our results indicate that there does not appear to be a threshold at which comprehension increases dramatically at a specific point in vocabulary knowledge growth. Instead, it appears that there is a fairly straightforward linear relationship between growth in vocabulary knowledge for a text and comprehension of that text. This relationship (based on the particular texts and assessment tasks in this study) can be characterized as an increase in comprehension growth of 2.3% for each 1% growth in vocabulary knowledge (between 92% and 100% of vocabulary coverage).⁷ Moreover, this relationship should be reasonably reliable. Our study included a large and varied participant pool (661 participants with a number of L1s), involved a dense sampling of vocabulary knowledge for each text in the study, included texts of both higher and lower degrees of background knowledge, and employed a comprehension measure that generated a wide range of variation in the resulting scores. Given the numerous articles and studies discussing the relationship between vocabulary and comprehension, our findings should provide some clarity to a still unresolved issue. Our conclusion is simple: There

does not appear to be a threshold level of vocabulary coverage of a text. Rather, as a higher level of comprehension is expected of a text, more of the vocabulary needs to be understood by the reader. As to what represents an adequate level of comprehension, this issue goes beyond the matter of a given text and task, and includes reader purpose, task goals, and a reader's expected "standard of coherence" (see Linderholm, Virtue, Tzeng, & van den Broek, 2004; Perfetti, Marron, & Foltz, 1996; van den Broek, Ridsen, & Husebye-Hartmann, 1995).

A more specific conclusion from our study (and also from Laufer, 1989, 1992, and Hu & Nation, 2000) is that high vocabulary coverage is an essential, but insufficient, condition for reading comprehension. Even high levels of vocabulary coverage (98% to 100%) did not lead to 100% comprehension, demonstrating, appropriately, that vocabulary is only one aspect of comprehension. Other factors play a part and need to be addressed by pedagogy that helps learners improve their comprehension. Nevertheless, vocabulary coverage was shown to be a key factor, as participants could only manage comprehension scores of about 50% with lower levels of coverage. Clearly, any reading abilities and strategic skills the readers had at these lower levels were not sufficient to fully overcome the handicap of an inadequate vocabulary. It seems that readers need a very high level of vocabulary knowledge to have good comprehension of the type of academic-like text used in this study. If one supposes that most teachers and learners aspire to more than 60% comprehension, vocabulary coverage nearing 98% is probably necessary. Thus, Nation's (2006) higher size estimates based on 98% coverage seem to be more reasonable targets than the lower size estimates based on the earlier coverage figure of 95%, especially in cases of students working with academic texts. This means that learners need to know something on the order of 8,000–9,000 word families to be able to read widely in English without vocabulary being a problem. This higher size target surely indicates that vocabulary learning will need to be given more attention over a more prolonged period of time if learners are to achieve the target. This is especially so because pedagogy needs to enhance depth of knowledge as well as vocabulary size. In the end, more vocabulary is better, and it is worth doing everything possible to increase learners' vocabulary knowledge.

Our results also showed that the relationship between vocabulary and reading was generally congruent for texts at two different levels of

background knowledge from 94% vocabulary coverage to 100% coverage (with one dip at 95%). Figure 5 illustrates the reasonably parallel paths of growth in comprehension for both the higher and lower background knowledge texts at each percentage of increasing vocabulary knowledge (from 94% vocabulary coverage on). This parallel growth pattern is apparent with the increasing n size, beginning at 94% of vocabulary coverage. One implication of this outcome is that comprehension tests drawing on varying levels of background knowledge should generate more or less the same relationship between vocabulary growth and reading comprehension growth, although higher background knowledge should generally lead to some advantage at higher coverage levels.

A final outcome of this study is the demonstration of intensive text measurement. It proved possible to develop a dense sampling of vocabulary knowledge for a text of reasonable length, a text that might be used in an authentic reading task in high intermediate to advanced L2 settings. The approach used in this study would not be difficult to replicate in further studies involving texts of similar length and general difficulty levels. Likewise, it was possible to create a dense, yet reliable, comprehension measure for a text without resorting to specific vocabulary items to increase the number of test items per text. It is a challenge to generate 30 test items for a 700-word passage, not retesting exactly the same points, while maintaining a reasonable level of test reliability. The use of GOs that reflect the discourse structure of the text provided an effective means to produce a comprehension measure that densely sampled from the text. As a fairly new type of comprehension test, GO fill-in items forced participants to recognize the organizational structure of the text and provide responses that were compatible with that organization. Moreover, the GO items were superior in terms of reliability to the more traditional MC items ($GO = .81$; $MC = .59$).

One further research goal that could be pursued in future studies is to identify the range of vocabulary knowledge coverage (in relation to comprehension) that would be appropriate for texts that are the focus of reading instruction and teacher support. The present study suggests that readers should control 98%–99% of a text's vocabulary to be able to read independently for comprehension. But what is a reasonable range of vocabulary coverage in a text that will allow students to comprehend that text under normal reading instruction conditions (perhaps 2 hours of instruction devoted to a 600- to 700-word text)? For example, can a student, with instructional support, learn to read and understand

texts for which they originally know only 85% of the words? Or should instructional texts aim for a vocabulary benchmark percentage of 90%–95%? At present, we do not have reliable coverage estimates that would be informative for instruction or materials development. The present study can provide one baseline for the beginnings of such research.

NOTES

¹The various studies incorporated different measurements and different units of counting, for example, individual words and word families.

²Although 95% coverage is probably not high enough for unassisted reading, it may be a good level for teacher-supported instructional texts.

³One reviewer raised questions about the potential problems readers of cognate languages might have with nonwords in checklist tests. One question concerns the type of nonwords in which unacceptable affixation was attached to real words (e.g., *fruital*; *fruit* is a real word). Learners could know the root form of the word, but not that the affixation was wrong, and so mistakenly check such a nonword. They would, therefore, be penalized, even though they knew the root form. Our checklist test avoided this problem by only using plausible nonwords that do not exist in either root or derived form in English (e.g., *perin*). A second potential problem is that a nonword may resemble a real word in the learner's L1 by chance. A learner may think that it is a cognate, and believe they know the nonword. It may seem that marking this as incorrect would be unfair to the learner, but in fact, it is important that this behavior be penalized. It is similar to reading a real text, and assuming that a "false friend" actually has a cognate meaning. If this incorrect meaning is maintained during reading, it can lead to much confusion and misunderstanding. For example, Haynes (1993) has shown that learners are quite resistant to changing their minds about an assumed meaning of an unknown word, even if the context indicates the meaning is not viable (the interpretation of the context usually changed instead). The checklist test is about accurate recognition of word forms. If learners mistake a nonword for a real L2 word (even if it resembles a real L1 word), this needs to be accounted for in the test, as the learners do not really know the word. Thus, we believe that neither of the above problems were applicable to our nonwords.

⁴K-R 21 is a simpler formula for calculating reliability coefficients than K-R 20; however, the assumption under K-R 21 is that all the test items are equal in difficulty. K-R 21 tends to underestimate the reliability coefficients if the assumption of equal difficulty cannot be met.

⁵The data were not normal, due to most participants scoring vocabulary coverage rates in the higher 90%. Thus, there was a strong ceiling effect for vocabulary coverage, confirmed by scatterplots, which probably served

to depress the correlation figure somewhat. Actually, the $\rho = .41$ correlation is reasonably strong given that approximately 60% of the participants were in the 98%–100% vocabulary coverage range. Despite a considerable compression of vocabulary variance, there is a reasonable correlation and a strong linear plotline from the data. We speculate that if we were able to include more of the participants with lower vocabulary levels, the added variance may have led to a higher correlation figure.

⁶We disregarded the 95% and 96% coverage levels, as we only had one NS participant at each of these levels.

⁷This rate of increase is similar to that found in Hu and Nation's (2000) study.

REFERENCES

- Alderson, J. (2000). *Assessing reading*. New York: Cambridge University Press.
- Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Huston (Ed.), *Advances in reading/language research, Vol. 2* (pp. 231–256). Greenwich, CT: JAI Press.
- Bernhardt, E. (1991). *Reading development in a second language*. Norwood, NJ: Ablex.
- Bowey, J. (2005). Predicting individual differences in learning to read. In M. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 155–172). Malden, MA: Blackwell.
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*, 136–163.
- Chapelle, C., Enright, M., & Jamieson, J. (Eds.). (2007). *Building a validity argument for the Test of English as a Foreign Language*. New York: Routledge.
- Circuit training. (2005, September 24). *The Economist, 376*, 97.
- Clapham, C. (1996). *The development of IELTS: A study of the effects of background knowledge on reading comprehension*. Cambridge Studies in Language Testing 4. Cambridge: Cambridge University Press.
- Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly, 38*, 78–103.
- Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 11*, 341–363.
- Grabe, W. (2004). Research on teaching reading. *Annual Review of Applied Linguistics, 24*, 44–69.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.
- Haynes, M. (1993). Patterns and perils of guessing in second language reading. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 46–65). Norwood, NJ: Ablex.
- Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*, 689–696.
- Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language, 23*, 403–430.
- Hudson, T. (2007). *Teaching second language reading*. New York: Oxford University Press.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). New York: Cambridge University Press.
- Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes–no vocabulary test: Correction for guessing and response style. *Language Testing, 19*, 227–245.
- Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 5*, 539–558.
- Johns, A. (1981). Necessary English: A faculty survey. *TESOL Quarterly, 15*, 51–57.
- Koda, K. (2005). *Insights into second language reading*. New York: Cambridge University Press.
- Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. In K. Koda (Ed.), *Reading and language learning* (pp. 1–44). Special issue of *Language Learning Supplement, 57*.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans to thinking machines* (pp. 316–323). Clevedon, England: Multilingual Matters.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). London: Macmillan.
- Laufer, B. (2000). Task effect on instructed vocabulary learning: The hypothesis of “involvement.” *Selected papers from AILA '99 Tokyo* (pp. 47–62). Tokyo: Waseda University Press.
- Laufer, B., & Sim, D. D. (1985). Taking the easy way out: Non-use and misuse of contextual clues in EFL reading comprehension. *English Teaching Forum, 23*, 7–10, 22.
- Linderholm, T., Virtue, S., Tzeng, Y., & van den Broek, P. (2004). Fluctuations in the availability of information during reading: Capturing cognitive processes using the Landscape Model. *Discourse Processes, 37*, 165–186.
- Long, D., Johns, C., & Morris, P. (2006). Comprehension ability in mature readers. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 801–833). Burlington, MA: Academic Press.
- Meara, P. (1992). *EFL vocabulary tests*. Swansea: Centre for Applied Language Studies, University of Wales, Swansea.
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4*, 142–151.

- Meara, P., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (Ed.), *Applied linguistics in society* (pp. 80–87). London: CILT.
- Milton, J., & Meara, P. M. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics*, 107–108, 17–34.
- Mochida, A., & Harrington, M. (2006). The yes/no test as a measure of receptive vocabulary knowledge. *Language Testing*, 23, 73–98.
- Nagy, W. E., Herman, P., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20, 233–253.
- Nassaji, H. (2003). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *Modern Language Journal*, 87, 261–276.
- 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? *Canadian Modern Language Review*, 63, 59–82.
- National Commission on Excellence in Education. (1983). *A nation at risk*. Washington, D.C.: National Commission on Excellence in Education.
- Oakhill, J., & Cain, K. (2007). Issues of causality in children's reading comprehension. In D. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 47–71). New York: Erlbaum.
- Ostyn, P., & Godin, P. (1985). RALEX: An alternative approach to language teaching. *Modern Language Journal*, 69, 346–355.
- Perfetti, C., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 227–247). Malden, MA: Blackwell.
- Perfetti, C., Marron, M., & Foltz, P. (1996). Sources of comprehension failure: Theoretical perspectives and case studies. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 137–165). Mahwah, NJ: Erlbaum.
- Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *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. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Rosenfeld, M., Leung, S., & Oltman, P. (2001). *The reading, writing, speaking, and listening tasks important for academic success at the undergraduate and graduate levels*. (TOEFL Monograph Series MS-21). Princeton, NJ: Educational Testing Service.
- Schmitt, N. (Ed.). (2004). *Formulaic sequences*. Amsterdam: Benjamins.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12, 329–363.
- Schmitt, N. (2010). *Researching vocabulary*. Basingstoke, England: Palgrave Macmillan.
- Sherwood, R. (1977). A survey of undergraduate reading and writing needs. *College composition and communication*, 28, 145–149.
- Snow, C., Burns, S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, D.C.: National Academy Press.
- van den Broek, P., Risden, K., & Husebye-Hartmann, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In R. F. Lorch, E. J. O'Brien, & R. F. Lorch, Jr. (Eds.), *Sources of coherence in reading* (pp. 353–373). Hillsdale, NJ: Erlbaum.
- van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., et al. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology*, 96, 19–30.
- van Gelderen, A., Schoonen, R., Stoel, R., de Glopper, K., & Hulstijn, J. (2007). Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology*, 99, 477–491.
- Weir, C., & Milanovic, M. (Eds.). (2003). *Continuity and innovation: Revising the Cambridge Proficiency in English Examination 1913–2002*. Cambridge: Cambridge University Press.
- What's wrong with our weather. (1999). In X. Zhai, S. Zheng, & Z. Zhang (Eds.), *Twenty-first century college English, Book 2* (pp. 269–271). Shanghai, China: Fudan University Press.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Zwann, R., & Rapp, D. (2006). Discourse comprehension. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 725–764). Burlington, MA: Academic Press.

 APPENDIX

A Vocabulary and Reading Comprehension Test

(Please finish this test in 100 minutes.)

Please fill in the following information before you take the test (5 minutes).

Name _____

University _____

Major _____

Year at university _____

Sex _____

Age _____

Mother tongue _____

- 1) How long have you been learning English? _____ years and _____ months.
- 2) Do you use English materials for classes other than English? Check one.
Yes _____ No _____
- 3) Have you travelled to an English-speaking country? Check one.
Yes _____ No _____
If yes, for how long? _____ years and _____ months.
- 4) Have you taken any national or international English tests? Check one.
Yes _____ No _____
If yes, what is the name of the test? _____.
What is your most recent score? _____.

Supporting Information

Additional supporting information may be found in the online version of this article:

Part I: Vocabulary Test for Reading (15 minutes)**Part II:** Reading Comprehension Test (80 minutes)

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

 Upcoming in *Perspectives*

Among our favorite feel-good words these days, in a world that all too often seems short of sentiments of sharing and collaboration, is that of ‘community.’ Beyond its general meaning, foreign language educators have over the last decade and a half been challenged to consider the particular meanings that the *Standards for Foreign Language Teaching* document and educational practice have foregrounded with the term ‘communities.’ How educators have done so imaginatively, energetically, and critically—that is the topic to be explored in *Perspectives* 95,2 (July 2011) under the title “*Connecting language learning to the community.*” As usual, the column includes voices that consider the topic from different vantage points, revealing its timeliness as well as its, perhaps, surprising complexity.
