Reading in a Foreign Language April 2016, Volume 28, No. 1
ISSN 1539-0578 pp. 63–78

What can readers read after graded readers?

Jeff McQuillan
Center for Educational Development
United States

Abstract

Nation (2014) concluded that most of the vocabulary one needs to read challenging texts in English can be acquired incidentally through voluminous reading. This study examines possible texts that second language (L2) readers can use to move from controlled-vocabulary materials such as graded readers, which go up through approximately the 4,000-word-family level, to more challenging texts such as newspapers, classic novels, and academic texts, at the 9,000-word-family level. An analysis of a set of popular fiction series books found that such books can provide a sufficient amount of input, with 98% vocabulary coverage, so as to serve as one possible "bridge" to more challenging texts.

Keywords: extensive reading, graded readers, vocabulary acquisition, text coverage, comprehensible input

Studies in both first and second language acquisition have shown that new vocabulary can be acquired incidentally through the reception of comprehensible input via reading (Krashen, 2004a). While the importance of reading in vocabulary acquisition is generally acknowledged among second language (L2) researchers, there has been disagreement as to whether the vocabulary one can acquire solely through reading can "take you all the way," to a point where you acquire a sufficient number of words to understand more challenging texts, including classic novels, newspapers, and academic writing.

Cobb (2007, 2008) held that reading alone cannot provide enough input to allow L2 readers to acquire enough words to handle challenging texts within a reasonable amount of time. McQuillan and Krashen (2008), however, argued that L2 acquirers can indeed get enough input to acquire most of the vocabulary they need through voluminous, self-selected reading, such as that provided by extensive reading programs (Day & Bamford, 1998; Mason, 2013).

Acquiring Sufficient Vocabulary to Read Challenging Texts

Nation (2014) attempted to settle this debate through a corpus analysis. His approach to the question was based upon three assumptions:

(a) In order to have "adequate comprehension" of text, one needs to know at least 98% of the

words in that text (Hu & Nation, 2000; Schmitt, Jiang, & Grabe, 2011);
(b) To achieve this 98% vocabulary coverage for challenging texts, one must know the 9,000 most frequently occurring word families in English (Nation, 2006); and
(c) To have a reasonable chance of acquiring an unknown word family, one must encounter it at least 12 times in text.

The logic Nation presents here is straightforward: If we know how many words one must read to encounter the first 9,000 word families at least 12 times, we can provide estimates of the amount of text and time L2 readers need to acquire a sufficient vocabulary to handle challenging texts.

Are Nation's assumptions reasonable? Several researchers (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Schmitt et al., 2011) have argued that "adequate comprehension" of text requires somewhere between 95% and 98% vocabulary coverage. While vocabulary knowledge is not the sole factor in determining reading comprehension, it has clearly been shown to be an important one in both the first language (L1) and L2 research (Anderson & Freebody, 1981; Hu & Nation, 2000). Laufer and Ravenhorst-Kalovski (2010), for example, found that vocabulary knowledge accounted for 64% of the variance in reading comprehension scores.

The exact percentage of words a reader needs to know to understand a text depends on how "adequate" comprehension is defined. Laufer and Ravenhorst-Kalovski suggest that 95% is the "minimum" coverage needed, with 98% (or more) being "optimum." In choosing the higher figure of 98%, Nation attempted to provide a conservative estimate of the percentage of words a reader needs to understand text independently.[1]

The other key assumption made by Nation – that 12 exposures to an unknown word are sufficient to acquire the word – is based on previous studies that produced differing estimates both above and below that figure (e.g., Brown, Waring, & Donkaewbua, 2008; Pellicer-Sanchez & Schmitt, 2010; Waring & Takaki, 2003). These analyses attempted to determine the number of exposures to a word needed to, in Nation's words, "develop something approaching rich knowledge" of a word (p. 2). In Pellicer-Sanchez and Schmitt (2010), for example, unknown words that occurred at least 10 times in the text were acquired 80% of the time, as measured by a meaning recognition test (Table 1, p. 41). In Waring and Takaki (2003), at least 15 repetitions were required for a similar level of success (72%).

Based on these and other similar studies, Nation's use of 12 occurrences as a threshold for acquisition appears to be an attempt to find a middle ground between competing estimates. Nation and other researchers acknowledge that acquisition of vocabulary depends on more than just the number of exposures to the acquired word, and any estimate depends on one's criteria for determining the "depth" of knowledge as well as its breadth (Wesche & Paribakht, 1996).

Nation (2014) analyzed a corpus comprised of 25 novels taken from Project Gutenberg (http://www.gutenberg.org). He also provided estimates of how long it would take a reader to read that amount of text, assuming a reading speed of 150 words per minute.

Table 1 shows Nation's results for the number of words that one would need to read in order to

encounter a word family at least 12 times in his chosen corpus of novels, and a calculation of the time required to read them. Estimates are broken down by 1,000-word-family groups, from the 2nd to the 9th 1,000-word-family levels.

Table 1. *Amount of input and time needed to acquire the 2nd through the 9th most frequently occurring 1,000-word families in English*

1,000- Word-Level List	Amount to Read	Hours Needed Per Level (@150 wpm)	Cumulative Hours
2,000	200,000	22	22
3,000	300,000	33	55
4,000	500,000	56	111
5,000	1,000,000	112	223
6,000	1,500,000	167	390
7,000	2,000,000	222	612
8,000	2,500,000	278	890
9,000	3,000,000	333	1,223
Total	11,000,000	1,223	

Note. Data from Nation (2014), Table 4

Assuming you know the 1,000 most frequently occurring words in English already, Nation estimated that you would need to read approximately 200,000 words in order to have a reasonable chance of acquiring most of the words in the 2,000-word-family level. After acquiring most of the words in the 2,000-word-family level, you would then need to read another 300,000 words in order to encounter most of the words in the 3,000-word-family level at least 12 times, and so on.

As shown in Table 1, one would need to read approximately 11,000,000 words to reach the 9,000-word-family level, and that this feat would take about 1,200 hours to complete. At one hour per day, this represents a little over three years of reading, very doable for a motivated adult or adolescent acquirer. One hour per day of reading is in line with what is expected of university students in the United States for out-of-class assignments. Nearly half of all American professors expect their students to do at least six hours of homework outside of school per week, or nearly an hour per day (Sanoff, 2006).

If Nation's analysis and the assumptions behind that analysis are correct, it appears that free reading can indeed provide L2 readers with the opportunity to acquire the necessary vocabulary to handle challenging texts.

Between Graded Readers and Challenging Text

If free reading is sufficient, the next step is to determine what sort of texts L2 acquirers should read. Nation (2014, p. 11) noted that controlled-vocabulary materials such as graded reader series can provide students with enough input to reach approximately the 4,000-word-family level. But what should readers read after graded readers? How is this "gap" between the 4,000- and 9,000-word-family levels to be filled? The problem can be summarized this way:

Graded Readers > ?? > Challenging Texts

Mid-Frequency Readers

Nation (2014) proposed that the gap can be made up in part by the use of "mid-frequency readers." Mid-frequency readers are adaptations of texts that meet the 98% vocabulary coverage criterion at lower vocabulary levels than the texts were originally written. The texts are created by substituting the less frequently occurring words in the stories and novels with more frequently occurring synonyms, as well as by controlling the number of new word families the reader will encounter in the text (Nation & Anthony, 2013; Schmitt & Schmitt, 2014). Texts have been developed by Nation and others at the 4,000-, 6,000-, and 8,000-word-family levels. Nation's proposal to fill the gap is thus:

Graded Readers > 4K Readers > 6K Readers > 8K Readers > Challenging Texts

Mid-frequency readers can be an important source of input for English as a Second Language (ESL) and English as a Foreign Language (EFL) acquirers. However, the number of such readers is still small, and for copyright reasons, the mid-frequency readers have thus far been limited to adaptation of works that are in the public domain. There is also the question of interest: not all L2 readers will find the texts chosen for adaptation to be sufficiently engaging to do the kind of voluminous reading required to read several million words. But it is one possible path, and given enough adapted texts, one that could allow readers to acquire sufficient vocabulary to read more challenging texts.

Light Reading, Narrow Reading

Krashen has long advocated the use of self-selected "light reading" to bridge the gap between modified texts such as graded readers and challenging, academic texts (2004a, 2010). Light reading refers to the materials being read, and may include comic books, children's books, young adult fiction, popular adult fiction, and popular magazines. In particular, Krashen (2004b) advocates a specific approach to light reading called "narrow reading." In narrow reading, readers read books by the same author or on the same topics. An example of narrow reading is the use of series books, texts written by the same author and usually involving the same main characters in the same or similar settings (Hwang & Nation, 1989; Schmitt & Carter, 2000).

Narrow reading of series books takes advantage of the powerful influence of prior knowledge on comprehension (Eidswick, 2010). Once readers finish the first book or story in the series, they have considerable background knowledge about the characters and setting that in turn can facilitate comprehension of subsequent stories.

In narrow reading, readers also become familiar with the writer's style and word choices, as well as the proper nouns (character names, places). This in effect reduces the vocabulary load required for reading additional novels in the series. This vocabulary "recycling" is particularly

strong with narrative fiction written by a single author (Gardner, 2008).

Previous research with L2 adults confirms that popular series books are an effective way of promoting language acquisition. Cho and Krashen (1994, 1995a, 1995b), for example, studied a group of adult women immigrants to the United States who began reading series books as a way of improving their English. They read books in the *Sweet Valley* collection by Francine Pascal, a series of children's books about the adventures of two twin girls. They started with the easiest books in the series, *Sweet Valley Kids*. After finishing *Sweet Valley Kids*, they graduated on to the next set of books in the series, written at a slightly higher vocabulary level, *Sweet Valley Twins*.

From there, some of the women in the Cho and Krashen studies continued on to *Sweet Valley High*, written at a slightly more difficult level than *Sweet Valley Twins.* One reader continued on further (Cho & Krashen, 1995b). After reading dozens of the *Sweet Valley* series books, she read adult novels by best-selling author Danielle Steele, all within the space of one year. Not only did the women enjoy their reading, they made impressive gains in vocabulary knowledge as a result. The series books provided a bridge to more challenging texts written for adult native speakers. We can summarize Krashen's proposed path this way:

Graded Readers > Light Reading > Challenging Texts

While there is some case study evidence that L2 readers can move from graded readers to "ungraded," unsimplified texts (Uden, Schmitt, & Schmitt, 2014), at least two additional research questions are raised by Nation's results:

1. Is there an adequate amount of reading material to satisfy Nation's recommended amount of input up through the 9,000-word-family level?
2. Can these texts be read with sufficient vocabulary coverage (at or above 98%) to provide a smooth transition from where graded readers leave off (between the 3,000- and 4,000-word-family levels) and more challenging texts begin (the 8,000- and 9,000-word-family levels)?

This study seeks to answer both questions by analyzing a set of popular fiction series books in terms of the quantity of input they can provide, and the levels of vocabulary coverage they require.

Method

Materials

Selections were analyzed from a number of popular fiction series written for children, young adults, and adults, all of which are either freely available on the Internet or widely available commercially (see Appendix).[2] As in the case of Nation's corpus of 25 novels, text selection in this study did not follow any strict criteria for selection other than that the text might be of

interest to adult language acquirers. Various genres (adventure, detective, Western) were chosen to appeal to a wide range of readers, perhaps slightly more so than Nation's selection of more "classic" novels currently in the public domain. The texts were hoped to reflect the kind of reading adults do for pleasure, as noted in previous reader preference studies (Nell, 1988). Also included were popular teen and children's books that previous research has shown can appeal to adult L2 readers (e.g., Cho & Krashen, 1994). The analysis aimed to determine the percentage of vocabulary coverage from the 3,000- to the 8,000-word-family level for each series of novels, as well as the total number of words in the series.

Vocabulary Coverage

In some cases, an entire text (a complete novel) was analyzed; in other cases, a selection from the text of between 1,500 and 5,000 words was used from one of the novels in the series. It was assumed that most of the novels in a given series would be of roughly similar vocabulary difficulty, recognizing that variations might take place from book to book within a series. To check the assumption that smaller samples of text would produce equivalent results as a fuller analysis, small samples of text (1,500 words) were analyzed from the first novel of the *Twilight* series (Meyers, 2011) and compared to an analysis of the entire text. The results in terms of determining the 1,000-word-family level at which 98% coverage was obtained were identical for the complete text and the sample texts, indicating it was not necessary to analyze an entire novel in order to arrive at a reasonably accurate estimate of the vocabulary coverage needed to read it.

The texts were analyzed with either the VocabProfile-Compleat (VP-Compleat), online software available on Tom Cobb's Lextutor website (http://www.lextutor.ca) (for shorter texts), or the AntWordProfiler (Anthony, 2012, available from http://www.laurenceanthony.net/software/antwordprofiler/) (for longer samples and entire novels). Both programs provide the same breakdown of word-family frequency based on a classification of the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) into 1,000-word families, as was used by Nation (2014) for his analysis, and both programs yield identical or very similar results. Proper nouns were included in the percentage of vocabulary coverage, following Nation (2006). (For a fuller discussion of the BNC itself, see Nation (2004); for the BNC and COCA, see Nation (2014)).

Total Number of Words

In addition to vocabulary coverage, estimates were also made of the number of total words (tokens) included in all the books of the book series. For some of the series books used in the analysis, total series word count was based on the average number of words per page from one of the books in the series multiplied by the total number of pages in the entire series as found on an e-book vender website (Amazon.com). This was used for series books where the length of the books in the series varied considerably, including the legal thrillers of John Grisham, the Suzanne Collins's *Hunger Games* series, the *Child Called It* series by Dave Pelzer, and J.K. Rowling's *Harry Potter* series.

For series books that appeared to have a fairly consistent number of words and pages in each book in the series, the word count was calculated from a sample book, multiplied by the number

of books in the series (Victor Appleton's *Tom Swift* books, Zane Grey's Westerns, R.L. Stine's *Goosebumps*, *Sweet Valley High*, *Sweet Valley Twins*, *Sweet Valley Kids*, Gertrude Chandler Warner's *The Boxcar Children*, and Agatha Christie's mysteries). "Fairly consistent" was defined as having no more than a 10% variation in total pages or total words from the average page or word count for the series, determined by examining at least five different books from each series.

For two of the series available in electronic format (the *Twilight* series and the Detective Larose series by Arthur Gask), all the books of the series were analyzed in order to check the accuracy of the methods of word count estimation used with the other series. For the *Twilight* series, the actual word count from the books in electronic format was 586,748. The estimated word count, using the number of pages per book in the series reported on Amazon.com (2,752) multiplied by the average number of words on a single printed page of the novel (200), was 550,400, a difference of 6%. For the Detective Larose series, the actual word count was 2,400,002 from electronic versions of the books. The estimated word count, using the number of words in a sample book (83,900) multiplied by the total number of novels in the series (27), was 2,265,300, a difference of only 5%. The methods of estimating word counts for the series were considered sufficiently accurate for the purposes of this study.

Results

Vocabulary Coverage

Table 2 lists all of the series books analyzed, sorted using Nation's criterion of 98% vocabulary coverage percentage, from the 3,000- up through the 8,000-word-family level. Vocabulary coverage is reported at each 1,000-word-family-level, with bolded figures indicating the level at which the text reaches 98% vocabulary coverage.

Books originally written for children and "tween" audiences fall mostly in the 4,000- and 5000-word-family levels (*The Boxcar Children Mysteries*, *Sweet Valley Kids* and *Sweet Valley Twins*, and *Goosebumps* series). The *Harry Potter* series is also found at these lower levels, but surprisingly, so are the Hercule Poirot mysteries of Agatha Christie, written for adult readers.

Three series written largely for teens (*Child Called It*, *Twilight*, and *Sweet Valley High*) have 98% vocabulary coverage at the 6,000-word-family level, as does another popular series written for adults, the legal thrillers of John Grisham. At the top end of the coverage rankings, at the 7,000- and 8,000-word-family levels, are three older series written during the early and middle parts of the 20th century: the juvenile adventure series *Tom Swift*, and two series written for adults (Arthur Gask's detective stories and Zane Grey's Westerns). Perhaps most surprising is the rank of the popular trilogy *Hunger Games* by the American writer Suzanne Collins, which despite having an intended audience of teenagers, also comes in at the 8,000-word-family level for 98% vocabulary coverage.

Coxhead (2012) also included an analysis of the *Hunger Games* trilogy in her study, using a larger sample of text and drawing from all three books in the series instead of just the first book,

as was done in this analysis. She determined readers would need at least the 9,000-word-family level for 98% vocabulary coverage, a somewhat higher estimate than the result obtained here. This difference may in part be due to variations in the vocabulary level across books in the series, as well her use of the British National Corpus rather than the BNC-COCA list used in this analysis, that latter incorporating both British and American texts.

Table 2. *Vocabulary coverage of selected popular fiction series books*

Popular Fiction Series Books	3K	4K	5K	6K	7K	8K
The Boxcar Children Mysteries (*The Boxcar Children*)	97.4	**98.1**	98.6	99	99.3	99.3
Sweet Valley Kids (*Lila's Secret*)	96.5	**98.0**	98.4	98.7	98.8	98.8
Goosebumps (*Welcome to the Dead House*)	96.9	97.8	**98.9**	99.4	99.5	99.6
Sweet Valley Twins (*Jessica On Stage*)	98.8	97.8	**98.4**	98.6	99.1	99.1
Harry Potter (*Harry Potter and the Sorcerer's Stone*)	95.1	97.1	**98.3**	98.8	99.1	99.2
Agatha Christie's Poirot Mysteries (*The Mysterious Affair at Styles*)	96.1	97.5	**98.3**	98.8	99.1	99.4
Child Called It (*Child Called It*)	96.7	97.1	97.9	**98.6**	98.9	99.3
Sweet Valley High (*Double Love*)	94.2	96.3	97.9	**98.6**	98.9	99.2
John Grisham's Legal Thrillers (*The Firm*)	95.9	97.1	97.7	**98.3**	99.0	99.3
Twilight (*Twilight*)	95.3	96.7	97.5	**98.0**	98.6	98.9
Tom Swift (*Tom Swift and His Electric Rifle*)	93.2	95.5	96.9	97.8	**98.3**	98.5
Arthur Gask's Detective Gilbert Larose (*The Master Spy*)	94.5	96.3	97.1	97.7	**98.1**	98.3
Hunger Games (*Hunger Games*)	93.1	95.3	96.8	97.4	97.8	**98.7**
Zane Grey's Westerns (*Betty Zane*)	91.6	93.4	95.9	96.9	97.6	**98.0**

Note: The names of works from which the text selections analyzed were taken are shown in parentheses, with references found in Appendix.

Total Number of Words

Table 3 includes the number of books in each series, the 1,000-word-family level at which they can be read with 98% coverage (taken from Table 2), and an estimate of the total word count for that series.

The number of books and total word count vary widely across series, as would be expected. Series written for children and teens generally have the greatest number of texts in them, although Zane Grey's Westerns have the highest total word count of the series analyzed, at just over five million words. Table 3 also shows how one related set of series (*Sweet Valley Kids*, *Sweet Valley Twins*, and *Sweet Valley High*) has a sufficient number of texts to provide adequate input for acquiring the word families of the 4,000-, 5,000-, and 6,000-word-family levels. This is

consistent with Cho and Krashen's (1994, 1995a, 1995b) results in improving the reading proficiency of their adult ESL subjects.

Table 3. *Estimated word count for popular series books*

Popular Series Books	Level @ 98% Vocab. Coverage	Number of Books in Series	Estimated Word Count
The Boxcar Children Mysteries	4K	139	1,400,000
Sweet Valley Kids	4K	88	528,000
Goosebumps	5K	179	4,800,000
Sweet Valley Twins	5K	118	2,400,000
Harry Potter	5K	7	1,000,000
Agatha Christie's Poirot Mysteries	5K	42	3,300,000
John Grisham's Legal Thrillers	6K	22	3,200,000
Twilight	6K	4	586,000
Child Called It	6K	3	194,000
Sweet Valley High	6K	143	4,300,000
Tom Swift	7K	29	1,200,000
Arthur Gask's Detective Gilbert Larose	7K	27	2,400,000
Zane Grey's Westerns	8K	52	5,200,000
Hunger Games	8K	3	240,000

Adequacy of Series Books as a Source of Input

Table 4 combines the information from Table 1 on Nation's recommended volume of reading for the 5,000- to 9,000-word-family levels with the total number of words from the selected series books found in Table 3 that would be appropriate for that level. Note that texts that can be read at 98% coverage at a given 1,000-word-family level are used to help readers acquire words in the *next* 1,000-word level. For example, texts that can be read at 98% coverage at the 4,000-word-family level are used to help the reader acquire the word families at the 5,000-word-family level, and so forth. A similar logic is used by Nation (2014) in the creation of the mid-frequency readers: the 4,000-, 6,000-, and 8,000-level readers are intended to help the reader acquire words at the 5,000-, 7,000-, and 9,000-word-family levels, respectively.

In Table 4, the total word count for the 5,000-word-family level shown in the last column is the sum of the word counts for the series books that can be read at 98% at the 4,000-word level (that is, *Boxcar Children* (1,400,000 words) plus *Sweet Valley Kids* (528,000 words), for a total of 1,980,000 words). The total word count shown for the 6,000-word-family level is the sum of all those books that can be read at the 98% at the 5,000-word-family level, and so forth.

Table 4 shows that for each 1,000-word-family level from 5,000 to 9,000, popular series books can provide sufficient input to meet Nation's recommended amount of reading to acquire most of the word families at those levels. For some levels, a single popular fiction series could theoretically provide enough input to acquire the majority of the word families. Readers could, for example, get all 1,500,000 words of input needed to acquire words at the 6,000-word-family-level by reading the Agatha Christie mysteries, which have a total of more than three million words. Nation (2014) points out, however, that exposure to a mix of reading genres may offer a

better chance to acquire the widest variety of word families up through the 9,000-word-family-level.

Table 4. *Minimum number of words needed and corpus size of series books for the 5th through 9th 1,000-word families*

1,000 Word List Level	Nation's Minimum Number of Words to Read	Estimated Word Count for Series Books
5,000	1,000,000	1,928,000
6,000	1,500,000	11,500,000
7,000	2,000,000	8,280,000
8,000	2,500,000	3,600,000
9,000	3,000,000	5,440,000

Discussion

The results provide support for the position that second language acquirers can indeed move from modified texts such as graded readers to challenging texts in English through the use of popular fiction series books. Table 2 shows that there is sufficient input each step of the way, all at Nation's recommended 98% vocabulary coverage, such that readers can follow a "smooth path" on their way to reading challenging texts.

Moreover, this reading can be done in a reasonable amount of time. After a little more than one year of reading an hour per day, L2 acquirers would be able to read popular novels such as Agatha Christie's Hercule Poirot mysteries, John Grisham's legal thrillers, and the teen vampire series, *Twilight*.[3] A little over three years of reading takes readers all the way to the 9,000-word-family level.

The results of the present study with regard to the suitability of children's books for ESL readers appear to conflict with the findings of Webb and Macalister (2010). In that study, the researchers found that the vocabulary knowledge needed to read "children's literature" was similar to that required by challenging adult texts. However, Webb and Macalister's study dealt with a very specific type of children's reading material which, one could argue, is not typical of the category: "quality" stories from a literary magazine for children, prepared and distributed by a government office of education. These stories are quite different from the sort of popular reading materials that, if one goes by book sales figures, most children actually read for pleasure outside of school. For Webb and Macalister's sample of texts, a 98% vocabulary coverage required knowledge of the first 10,000 most frequently occurring word families. This is far above the level of vocabulary required to read popular series books such as the *Harry Potter*, *Goosebumps*, and *Sweet Valley* novels, as reported in Table 2.

Not every adult reader will be interested in books and stories written for children and adolescents, of course, or even in reading fiction. The particular selection of series books analyzed in this study is just one possible path from graded readers to challenging text.

Readers could choose, instead, other combinations of simplified and unsimplified texts. An

alternate path, especially at the 3000- to 5,000-word-family levels where texts intended for adults are more difficult to find, could include modified input from language teaching podcasts (McQuillan, 2006) such as those provided by the British Council (http://www.britishcouncil.org); expository text from the more than 115,000 entries in Wikipedia's Simple English website (http://simple.wikipedia.org); and, for those inclined more to current events, Voice of America's "Learning English" service, which provides several controlled vocabulary news stories each day, with more than a million words posted on its website each year (http://www.voanews.com). All three sources are aimed at intermediate adult acquirers and contain materials in the 3,000- to 5,000-word-family levels. Once readers reach the 5,000-word-family level, the number of options for unsimplified input increases, including the use of series books written for adults.

Not all readers will read graded readers or other simplified texts to the point of a smooth transition to unsimplified texts. Waring (2008, cited in Uden, Schmitt, & Schmitt, 2014) noted many L2 readers often prefer to "wean themselves off" graded readers and try more difficult "authentic" texts, finding them more motivating and interesting. Thus, while having less than 98% vocabulary coverage may be more of a struggle, some readers apparently find it worth the effort. As shown in Table 2, many popular books can be read at 95% coverage at only the 3,000- to 4,000-word-family levels, such as the best-selling *Twilight* and *Harry Potter* series, as well as John Grisham's legal thrillers.

This desire to read above one's proficiency level may occur not only among those reading at the level of graded readers (up through the 4,000-word-family level), as Waring suggested, but at any point along the path toward fluency. Some readers who are able to read juvenile fiction written at the 5,000 word-family-level, for example, may be more interested in a crime novel written at the 6,000- or 7,000-word-family-level. Still other readers may not have commercial graded readers available, or access to texts that fall into the next level of difficulty above their current level. For these readers, there may be little choice but to attempt to read texts for which the vocabulary coverage is less than 98%.

Reading Below 98% Vocabulary Coverage

Is it possible for these L2 readers to read books at a level *below* the 98% vocabulary coverage threshold that Nation and others recommend, and thus to access texts potentially more interesting to an adult reader much sooner? Uden et al. (2014) provided evidence that students who "graduate" from the highest-level graded readers are in fact able to make the transition directly to what they term "unsimplified" novels written for native speakers of English, even when reading below a 98% vocabulary coverage level. The researchers provided profiles of four L2 readers who begin the study with vocabularies in the 5,000- to 6,000-word-family level range. The participants first read advanced Cambridge graded readers (the most advanced of which had 98% coverage at around the 4,000-word-family level), and then two adult novels in English, *The Innocent* by Ian McEwen, and *Peaceful Warrior* by Dan Millman, which have 98% coverage at the 7,000- and 8,000-word-family levels, respectively.

The participants' vocabulary coverage for the most advanced graded readers was 98-99%, but only around 95% for the unsimplified novels. Measures of reading speed and comprehension were given on two of the advanced graded readers and the two unsimplified novels. Participants

were also asked to rate their enjoyment of the books and the ease or difficulty of the text.

Both comprehension scores and a rating of "reading ease" dropped when readers moved from the graded readers to the novels, but for three of the participants, not dramatically so. Reading ease was measured on a six-point scale. The average reading-ease score for all four participants for the graded readers was 5.93 (SD = 0.32), while for the novels it was 4.27 (SD = 1.20). This indicates that, while there was a difference in perceived reading difficulty between 98-99% coverage and 95% vocabulary coverage, participants still thought the novels were within their reach. Uden et al. concluded that most of the participants "made the jump to the ungraded novels without sacrificing much comprehension, reading speed, or satisfaction" (p. 19).

Note that the participants in Uden et al. read unrelated novels – each book was by a different author writing on a different theme, making the readers' task much more difficult than if they had read series books by a single author. A reader at the 3,000- to 4,000-word-family level, for example, may find the first Agatha Christie mystery or *Twilight* series novel a challenge at 95-96% vocabulary coverage, but the background knowledge about the characters and setting gained from the first book would likely make the reading of the subsequent books in the series easier.[4]

While the results of the current study showed a clear path from graded readers to higher-level texts, future research should examine successful L2 readers who've "made it" to more challenging texts to see if in fact they follow similar routes to fluency using popular reading materials, as some case studies suggest they do (Tse, 1996).

Pedagogical Implications

It would be wrong to conclude based on the results of this study that adult L2 readers should test their vocabulary levels and attempt to "match" themselves exactly to texts using a 95% or 98% vocabulary coverage criterion. A variety of factors can affect comprehension (and enjoyment) of a text in addition to the percentage of unknown words, including individual interest and background knowledge. As Schmitt et al. (2011) noted, even 100% vocabulary coverage of a text does not ensure perfect comprehension.

Fortunately, Nell (1988) found that readers can make judgments about the suitability of texts for their own pleasure reading fairly quickly, often with a very short sample of a text. A wide variety of books can now be sampled for free on electronic book websites to help individual adult L2 readers decide on an appropriate text. Many of the books included in this analysis, and all those in Nation's, are also available for free on sites such as Project Gutenberg.

For the classroom teacher, the best approach may be simply to provide lots of texts for students to choose from in a classroom library, allowing students to sample different books and make their own determinations on what to read. Such free reading approaches have already been shown to be very successful with L2 readers (Krashen, 2004a; Mason, 2013; McQuillan, 1998).

Reading popular fiction series books is not the only way L2 readers can move from graded readers to more challenging text, but given the variety of series available at different levels of

vocabulary difficulty, and the quantity of text that they typically provide, they should be given serious consideration by teachers and L2 acquirers alike.

Acknowledgments

I thank Warren Ediger, Stephen Krashen, Beniko Mason, Ken Smith, Paul Nation, and several anonymous reviewers for their helpful comments and suggestions on earlier drafts of this paper.

Notes

1. Different definitions of "adequate comprehension" have resulted, unsurprisingly, in a variety of "minimum" estimates by researchers. Schmitt et al.'s (2011) data suggest, however, that there exists no discernible break or threshold in vocabulary coverage below which comprehension can be clearly judged to be "inadequate" (p. 39), at least by any means other than an arbitrary one. Thus, debates about which vocabulary coverage percentage should be considered "minimum" or "optimum" are probably not very productive. Nation's threshold of 98% should be treated as a "safe harbor" estimate in his argument, a rough approximation likely to cover most readers reading most texts. For an argument that all "objective" determinations of minimal competence in educational measurement are psychometrically dubious, see Glass (1978); for a treatment of the issue of competency thresholds as they apply specifically to reading comprehension, see McQuillan (1997), especially his discussion of criterion cut-scores in reading assessments (pp. 3–5).

2. Note that Grisham's books are not strictly speaking series books, in that the characters and settings usually change from book to book, but were thought to be sufficiently similar in writing style and theme to be included here.

3. An hour of reading per day for 390 days brings one to the 6,000-word-family level. Interestingly, Beglar and Nation (2007), reporting on their Vocabulary Size Test, found that "initial studies using the test indicate that undergraduate non-native speakers successfully coping with study at an English speaking university have a vocabulary of around 5,000-6,000 word families" (p. 12). No further information on these studies is provided, however.

4. A possible objection to the use of popular series books, as opposed to Nation's mid-frequency readers, is that a significant percentage of the potentially unknown word families - i.e., those word families that are above the 98% coverage level - occur only once in the text. For example, in a separate analysis done on the *Twilight* series using the BNC lists, it was found that 62% of the 1,035 word families in the 7,000- to 20,000-word-family levels occur only once in the first book of the series. This clearly presents a greater burden on the reader than that encountered in Nation's mid-frequency readers, which by design do not contain low-frequency "one-timers." However, continued reading of the series helps mitigates this problem. By the end of the fourth book of the *Twilight* series, only 26% (552/2,133) of the words that are in 7,000- to 20,000-word-family levels will have been encountered only once, with more than a third (35%) appearing four or more times.

References

Anderson, R., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews* (pp. 77–117). Newark, DE: International Reading Association.

Anthony, L. (2012). AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software/antwordprofiler/

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

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.

Cho, K. S., & Krashen, S. (1994). Acquisition of vocabulary from the *Sweet Valley High Kids* series: Adult ESL acquisition. *Journal of Reading*, *37*, 662–667.

Cho, K. S., & Krashen, S. (1995a). From *Sweet Valley High Kids* to *Harlequins* in one year. *California English*, *1*, 18–19.

Cho, K. S., & Krashen, S. (1995b). Becoming a dragon: Progress in English as a second language through narrow free voluntary reading. *California Reader*, *29*, 9–10.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, *11*, 38–63.

Cobb, T. (2008). Commentary: Response to McQuillan and Krashen. *Language Learning & Technology*, *12*, 109–114.

Coxhead, A. (2012). Researching vocabulary in secondary school English texts: *The Hunger Games* and more. *English in Aotearoa*, *78*, 34-41.

Day, R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge: Cambridge University Press.

Eidswick, J. (2010). Interest and prior knowledge in second language reading comprehension. *JALT Journal*, *32*, 149–168.

Gardner, D. (2008). Vocabulary recycling in children's authentic reading materials: A corpus-based investigation of narrow reading. *Reading in a Foreign Language*, *20*, 92–122.

Glass, G. (1978). Standards and criteria. *Journal of Educational Measurement*, *15*, 237–261.

Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, *23*, 403–430.

Hwang, K., & Nation, I. S. P. (1989). Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers. *Reading in a Foreign Language*, *6*, 323–335.

Krashen, S. (2004a). *The power of reading: Insights from the research* (2nd ed.). Portsmouth, NH: Heinemann.

Krashen, S. (2004b). The case for narrow reading. *Language Magazine*, *3*, 17–19.

Krashen, S. (2010). Academic language proficiency: Acquired or learned? In *Selected papers from the Nineteenth International Symposium on English Teaching* (pp. 34–43). Taipei: Crane.

Laufer, B., & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*, 15–30.

Mason, B. (2013). The efficient use of literature in second language education: Free reading and listening to stories. In J. Bland & C. Lutge (Eds.), *Children's Literature in Second Language Education* (pp. 25–32). London: Bloomsbury Publishing.

McQuillan, J. (1997). *The literacy crisis: False claims, real solutions*. Portsmouth, NH: Heinemann.

McQuillan, J. (1998). The use of self-selected and free voluntary reading in heritage language programs: A review of research. In S. Krashen, L. Tse, & J. McQuillan (Eds.), *Heritage Language Development* (pp. 73–88). Culver City, CA: Language Education Associates.

McQuillan, J. (2006). Languages on the go: Tuning in to podcasting. *International Journal of Foreign Language Teaching*, *2*, 16–18.

McQuillan, J., & Krashen, S. (2008). Commentary: Can free reading take you all the way? A response to Cobb (2007). *Language Learning & Technology*, *12*, 104–108.

Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a Second Language: Selection, Acquisition, and Testing* (pp. 3–13). Amsterdam: John Benjamins.

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

Nation, I. S. P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, *26*, 1–16.

Nation, I. S. P., & Anthony, L. (2013). Mid-frequency readers. *Journal of Extensive Reading*, *1*, 5–16. Available at: http://jalt-publications.org/access/index.php/JER/issue/view/7

Nell, V. (1988). The psychology of reading for pleasure: Needs and gratifications. *Reading Research Quarterly*, *23*, 6–50.

Pellicer-Sanchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do *Things Fall Apart*? *Reading in a Foreign Language*, *22*, 21–55.

Sanoff, A. (2006, March 10). A perception gap over students' preparation. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/A-Perception-Gap-Over/31426

Schmitt, N., & Carter, R. (2000). The lexical advantages of narrow reading for second language learners. *TESOL Journal*, *9*, 4–9.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*, 484–503.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in text and reading comprehension. *The Modern Language Journal*, *95*, 26–43. doi: 10.1111/j.1540-4781.2011.01146.x

Tse, L. (1996). When an ESL adult becomes a reader. *Reading Horizons*, *31*, 16–29.

Uden, J., Schmitt, D., & Schmitt, N. (2014). Jumping from the highest graded readers to ungraded novels: Four case studies. *Reading in a Foreign Language*, *26*, 1–28.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*, 130–163.

Webb, S., & Macalister, J. (2010). Is text written for children useful for L2 extensive reading? *TESOL Quarterly*, *47*, 300–322. doi: 10.1002/tesq.70

Wesche, M., & Paribakht, T.S. (1996). Asessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review*, *53*, 11–40.

Appendix

List of Works Analyzed

Appleton, V. (1911). *Tom Swift and his electric rifle; Or, daring adventures in Elephant Land.* New York: Grosset & Dunlap.
Christie, A. (1920). *The mysterious affair at Styles*. New York: John Lane.
Collins, S. (2008). *The hunger games*. New York: Scholastic.
Gask, A. (1937). *The master spy*. New York: The Macaulay Company.
Grey, Z. (1903). *Betty Zane*. New York: Charles Francis Press.
Grisham, J. (1991). *The firm*. New York: Bantam Dell.
Meyers, S. (2011). *Twilight*. New York: Hachette.
Pascal, F. (1990). *Lila's secret*. New York: Bantam Skylark.
Pascal, F. (1984). *Double love*. New York: Bantam Skylark.
Pascal, F. (1989). *Jessica on stage*. New York: Bantam Skylark.
Pelzer, D. (1995). *A child called it: One child's courage to survive*. Deerfield Beach, Florida: Health Communications, Inc.
Rowling, J. K. (1998). *Harry Potter and the sorcerer's stone*. New York: Scholastic.
Stine, R. L. (2011). *Welcome to the dead house*. New York: Scholastic.
Warner, G. C. (1924). *The boxcar children*. New York: Rand McNally and Company.

About the Author

Jeff McQuillan, Ph.D. is a Senior Research Associate at the Center for Educational Development in Los Angeles, California, co-producer of *English as a Second Language Podcast*, and editor of *Learn English Magazine*. E-mail: jeff@eslpod.com