SECTION B: *Excerpt from:* ***Advancing Our Students' Language and Literacy: The Challenge of Complex Texts***

Marilyn Jager Adams is a research professor in the Cognitive, Linguistic, and Psychological Sciences Department of Brown University and former chief scientist for Soliloquy Learning Inc. She is the author of numerous scholarly papers and several books, including two landmark works: Beginning to Read: Thinking and Learning about Print *and* Phonemic Awareness in Young Children*. This article is adapted with permission of the Guilford Press from "The Challenge of Advanced Texts: The Interdependence of Reading and Learning," which Adams wrote for* Reading More, Reading Better, *edited by Elfrieda H. Hiebert, copyright 2009 by Guilford Press.*

The Vocabulary of Written Language

Reading educators have long appreciated that there is a very strong relationship between vocabulary and reading comprehension. But what exactly is it about the wording of texts that underlies this relation? Part of the answer is that written texts draw upon many more words than normally arise in

oral language situations.[20]

To gain insight into this phenomenon, Hayes and colleagues compared spoken language with texts.[21] For this study, they focused on trade publications rather than school materials, and the texts they used included preschool books, children's books, comic books, adult books, magazines, newspapers, and abstracts from scientific magazines. For comparison, they compiled and analyzed a variety of oral language samples, including language from prime-time adult television shows, children's television shows, mothers' speech to children ranging in age from infancy to adolescence, conversations among college-educated adults (including from the Oval Office), and adults providing expert wit- ness testimony for legal cases. Regardless of the source or situation and without exception, the richness and complexity of the words used in the oral language samples paled in comparison with the written texts. Indeed, of all the oral language samples evaluated, the only one that exceeded even preschool books in lexical range was expert witness testimony.

This difference between the wording of oral and written language must lie at the crux of the advanced literacy challenge, as it points to a profound dilemma. On the one hand, the extent of this disparity implies that the great majority of words needed for understanding written language is likely to only be encountered— and thus can only be learned—through experience with written text. On the other hand, research has taught us that written text is accessible—and thus permits learning—only if the reader or listener already knows the vast majority of words from which it is constructed. Indeed, research indicates that reading with com- prehension depends on understanding at least 95 percent of the words of a text.[22]

How Many New Words Do Readers Need to Learn?

So roughly how many words do kids need to learn in order to be proficient readers? This question raises the second key part of the vocabulary problem.

Suppose you counted the number of times each different word in this article occurred. What you would find is that there are a few words that I have used quite a number of times, and many, many others that

I used only once or twice. This distribution of word counts or frequencies is an example of what is known as Zipf's law.[23]

According to Zipf's law, every natural language sample is made up of relatively few words that recur over and over again, and many, many words that arise very infrequently. The type of natural language sample does not matter and, provided that it is not too short, neither does its size. That is, whether you counted all the words in a casual conversation, a lecture, a newspaper article, a whole book, or even a whole library's worth of books, you would find the same thing: of all the different words in your sample, a small number would occur over and over again, while many, many others would occur only once.

Zipf 's law may feel intuitively obvious. Less obvious, however, are its implications with respect to the vocabulary challenge.

An example may vivify the issue. Counting words that appear in relevant text is a common approach to making dictionaries. For example, if you wanted to make a dictionary for geologists, you might begin by gathering a sample of the kind of articles about geology that you think your customers would like to read and then counting the number of occurrences of all the different words within them. The goal is to make sure your dictionary contains all the words that your customers will want to look up most.

Similarly, as part of creating *The American Heritage School Dictionary*,[24] John Carroll and his colleagues were asked to figure out which words should be included by examining children's reading materials. To do this, the team gathered texts that had been written especially for children in grades 3 through 8, taking care that the collection as a whole captured the range of different kinds of text and topics that the children might read in amounts that were proportionate to how often they could be expected to read them. From across these materials, the team then extracted 10,000 excerpts, totaling 5 million words of text in all, which, after sorting, turned out to include 86,741 different words. Their job was then to figure out which of these 86,741 words arose sufficiently often to warrant inclusion in the dictionary.[25]

Enter Zipf's law. Just 109 very frequent words accounted for fully half of the vast sample of children's reading material that Carroll and colleagues had put together. Indeed, 90 percent of the sample was accounted for by just 5,000 relatively common words. At the other extreme, more than half of the words appeared only once. Still worse: the team estimated that the actual number of different words in the children's reading materials—that is, the number of different words that would have turned up if they had counted such texts exhaustively rather than just working with excerpts—would have totaled 609,606. Due to Zipf 's law, a sample of 5 million words was just plain too small even to identify—much less to judge the relative frequency of—the vast majority of words that might well have belonged in the dictionary.

But hold it. We are talking about materials that are specifically written for and meant to be understood by schoolchildren in grades 3 through 8. How can they possibly be expected to know more than 600,000 different words?

In fact, many of these words are cousins of each other. For example, if a child knows the word *shoe*, then she or he is unlikely to experience difficulty with *shoes*. Similarly, a child probably won't have trouble with word families like *walk*, *walked*, and *walking*. Pushing this reasoning further, vocabulary researchers Bill Nagy and Richard Anderson[26] have argued that students shouldn't have problems with

any sort of prefixing, suffixing, or compounding of a word, provided that the meaning of the word's base is preserved. As examples, they suggested that if children know the word *elf*, they should have little problem with *elfin* or with pairs such as *cow*/*cowhand*, *know*/*knowledge*, *therapy*/*therapeutic*, and *represent*/*misrepresent*. Eliminating all such "closely related" words from the word count that Carroll and colleagues had done for the dictionary, and keeping only base words plus affixed or compound words whose meanings are harder to figure out from their base words (such as *vice/vicious*, *well*/*farewell*, *shift*/*shiftless*, *fix*/*prefix*), Nagy and Anderson estimated that the actual number of separate words that children need be taught is closer to 100,000. If Nagy and Anderson's elimination rules were too aggressive given children's word sense, then the actual number might be double or triple their estimate. And, of course, if we extend concern from grade-school materials to advanced texts, the actual number must be larger still.

Developing Students' Vocabulary: Examining the Options

So, what is the best way to help students master the many, many words they must know to understand advanced texts? In broad terms, there appear to be only two options: (1) to endeavor to teach students the words they will need to know, and (2) to expect students to learn new words through reading.

Is direct vocabulary instruction worthwhile? Based on a highly regarded meta-analysis, the answer seems to be a resounding "yes."[27] Across studies involving a variety of students, instructional specifics, and outcome measures, the meta-analysis showed that direct vocabulary instruction significantly increases knowledge of words that are taught. Just as importantly, students who received vocabulary instruction were found to perform significantly better on global nonspecific vocabulary measures such as standardized tests, indicating that such instruction promotes learning of words beyond those that have been explicitly taught (e.g., being taught a word like *aquarium* helps with indirectly learning words like *aquatic*, *aqueduct*, and *aqueous*).

However, we must bear in mind that, by its very nature, direct vocabulary instruction admits coverage of precious few words relative to the magnitude of the challenge. Even if, beginning in grade 1 and continuing through grade 12, teachers consistently taught—and students perfectly retained—20 vocabulary words each and every week, the gain in vocabulary would total only 8,640 words in all (20 words × 36 weeks of school × 12 years), many times fewer than what is required.

Such considerations have led some scholars to argue that the only feasible means by which students might acquire an adequate reading vocabulary is through the process of inferring the meaning of each new word from its context in the course of reading.[28] Indeed, research shows that the probability that students understand and retain any given new word that they encounter in print is 0.05.[29]

So how far will this get them? Researchers have (generously) estimated that median, middle-class, fifth-grade students read close to 1,000,000 words of text per year, in school and out.[30] Based on Carroll and colleagues' research, we can expect a million words of reading to include roughly 17,200 different words. If we suppose that the students already know one-quarter of the words in their texts, then the number of new words they should encounter through this reading would equal 12,900 per year. Yet, if the likelihood that the students will understand and retain each of these words is only 0.05, then their vocabulary can only be expected to grow by 645 per year, giving them but 5,160 new words by the time they graduate from high school.

Recalling that even texts that are for students in grades 1 through 8 presume knowledge of at least 100,000 different words, it is clear that both estimates for learning vocabulary fall way short of the need. At the same time, however, both estimates also seem at odds with the intuitive sense that a high school student need be neither a genius nor a tireless scholar to read and understand most materials written for grade-school children.