The Challenge of Advanced Texts:

The Interdependence of Reading and Learning

Marilyn Jager Adams

Department of Cognitive and Linguistic Sciences

Brown University

Providence, Rhode Island

In:

Elfrieda H. Hiebert (Ed.),

Reading more, reading better:

Are American students reading enough of the right stuff?

New York: Guilford Publications, 2009

Author Contact:

Marilyn Jager Adams
421 Pleasant Street
Belmont, MA 02478

Phone: 617-489-3044
Email: marilyn.adams@verizon.net

The Challenge of Advanced Texts:

The Interdependence of Reading and Learning

As I write this chapter, it has been 25 years since the release of *Our Nation at Risk* by the National Commission on Excellence in Education (NCEE;1983). NCEE was specifically authorized by the U. S. Congress and created by then Secretary of Education, Terrence Bell, to examine the quality of teaching and learning in our nation's schools, with special attention to the experience of teen-aged youths.

The principal motivation for the report was a growing concern that the United States's "once unchallenged preeminence in commerce, industry, science, and technological innovation is being overtaken by competitors throughout the world" (NCEE, 1983, p. 1). This, NCEE warned, was the wave of the future, as we were to expect an ever increasing redistribution of competitive capability throughout the globe. "Knowledge, learning, information and skilled intelligence are the new raw materials of international commerce… If only to keep and improve on the slim competitive edge we still retain in the world markets, we must dedicate ourselves to the reform of our educational system" (NCEE, 1983, p. 2).

NCEE also concluded that such educational reform was long overdue and direly needed:

> Our Nation is at risk. Our once unchallenged preeminence in commerce, industry, science, and technological innovation is being overtaken by competitors throughout the world… We report to the American people that the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people.
>
> If an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today we might well have viewed it as an act of war… As it stands, we have

allowed this to happen to ourselves... We have, in effect, been committing an act of unthinking, unilateral educational disarmament. (NCEE, 1983, p. 1)

Particularly worrisome at the time was a prolonged downward trend in the scores of U.S. high school students on the Scholastic Aptitude Test (SAT), as shown in Figure 8.1. From 1963 to 1980, the average score on the math section of the SAT had fallen by 36 points; on the verbal section, it had fallen by 54 points, an equivalent of 0.49 standard deviations (Price & Carpenter, 1978).

[Insert Figure 8.1]

When the SAT score decline stretched into the 1970s, the College Board engaged a panel to try to identify the underlying causes (College Entrance Examination Board, 1977). A first hypothesis to be checked, of course, was that the difficulty of the test had somehow shifted to students' disadvantage. But, no, to the contrary, indications were that scoring had become more lenient (Beaton, Hilton, & Schrader, 1977) and that the verbal passages had become slightly easier (Chall, Conard, & Harris, 1977). A second prominent hypothesis was that the decline was due to changes in the demographics of the test takers. In this case, the answer was positive, but only in part. Statistics showed that, over the 1960s, changes in the composition of the tested population accounted for as much as three-quarters of the test score decline – and, no wonder, for during this period the number of students taking the SAT tripled. However, when the test-taking population stabilized over the 1970s, the scores did not. Instead, the decline continued, even steeper than before, while the extent to which it could be ascribed to demographic shifts shrank to 30%, at best (Stedman, 1993). Furthermore, it was the scores of the strongest students, those in the top 10% of their class, that dropped the most; the scores of students toward the bottom of the distribution were holding steady or even gaining (Turnbull, 1985).

By the early 1990s, SAT scores appeared to have plateaued. The College Board decided to "recenter" the scale, basically adding about 25 points to the math scores and about 80 to the verbal scores so as to return the mean of each test to a value close to 500 points. (The scores in Figure 8.1 have been adjusted so that all are on the current "recentered" scale.) Beleaguered, the College Board also changed the name of the test from the Scholastic Aptitude Test to simply the SAT, with the letters standing for nothing.

Unfortunately, if the downward trend of the SAT scores had slowed, their levels have even today not recovered (see Figure 8.1). "SAT Scores Record Biggest Score Dip in 31 Years," headlined the *Washington Post* in 2006 (Matthews, 2006). "SAT Scores Drop for Second Year in a Row," headlined *U.S. News and World Reports* in 2007 (Kingsbury, 2007).

Not Just the SAT Scores

To be sure, whether scores on the SAT exams truly reflect relevant or important intellectual or academic proficiencies remains a topic of discussion (e.g., College Entrance Examination Board, 1977; Lewin, 2006; Rothstein, 2004; Stedman, 1996). On the other hand, there are a number of other indications that the development of more advanced levels of literacy is a special problem in the United States, and this is so whether compared with the literacy proficiency of other countries around the world or measured against our country's own standards and expectations.

Recent international studies of fourth graders' reading and literacy development show children in the United States to be above average, ranking ninth among the 35 countries participating in the 2001 assessment (Mullis, Martin, Gonzalez, & Kennedy, 2003). By contrast, relative to the 30 OECD (Organization for Economic Cooperation and Development, a.k.a. "developed") countries, U.S. high school students performed below average, ranking 17th of 30 and significantly outperforming only 5 of the others. The top 10% of U.S. high school students scored comparably to the top 10% of students in the other countries. However, when the scores are broken into reading levels, the United States showed a significantly smaller proportion of high school students at intermediate levels of proficiency and a significantly greater proportion at the lowest levels. Thus, while it seems the best of our students are world-class, our educational system is somehow failing the majority. With time, as so strongly voiced in *Our Nation at Risk*, this will matter increasingly, both personally and nationally.

Based on data from the International Adult Literacy Survey (IALS), the average literacy of U.S. adults ranked 12th among 20 countries of comparably high income (Sum, Kirsch, & Taggart, 2002). Moreover, within the five performance categories defined by the assessment framework, 45% of U.S. adults fell within the lowest two, placing them 16th among the 20 countries in the study. This relative underperformance occurred despite the fact that U.S. adults ranked first in the number of years of schooling and in academic degree completion. A particularly troubling aspect of Sum et al.'s (2002) analysis was the precipitous rate at which the relative literacy levels of U.S. adults seemed to be falling. As compared with their peers in other countries, the literacy levels of older U.S. adults (ages 36-45, 46-55, and 56 and older) ranked in the top five. In contrast, the average literacy performance of U.S. adults younger than 35 years old ranked in the bottom half of the distribution by every measure. Closer analyses of these younger adults showed that, among participants with just a high school diploma or less, those from the United States fell at the bottom of the pile, ranking 19th out of 20. Although scores were far higher for young U.S. adults who had completed 4 or more years of postsecondary education, they were still below the average of their same-aged and like-educated peers in the other countries.

Results of domestic assessments are not inconsistent with the international results. The 2003 National Assessment of Adult Literacy (Kutner et al., 2007) found more than 40% of U.S. adults unable to comprehend texts of moderate everyday difficulty. Only 13% of U.S. adults could read and understand the longer and more complex documents included in the assessment – and these documents were neither very long nor very complex.

Returning to school students, the National Assessment of Educational Progress (NAEP) documents slight but significant improvement in the average reading scores of fourth graders (4 points) and eighth graders (3 points) between 1992 and 2007. Similarly, the percentage of fourth and eighth grade students performing at grade level ("proficient") or above has increased over that period from 29% to 33% and from 29% to 31%, respectively (Lee, Grigg, & Donahue, 2007).

For twelfth graders, by contrast, the average reading score on the NAEP *fell* by 6 points between 1992 and the most recent assessment in 2005. Between 1992 and 2005, the percentage of twelfth graders performing at or above grade level ("proficient") fell from 40% to 35%; at the other end of the spectrum, those who scored *below* the basic level rose from 20% to 27% (Grigg, Donahue, & Dion, 2007).

The NAEP scores of the twelfth graders are jarring not just because they are so low but also because they are low compared to the fourth and eighth graders and even, working backwards in time, within same cohorts when in the fourth and eighth grades. Bear in mind, too, that 25% of eighth graders nationwide drop out of school before completing high school (Seastrom, Hoffman, Chapman, & Stillwell, 2005); presumably, then, those who stay in school, and therefore participate in the NAEP in the twelfth grade, disproportionately include the more successful and motivated students. In short, one can't help but wonder whether the twelfth graders were trying. After all, twelfth graders are an ultraworldly group, and there is little personal consequence for doing well or poorly on the NAEP.

Yet, college entrance examinations are voluntary, and performing well on them is the very point of taking them. In analyzing scores from its own college admissions exam, ACT, Inc. (known until 1996 as the American College Testing Program) compared them within-cohort to scores on its tests for eighth and tenth graders. For each of the cohorts examined and regardless of gender, race/ethnicity, or household income, the students were collectively on track in the eighth and tenth grades for better scores than they ultimately obtained in the twelfth grade. The report (ACT, 2006) concludes that there is a specific problem at the high school level. The same conclusion was drawn by the College Entrance Examination Board (1977) in the mid-1970s and again in the mid-1980s (Turnbull, 1985).

What Could Be the Problem?

The College Board's 1977 panel examined a number of factors that might have contributed to the SAT score decline. One of these, proposed by Jeanne Chall, was that the reading selections on the tests had somehow become too hard for the students. To test this hypothesis, Chall and her colleagues (1977) sampled passages from SAT tests administered between 1947 and 1975, using readability analyses to compare their difficulty. Yet, the data indicated that the SAT passages had become easier, not harder. Between 1963 and 1975, during the years of the score decline, the average difficulty of the test passages lay at the 11^{th}-grade level, which should have been solidly in range for 12^{th}-grade college-bound students. However, Chall et al. also evaluated popular 11^{th}-grade textbooks in history, literature, grammar, and composition. The average difficulty of the textbooks lay between the ninth- and tenth-grade levels. Could the SAT score decline have been due to this difference in the relative difficulty of the test and the school books? If students had neither practiced nor been instructed with reading materials as hard as the SAT passages, then one could hardly expect them to read the latter with competence and confidence.

Following on Chall et al.'s (1977) hypothesis, Hayes, Wolfer, and Wolfe (1996) undertook a complementary study in which they analyzed the difficulty of popular reading textbook series over time. Their results indicated that the difficulty of the text in these books, especially in grades 4 and up, had been reduced and, further, that this reduction was temporally aligned with the SAT score decline. As one indication, the average length of the sentences in books published between 1963 and 1991 was shorter than that of books published between 1946 and 1962. In the seventh- and eighth-grade textbooks, for example, the mean length of sentences had decreased from 20 words to 14 words – "the equivalent of dropping one or two clauses from every sentence," observed Hayes et al. (1996, p. 497). Meanwhile, the sophistication of the books' wording also declined. Hayes et al.'s analysis indicated that the wording of school books published from 1963 forward for eighth graders was as simple as that in books used by fifth graders before 1963, while the wording of twelfth-grade literature texts published after 1963 was simpler than seventh-grade texts published prior to 1963.

Such a disparity between the students' school books and the passages on the tests might well explain students' poor performance on their college entrance exams. More significantly, however, failing to provide instruction or experience with "grown-up" text levels seems a risky course toward preparing students for the reading demands of college and of life, in general.

This concern was recently raised again by ACT, Inc. (2006), in reviewing the poor performance of students on its college entrance exam. The maximum score on the reading component of the ACT college entrance exam is 36. ACT has found that scores of less than 21 predict reading difficulties in college coursework and also in the workplace. Among students who took the ACT exam in 2005, the scores of 51% -- more than half -- fell below the 21-point cut-off for college readiness in reading ability. Through analyses of student performance, ACT determined that the major stumbling block for the students was complex texts. More specifically, the ACT reading assessment is designed around three levels of textual complexity. For students whose overall performance fell below the 21-point benchmark, average performance on the complex texts was at chance levels. As students scored beyond the 21-point benchmark, their performance on the complex texts steadily increased but reached levels comparable to performance on texts classified as moderate and simple only among students who scored at least 35 of the 36 points possible.

The Wording of Natural Language

Hayes et al.'s (1996) analysis of shifts in the difficulty levels of school books was based on a clever and conceptually straightforward approach. As a reference against which to gauge usage of words in everyday "grown-up" text, they sampled words from a large number of English language newspapers. To represent the wording of the textbooks, they sampled approximately 1,000 words from each of more than 800 different school books, divided into sets corresponding to the particular era (prior to 1962 and subsequent to 1962) and the grade level (each of grades 1-12). They then developed separate word frequency corpora for each of the sets of school book samples and for the reference sample in the standard way, by counting the number of times each different word occurred and then ordering the words in each set of texts from most to least frequent (see, e.g., Carroll, Davies, & Richman, 1971, for a fuller explanation of the process of generating word frequency counts). Eliminated from the corpora were all proper nouns and also, so as to focus on meaning-bearing words, all closed-class words (i.e., grammatical words such as prepositions, determiners (e.g., the, this, some), conjunctions, and pronouns). In addition, all but the 10,000 most frequent words were eliminated from the reference sample (because word counts of low-frequency words are more sensitive to sample size and more susceptible to sampling error). After converting frequencies of the remaining words in each of the corpora to proportions (so that all were on the same scale regardless of the size of the sample), the researchers basically computed the extent to which words in the textbook samples were overused or underused relative to the reference sample taken from the newspapers and, by the same metric, relative to each other. Hayes et al.'s (1996) conclusion that the vocabulary of school books had been simplified was based on their finding that the wording of post-1962 school books had shifted toward more common words, as gauged by word frequencies in the reference sample.

Hayes et al. (1996) found that it was especially school books for students in grades 4 and up that were simplified in the years after 1962. They also found that, although the wording of school books for children generally increased with grade level across grades 1 through 8, the same was not true of high school books. Instead, across grades 9 through 12 (including texts for Advanced Placement courses), the difficulty levels of the literature books were shown to differ little from one another or from the grade 7 and grade 8 offerings. High school students' science texts were significantly more difficult than their English books. However, even among science texts, only those designated for Advanced Placement coursework evidenced difficulty levels comparable to the newspaper sample used as the benchmark reference. Because Hayes et al's high school text sample was relatively sparse, it is possible that it is not wholly representative. On the other hand, the measured simplicity of the high school books is consistent with the conclusions and speculations of others (e.g., ACT, 2006; Chall et al., 1977).

Over the years, following basically the same approach, Hayes and his colleagues quantified the relative lexical demands of a number of language domains . Many of his analyses affirmed that the lexical difficulty of texts varied predictably with the maturity or sophistication of the audience for which each was written. As examples, his analyses showed that the wording of scientific publications aimed at scientists, such as *Cell*, *Nature*, and *Science*, is more sophisticated than the wording of those written for lay persons, such as *Scientific American* (Hayes, 1992); that the wording of scientific publications, including college textbooks, is more difficult than that of newspapers (Hayes, 1992); that the wording of newspapers is significantly more difficult than that of popular adult novels (Hayes & Ahrens, 1988); that the wording of adult novels is generally more difficult than that found in novels written for grade-school children (Hayes & Ahrens, 1988); and that the wording of school children's books, with the exception of preprimers, is more difficult than that of books written for preschoolers (Hayes & Ahrens, 1988).

On the other hand, the results of some of Hayes's analyses are quite provocative with respect to the nature of literacy challenge. For example, while his analyses showed that textbooks had become progressively easier over this century (Hayes et al., 1996), they also indicated that the lexical difficulty of English-language newspapers had remained nearly constant (Hayes et al., 1996). Could this disparity be a factor in the declining circulation of newspapers?

In contrast, analyzing the wording of scientific magazines and journals published between 1930 to 1990, he found that sophistication of every single one that he evaluated, whether professional or lay, had increased dramatically over the decades (Hayes, 1992). If it is a national goal to inspire more students to become engineers and scientists, then shouldn't the difficulty of our schoolbooks have increased alongside? If a goal is to ensure that our students will be able to stay sufficiently informed about scientific progress to conduct business, reflect on policy, and manage their family's health and education, then at a minimum, shouldn't the difficulty of our school books keep pace with the difficulty of scientific publications aimed at the general public?

Even so, it was Hayes's comparisons of spoken and written language sources that seemed most telling. For these analyses, Hayes and Ahrens (1988) 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 witness testimony for legal cases. Regardless of the source or situation and without exception, the lexical richness of 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 literacy challenge, as it points up 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 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 comprehension depends on understanding at least 95% of the words of a text (Betts, 1946; Carver, 1994; Hu & Nation, 2000; Laufer, 1988).

The Word Frequency Spectrum

As it happens, the distribution of words in natural language conforms closely to certain mathematical functions. This fact has been very useful for computational linguists. It enables them, for example, to detect bias in their sampling and to determine the costs versus benefits of increasing their sample sizes. It has also been highly useful for corpus researchers. As a case in point, it enabled Hayes to derive quantitative indices of lexical richness; representing the distribution of the words in each corpus mathematically, he needed only to take the integral of the function describing each curve to find the area beneath and then to subtract the area beneath the first curve from the area beneath the second to compare the lexical density of the two distributions.

As another example, in building the *American Heritage Word Frequency Book* corpus, Carroll et al. (1971) sampled 500 words from each of about 1,000 texts written for children in grade 3 through 8, which included about 87,000 different words. Using the mathematics of word frequency distributions, Carroll et al. estimated that the actual number of different words in such materials—that is, the number that they would have found had they counted such texts exhaustively rather than sampling just 5 million words of excerpts—would have totaled more than 609,606.

Again, all of the materials from which Carroll et al. (1971) sampled were indeed specifically written for and meant to be understood by school children in grades 3 through 8. But how can grade school children possibly be expected to know more than 600,000 different words? Can Carroll et al.'s estimate possibly be correct?

Of the 5 million words of text from which Carroll et al. (1971) built their corpus, 50% of the sample was represented by just 109 very frequent words and 90% of the sample by just 5,000 frequent words. At the other extreme, 35,079 of the 86,741 distinct words in the corpus—more than 50%—turned up only once each in the entire sample. Further, the 86,741 distinct words that showed up at all in the sample represent only 15% of the total number of different words estimated to arise in texts for children in grades 3 through 8. The problem, explained Carroll et al., is that a sample of 5 million words is just plain too small to capture but a fraction of the total distribution.

In recent years, the proliferation of electronic texts has made it possible to compile much larger word frequency corpora. The British National Corpus was built from 100 million words taken from 4,124 different sources. Approximately 10% of the corpus was sampled from spoken language sources and 90% from written texts. Because the goal was to create a profile of present-day English, all of the sources were published after 1985, and the distribution of topics and text was guided by the pattern of print publishing in the United Kingdom. Ten percent of the corpus was drawn from imaginative texts and 90 percent from informative texts. (Leech, Rayson, & Wilson, 2001).

As described on the British National Corpus's website, this is a very large corpus: "To put these numbers into perspective, the average paperback book has about 250 pages per centimeter of thickness; assuming 400 words a page, we calculate that the whole corpus printed in small type on thin paper would take up about ten metres of shelf space. Reading the whole corpus aloud at a fairly rapid 150 words a minute, 8 hours a day, 365 days a year, would take just over four years" (British National Corpus, 2008).

Another drawback to Carroll et al.'s (1971) estimate of the total vocabulary in school children's texts is that words in the American Heritage corpus are distinguished solely by their spellings. Conflicting with this practice, research suggests that nouns and their regular plurals, such as cat and cats, share a common mental representation (Sereno & Jongman, 1997), as do the basic conjugations of the verbs, such as walk, walks walked, and walking (Stanners, Neiser, Hernon, & Hall, 1979). As Carroll et al. (1971) point out, had each noun and verb been counted only once rather than separately for each of its inflected forms, their estimate of the total number of words in schoolchildren's texts would have been considerably smaller. Thus, a further advantage of the British National Corpus for purposes of investigating the lexical demands of texts is that, within it, each word is tagged with its part of speech.

In a challenge to develop a vocabulary assessment for the National Assessment of Adult Literacy (Adams & Spoehr, in press), Kathryn Spoehr and I turned to the British National Corpus in quest of a means of examining the feasibility of selecting a spectrum of test items as a function of their frequencies. Because we were interested only in words that occurred sufficiently often that their relative frequency estimates might be reliable, we restricted attention to those that appeared at least 10 times in the full corpus or, equivalently, with a minimum probability of about once per 10 million words. Because our interest was in vocabulary, we focused exclusively on adjectives, adverbs, common nouns, and verbs. So as to afford a cleaner view of the number of different words available at each frequency, we included only the base forms of the nouns and verbs. The resulting distribution is shown in Figure 8.2.

[Insert Figure 8.2]

In Figure 8.2, a separate line is accorded to each of the four major parts of speech. The x-axis represents the frequency of the words, ranging from 100 or more per million to 1 per million. The y-axis indicates the number of words found at each frequency. As can be seen from the small spike on the far left of each curve, there are a few dozen words for each part of speech with frequencies of 100 per million or greater. Excepting those words, however, the graph shows that the frequency of occurrence of the vast majority of the words is extremely low. Fully 60% of the words that we counted had frequencies of less than one per million. Had we included proper nouns in our count, the number of infrequent words would have swelled still further and, in the present context, it is important to recognize that gaining familiarity with a great abundance of proper nouns is a key and core component of becoming literate. Further, though the total number of different words in the British National Corpus is 757,087, nearly 10 times more than captured in the American Heritage sample (Carroll et al., 1971), the percentage of different words that turned up only once was nearly identical in the two corpora (52% and 54%, respectively).

The formal explanation of this distribution of words is known as Zipf's law, named for the linguist who discovered it (Zipf, 1935). According to Zipf's law, the distribution of words in natural language discourse conforms to an inverse power function, and research confirms this to be true regardless of the topic, genre, language, level, or modality of the source of the word count.

In a nutshell, Zipf's law states is that 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—just as seen in the American Heritage and British National corpora. In turn, the J-shaped nature of such word frequency distributions developmentally divides their utility for guiding educational practice. Because large-scale corpora quite reliably index the relative frequency of the most common words (the initial "curl" on the J), they can be fruitfully used to design or evaluate texts and tests for primary grade students (see, e.g., Hiebert, 2005). In contrast, word frequency statistics from such corpora can offer little if any useful guidance where interest is shifted to the middle and upper grades, and this holds whether the focus is on vocabulary instruction, on methods or formulas for evaluating the readability or difficulty of text, or on assessments of vocabulary breadth. The impasse is that even in the largest corpora, the counted frequencies of the vast majority of words—nearly all beyond those few thousand and most common words—are too low to be statistically meaningful or trustworthy and are tied with thousands of other words and in any case.

Developing Students' Vocabulary: Examining the Options

Having a better sense of the magnitude of the challenge, let us turn to the question of how best to help students master enough words 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 on their own through reading.

Is direct vocabulary instruction worthwhile? Based on a meta-analysis by Stahl and Fairbanks (1986), the answer seems to be a resounding "yes." Across studies involving a variety of students, instructional specifics, and outcome measures, Stahl and Fairbanks found 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.

In the present context, however, we must also 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 x 36 weeks of school x 12 years), many times fewer than what is required.

Such considerations have led some scholars to conclude that the only feasible means by which students might acquire an adequate reading vocabulary is through the process of inferring the meanings of new words from their contexts in the course of reading (see Nagy, Herman, & Anderson, 1985). Under scrutiny, however, the promise of this proposal fades as well.

Let us suppose that, as Anderson, Wilson, and Fielding (1988) have estimated, median fifth-grade students read approximately 700,000 words of text per year. If these students were to read the 5 million words of text from which the American Heritage corpus was compiled, it would take them about 7 years—through the 12th grade. If we suppose that these students already know 12,000 of these words, then—again based on the American Heritage corpus—we can expect them to encounter about 75,000 new words in the course of this reading. Using Nagy, Anderson, and Herman's (1987) estimate of 0.05 as the probability that students retain the meaning of any given new word they encounter during reading, the upshot is that our fifth graders will have learned only 3750 new words by the time they graduate, equaling about 550 words per year. Counting time with informal text (e.g., mail) as reading and generously assuming that an estimated 15 minutes per day of in-school reading time is spent, without pause, reading at 200 words per minute, Nagy et al (1987) boost the estimated amount read per year to 1 million words, yielding a total of nearly 4,700 new words in just 5 years or about 840 per year. No matter, recalling that even texts that are for students in grades 1 through 8 presume at least 600,000 different words, it is clear that both estimates fall way short of the challenge. At the same time, however, both of these estimates 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.

Insights from Computational Models of Vocabulary Acquisition

Working from corpus statistics, the prospect of raising students' vocabulary to the demands of advanced texts seems a nearly impossible task. Yet, recent computation models such as the Latent Semantic Analysis (LSA) model developed by Tom Landauer and his colleagues (e.g., Landauer, 1998; Landauer & Dumais, 1997; see also Griffith, Stevers, & Tenenbaum, 2007) offer a different way of viewing the challenge.

The core mechanism underlying the LSA model is associative learning. The first step in training the model consists in inputting large quantities of text where the computer is programmed to remember each word and the context in which it occurred. The second step in training the model consists in creating associations or connections between the words and contexts and then mathematically collapsing or bundling them according to their commonalities and distinctions. The connections between words and their contexts are bi-directional and are weighted positively by the number of times the word occurs in the context and inversely by the number of different contexts in which the word occurs. By interconnecting all of the words and contexts in this way, a rich matrix of associations arises between the words, between the contexts, and between the words and the contexts in both directions.

As a concrete example, Landauer and Dumais (1997) trained the computer by having it "read" each of 30,473 articles (or, for long articles, the first 2,000 words) from the *Grolier Encyclopedia*. The sample totaled 4.6 million words of text in all, which the researchers estimated to be comparable in magnitude to the life long learning of a seventh grader. From these readings, the researchers created a matrix in which the 30,473 articles (the "contexts") stood as the columns and the rows were populated by each of the 60,768 words that had arisen in at least two of the encyclopedia articles. The researchers then used a mathematical procedure (singular value decomposition) to condense the separate connections between each word and context to a smaller set that optimally captured their overlap and separation. Again, the connections are bidirectional in the LSA model. As they extend both from words to contexts and from contexts to words, they also extend, by derivation, from words to other words, contexts to other contexts, and between words and contexts that did not co-occur in training.

To evaluate what the model had learned, Landauer and Dumais (1997) then tested it with 80 retired items from the synonym subtest of the Test of English as a Foreign Language (TOEFL). Each of these items presents an isolated test word for which the test taker is to select the best synonym from four alternative words. The model's performance, as based on the knowledge it gained from "reading" the encyclopedia articles, was then compared to performance of a large sample of applicants to U.S. colleges from non-English-speaking countries. The model's score was 64.4%; the people's score averaged 64.5%. Similarly, after the LSA model "read" an introductory psychology text, it performed nearly as well as college students on a multiple-choice exam (Landauer, 1998).

The LSA model has also been used with impressive success to gauge the quality and content of student essays and the coherence and conceptual density of reading materials. In one study, for example, students were first asked to write essays about the heart and circulatory system, and LSA was used to benchmark their prior knowledge based on what they wrote (Wolfe et al., 1998). Each student was then asked to read one of four passages on the topic, where the four passages ranged in sophistication from elementary level to medical school level. The results showed that students learned most when given a passage that was just a little—but not too much—more sophisticated than the knowledge shown in their essays.

To investigate LSA's vocabulary more closely, Landauer and Dumais (1997) compared the word-learning of the model to that of schoolchildren. In simulating schoolchildren, results showed that the probability that the model learned any given new word in any given new text was approximately 0.05, just like the students in Nagy et al.'s (1987) study.

Within the LSA model, any new input is represented in terms of the overall structure of the network or, equivalently, in relation to the representations of other words within it. It is for this reason that, like people, the amount that the LSA model learns from any set of readings depends on how much it already knows; for any given text, the larger its starting vocabulary, the more it learns.

But something else, too. Because the meaning is represented relationally, the connections that effectively define the meanings of words grow, shrink, and shift continuously, continually, and always in relation to one another. Thus, the addition or modification of any one connection impacts many others, pulling some closer together, pushing some farther apart, and otherwise altering the strengths and patterns of connections among words and contexts. Due to this dynamic, Landauer and Dumais (1997) found that with each reading the model effectively increased its understanding not just of words that were in the passage but also of words that were *not* in the passage. Measured in terms of total vocabulary gain, the amount the model learned about words that did *not* appear in a given reading was three times as much as what it learned about words that were in the reading.

"What?" we cry, "How can that be? How can reading a text produce increases in knowledge of words that it does not even contain! That is not credible! It makes no sense!" But wait. If we were talking about knowledge rather than words, then it would make lots of sense. Every concept—simple or complex, concrete or abstract—is learned in terms of its similarities, differences, and relationships with other concepts with which we are familiar. As a simplistic example, when we read about tigers, then, by dint of both similarities and contrasts, we learn more about all sorts of cats and, further, about every subtopic mentioned along the way. The more deeply we read about tigers, the more nuanced and complex these concepts and their interrelations become.

Words are not just words. They are the nexus—the interface—between communication and thought. When we read, it is through words that we build, refine, and modify our knowledge. What makes reading valuable and important is not the words themselves so much as the understandings they afford. The reason we need to know the meanings of words is that they point to the knowledge from which we are to construct, interpret, and reflect on the meaning of the text.

What is unique about the LSA model is not its implications about the essential structure of semantic memory. Cognitive psychologists broadly agree that the meaning of any word consists of bundles of features and associations that are the cumulative product of the reader's experience with both the word in context and the concepts to which it refers. In any given instance, only that subset of a word's meanings, usages, or features that are contextually relevant is activated (see, e.g., Gorfein, 2001). As examples, the activated meaning of the word fan differs, depending on whether the text is about a soccer fan, a ceiling fan, or a peacock's fan. What is unique about the LSA model is its demonstration that this structure and this dynamic can so richly and powerfully evolve through accrued experience with the various contexts in which words do and do not occur—that is, through reading.

There are potential extensions of the LSA model that also beg attention. For example, the concepts and relations that emerge and are strengthened through reading may belong to words that the student does not already know. That is exactly the point. As they grow in richness and complexity, the relationships that evolve within the network will increasingly support many new words and many new spheres of knowledge, whether in abstract or tied to related concepts or situations. Perhaps this is why people's vocabulary correlates so strongly with the amount and kinds of reading in which they have engaged (Gradman & Hanania, 1991; Stanovich & Cunningham, 1992, 1993).

Another thought that comes to mind is that, if reading results in so rich a network of knowledge through nothing more than overlaps and contrasts in association, then shouldn't students learn far more efficiently, given active, incisive comprehension strategies? Research indicates that comprehension strategies can be taught and suggests that doing so may improve comprehension (National Institute of Child Health and Human Devlopment [NICHD], 2000). However, comprehension strategies seem to do little to compensate for poor reading ability or weak domain knowledge (O'Reilly & McNamara, 2007). Instead, research repeatedly shows prior domain knowledge to be a prepotent predictor of students' ability to comprehend or to learn from advanced texts (Dochy, Segers, & Beuhl, 1999; Shapiro, 2004). In themselves, strong reading skills are also important, but they, too, seem to be of greatest advantage to students with strong domain knowledge (O'Reilly & McNamara, 2007).

Perhaps such findings should not be surprising. As broadly accepted by cognitive psychologists, there are two modes of reasoning (see Sloman, 1996; 2005). The first, most common, mode is knowledge-based. This sort of reasoning is rapid, extensive, and automatic; it is the sort of reasoning that LSA and neural networks statistically emulate. The second mode of reasoning is conscious and rule-based. Such logical analytic thought also warrants instructional attention in our schools, as it is our means of monitoring internal consistency and vetting our thoughts for bias and happenstance. However, no reasoning strategy, however well-structured, can rival the speed, power, or clarity of knowledge-driven understanding (Ericsson, Charness, Feltovich, & Hoffman, 2006); nor can it compensate for an absence of sufficient information.

Still another idea that raises itself is that the pathway to advanced texts might be well paved through other media such as educational videos. That is, if domain knowledge is the ticket to understanding advanced texts, then might the entry of students with less relevant knowledge or weaker reading skills be accelerated through well-designed videos? Unfortunately, evidence so far indicates that even while such educational media can be valuable for developing students' interest in a topic, their impact on the knowledge structures underlying language development and reading comprehension is minimal (Bus, de Jong, & Verhallen, 2008; Echols, West, Stanovich, & Zehr, 1996; West & Stanovich, 1991).

In keeping with this, the J-shaped curve that characterizes the word frequency distributions of large-scale corpora can be seen to represent two different categories of words. The high frequency words—those clustering at the curl of the J—are the nuclear or matrix words of the language. Writers depend on these words to carry the structure and flow of the language regardless of topic or genre. The low frequency words—those clustering at the staff of the J—are the information-bearing words. As such, the usage of these low frequency words is tied to specific topics and genre. Conversely, mastery of their meanings depends on learning about the specific topics and genre to which they pertain—and, in turn, such learning is strongly dependent on reading within those domains.

There may one day be modes and methods of information delivery that are as efficient and powerful as text, but for now there is no contest. To grow, our students must read lots, and more specifically they must read lots of "complex" texts—texts that offer them new language, new knowledge, and new modes of thought. Beyond the basics, as Hirsch (2006) has so forcefully argued, the reading deficit is integrally tied to a knowledge deficit.

Back to the Classroom: A Strategy for Developing Advanced Reading

The capacity to understand and learn from any text depends on approaching it with the language, knowledge, ideas, and reading skill that it presumes. It follows that, when assigning materials from which students are to learn, there are basically but two choices. Either the materials must be sufficiently accessible in language and concept for the students to read and understand on their own, or the students must be given help as they read. Some students receive such help in their homes, but many do not and, as I have argued elsewhere, this may be the major factor underlying the achievement gap (Adams, 2006; 2008). In any case, opportunity for one-on-one reading assistance in the typical school setting is limited, leaving for educators only the alternative of restricting assignments to materials that are within their students' independent reach. For weaker students, there follows the popularity of so-called high-low texts, intended to offer high interest or information along side low demands on vocabulary and reading skill.

Although the relaxation of school book complexity may be the consequence of our earnest efforts to ensure full curricular access to all, it is a solution with serious problems of its own. In terms of literacy growth, it is a solution that is vortically self-propagating and self-defeating, for it is a solution that denies students the very language, information, and modes of thought that they need most in order to move up and on.

Many have suggested using word-frequency information from large corpora to sequence vocabulary instruction or to evaluate readability, but this will not work. The five most frequent common nouns in the British National Corpus are <u>time,</u> <u>year,</u> <u>people,</u> <u>way,</u> and <u>man</u>. In the American Heritage Corpus, they are <u>time,</u> <u>people,</u> <u>way,</u> <u>words,</u> and <u>things</u>. These are not the sorts of words that distinguish advanced literacy status. Yet, as described earlier, the reader would need to cover massive amounts of text to encounter the sorts of words that do. Indeed, Carroll et al. (1971, p. xxii) estimate that the likelihood of encountering a word of just average frequency within the domain of grade school texts is about 1 in 13 billion. The problem with using these corpora for instructional guidance or readability estimates relates to the fact that in their creation, the goal was to capture as broad and representative a range as possible of their target domains, thus sampling reading materials so as to avoid redundancy and bias. In effect, then, both of these corpora are topic-neutral. If people were similarly to read a little of this and a little of that, and to retain the words they encounter independently of their contexts, then the likelihood of their knowing any given word might well resemble the probabilities reflected in these corpora.

An alternate and far more promising strategy follows directly from Zipf's law. Again, according to Zipf's law, every natural language discourse comprises a few words that recur again and again and many words that occur just once or only a few times. And, again, Zipf's law is shown to hold for virtually every natural language domain, regardless of its size, topic, modality, or sophistication. But do not be confused here: it is the shape of the distribution that does not change—the words underlying the curve vary from source to source.

In particular, the most frequent words in any language domain relate directly to its topic. Indeed, it is this fact that enables automatic topic spotting by computers. For example, a quick sampling of informational texts on Mars that I picked off the Internet affirms that, without exception and whether the intended audience was young children or scientists, the nouns <u>Mars</u> and <u>planet</u> are among the five most frequent in each. The balance of the dominant nouns in each text depend on the subtopic in focus – variously, its moons, its geography, our efforts at its exploration, etc.

In other words, combined with what else we know about literacy growth, Zipf's law prescribes a self-supporting strategy for developing the sorts of knowledge structures that complex texts require. We know that, even for young (Cunningham, 2006) and delayed (Share & Shaley, 2004) readers, any new word encountered (and attended) in print becomes a sight word with little more than a single encounter, provided its meaning is known (see Adams, 2008). We know that the more that students already know about the topic of a text, the greater their understanding and learning from its reading (O'Reilly & McNamara, 2007; Shapiro, 2004). We know that vocabulary strength predicts the speed and security with which students learn the meanings of unfamiliar words through direct instruction or study (Biemiller & Boote, 2006; Jenkins, Stein, & Wysocki, 1984; Perfetti, Wlotko, & Hart, 2005; Robbins & Ehri, 1994). Such prior knowledge also predicts the likelihood that students will learn the meanings of new words through context during reading (Daneman & Green, 1986; Herman et al., 1987; Shefelbine 1990; Sternberg & Powell, 1983). It predicts the probability with which readers correctly infer a new word's meaning from context (Morrison, 1996; Nassaji, 2004), and it predicts both the amount and the nature of the reasoning that is evidenced when they are asked to try to do so (Calvo, Estevez, & Dowens, 2003; Nassaji, 2004).

The challenge, then, lies in organizing our reading regimen such that each text bootstraps the language and knowledge that will be needed for the next. Zipf's law tells us that this can be done by scaffolding students' reading materials within topic.

Pick any topic about which you would like your students to learn—once started, there will be plenty of time for others. If the students are below-level, begin with shorter, simpler texts.

Teach the key words and concepts directly, engaging students in using and discussing them so as to be sure they are well anchored. As the students learn the core vocabulary, basic concepts, and overarching schemata of the domain, they will become ready to explore its subtopics, reading as many texts as needed or appropriate on each subtopic in turn. Gradually and seamlessly, they will find themselves ready for texts of increasingly greater depth and complexity. Better yet, as their expertise on, say, Mars, expands, they will find themselves in far better stead to read about Venus, Jupiter, earth sciences, and on and on.

Even while making incremental progress in this way, some may question whether offering advanced texts to reluctant readers is a realistic plan. Consider, however, that while no text on dinosaurs would pass a readability criterion for second-graders, many second-graders nonetheless read about dinosaurs with great satisfaction. Similarly, I have rarely met a Boston cabbie—no matter how much he decried reading—who wasn't quick to pick up and read a news article about the Red Sox. Knowledge is the prepotent determinant of reading comprehension. In theory—and as these two examples perhaps attest—the greatest cognitive and literacy benefits of text-based expertise depend on reading deeply in multiple domains and about multiple topics. We can and must do a better job of leading—and enabling—our students to do so. If education is the key to opportunity, then their options, in school and beyond, depend on it.

References

ACT, Inc. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading*. Iowa City, IA: Author.

Adams, M. J. (2006). The promise of automatic speech recognition for fostering literacy growth in children and adults. In M. McKenna, L. Labbo, R. Kieffer, & D. Reinking (Eds.), *Handbook of Literacy and Technology*, Volume 2, pp. 109-128. Hillsdale, NJ: Lawrence Erlbaum Associates.

Adams, M. J. (2008). The limits of the self-teaching hypotheses (and how technology might help). In S. B. Neuman (Ed.), *Literacy achievement for young children from poverty*, pp. 277-300. Baltimore, MD: Paul H. Brookes.

Adams, M. J., & Spoehr, K. T. (2008, in press). *Framework for the vocabulary assessment for the National Assessment of Adult Literacy (NAAL).* Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Anderson, R., Wilson, P., and Fielding, L. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly, 23*, 285-303.

Baroni, M. (2008, in press). Distributions in text. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook.* Berlin: Mouton de Gruyter.

Beaton, A. E., Hilton, T. L., & Schrader, W. B. (1977). *Changes in the verbal abilities of high school seniors, college entrants, and SAT candidates between 1960 and 1972.* New York: College Entrance Examination Board.

Betts, E. A. (1946). *Foundations of reading instruction.* New York, NY: American Book Co.

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44-62.

British National Corpus, version 2 (BNC World). (2001). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Available at <http://www.natcorp.ox.ac.uk/>.

British National Corpus (2008). The BNC in numbers. Retrieved June 15, 2008, from http://www.natcorp.ox.ac.uk/corpus/

Bus, A. G., de Jong, M. T., Verhallen, M. J. A. J., & van der Kooy-Hofland, V. A. C. (2008). Design features in living books and their effects on young children's vocabulary. In S. B.

Neuman (ed.), *Educating the other America*, pp. 263-276. Baltimore, MD: Paul H. Brookes Publishing.

Calvo, M. G., Estevez, A., & Dowens, M. G. (2003). Time course of elaborative inferences as a function of prior vocabulary knowledge. *Learning and Instruction, 13(6)*, 611-631.

Carroll, J. B., Davies, P., & Richman, B. (1971). *The American Heritage word frequency book*. Boston: Houghton Mifflin.

Carroll, J. B. (1971). Statistical Analysis of the corpus. In J. B. Carroll, P. Davies, & B. Richman, *The American Heritage word frequency boo*k, pp. xxi-xl. Boston: Houghton Mifflin.

Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behavior, 26,* 413-437.

Chall, J. S., Conard, S. S., & Harris, S. H. (1977). *An analysis of textbooks in relation to declining SAT scores*. New York: College Entrance Examination Board.

College Entrance Examination Board. (1977). *On further examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline.* New York: Author.

Cunningham, A. E. (2006). Accounting for children's orthographic learning while reading text: Do children self-teach? *Journal of Experimental Child Psychology, 95*, 56-77.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language, 25,* 1-18.

Dochy, F., Segers, M., & Buehl, M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69(1)*, 145-186.

Echols, L. D., West, R. G., Stanovich, K. W., & Zehr, K. S., (1996). Using children's literacy activities to predict growth in verbal cognitive skills: A longitudinal investigation. *Journal of Educational Psychology, 88*, 296-304.

Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (2006), *The Cambridge Handbook of expertise and expert performance.* Cambridge, U.K.: Cambridge University Press.

Gorfein, D. (2001). *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity.* Washington, DC: American Psychological Association.

Gradman, H., & Hanania, E. (1991). Language learning background factors and ESL proficiency. *Modern Language Journal, 75,* 39-51.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114(2)*, 211-244.

Grigg, W., Donahue, P., & Dion, G. (2007). *The nation's report card: 12th-grade reading and mathematics 2005 (NCES 2007-468).* Washington, DC: National Center for Educational Statistics, Institute of Education Sciences, U.S. Department of Education.

Hayes, D. P. (1988). Speaking and writing: Distinct patterns of word choice. *Journal of Memory and Language, 27*, 572-585.

Hayes, D. P. (1992). The growing inaccessibility of science. *Nature, 356,* 739-740.

Hayes, D. P., & Ahrens, M. G. (1988). Vocabulary simplification for children: A special case of "motherese"? *Journal of Child Language, 15*, 395-410.

Hayes, D. P., Wolfer, L. T., & Wolfe, M. F. (1996). Schoolbook simplification and its relation to the decline in SAT-verbal scores. *American Educational Research Journal, 33(2)*, 489-508.

Herman, P. A., Anderson, R. C., Pearson, P. D., & Nagy, W. E. (1987). Incidental acquisition of word meaning from expositions with varied text features. *Reading Research Quarterly, 22*, 263-264.

Hiebert, E. H. (2005). In pursuit of an effective, efficient vocabulary curriculum for the elementary grades. In E. H. Hiebert & M. Kamil (Eds.), *The teaching and learning of vocabulary: Bringing scientific research to practice,* pp. 243-263. Mahwah, NJ: Erlbaum.

Hirsch, E. D. (2006). *The knowledge deficit: Closing the shocking education gap for American children.* Boston: Houghton Mifflin.

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13(1)*, 403-430.

Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal, 21(4)*, 767-787.

Kingsbury, A. (2007, August 28). SAT Scores drop for the second year in a row. *U.S. News & World Report.* Retrieved from http://www.usnews.com/

Kurtz, S. (2002, August 21). Testing debate. *National Review Online.* Retrieved July 31, 2008, from http://www.nationalreview.com/kurtz/kurtz082102.asp.

Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y., & Dunleavy, E. (2007). *Literacy in everyday life: Results from the 2003 National Assessment of Adult Literacy (NCES 2007-480).* Washington, DC: National Center for Educational Statistics, Institute of Education Sciences, U.S. Department of Education.

Landauer, T. K. (1998). Learning and representing verbal meaning: The latent semantic analysis theory. *Current Directions in Psychological Science, 7(5)*, 161-164.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquistion, induction, and representation of knowledge. *Psychological Review, 104(2*), 211-240.

Laufer, B. (1988). What percentage of text-lexis is essential for comprehension. In C. Lauren & M. Nordmann (Eds.), *Special language: From humans to thinking machines,* pp. 316-323. Clevedon, UK: Multilingual Matters.

Lee, J., Grigg, W., & Donahue, P. (2007). *The nation's report card: Reading 2007 (NCES 2007-496).* Washington, DC: National Center for Educational Statistics, Institute of Education Sciences, U.S. Department of Education.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English based on the British National Corpus.* London: Longman.

Lewin, T. (2006, August 31). Students paths to small colleges can bypass SAT. *The New York Times*. Retrieved from http://www.nytimes.com.

Matthews, J. (2006, August 30). SAT records biggest score dip in 31 years. *Washington Post*. Retrieved from http://www.washingtonpost.com/.

Morrison, L. (1996). Talking about words: A study of French as a second language learners' lexical inferencing procedures. *Canadian Modern Language Review, 53*, 41-75.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools.* Chestnut Hill, MA: Boston College. Retrieved July 31, 2008, from http://timss.bc.edu/PIRLS2001.HTML.

Nagy, W. E., Herman, P., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20,* 233-253.

Nagy, W., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, *24*, 237-270.

Nassaji, H. (2004). The relationship between depth of Vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Canadian Modern Language Review, 61(1)*, 107-134.

National Center for Education Statistics (2007). *Digest of Education Statistics: 2007.* Washington, D.C.: U.S. Department of Education. Retrieved June 15, 2008, from http://nces.ed.gov/programs/digest/d06/tables/dt06_132.asp

National Commission on Excellence in Education (NCEE). (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: U.S. Government Printing Office. Retrived June 15, 2008, from http://www.ed.gov/pubs/NatAtRisk/.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Washington, DC: U.S. Government Printing Office.

O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. *American Educational Research Journal, 44(1)*, 161-196.

Perfetti, C. A., Wlotko, E. W., & Hart, L. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(6)*, 1281-1292.

Price, G. G., & Carpenter, T. P. (1978). Review: On further examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. *Journal for Research in Mathematics Education, 9(2)*, 155-160.

Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology, 86*, 54-64.

Rothstein J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics, 121(1-2),* pp. 297-317.

Seastrom, M., Hoffman, L., Chapman, C., & Stillwell, R. (2005). *The averaged freshman graduation rate for public high schools from the common Core of Data: School years 2001-02 and 2002-03 (NCES 2006-601).* Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Sereno, J., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory and Cognition, 25*, 425-437.

Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal, 41(1)*, 159-189.

Share, D., & Shalev, C. (2004). Self-teaching in normal and disabled readers. *Reading and Writing: An Interdisciplinary Journal, 17*, 769-800.

Shefelbine, J. (1990). Student factors related to variability in learning word meanings from context. *Journal of Reading Behavior, 22,* 71-97.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119(1),* 3-22.

Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives.* Oxford, UK: Oxford University Press.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*, 72-110.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representations for morphologically related words. *Journal of Verbal Learning and Verbal Behavior, 18,* 399-412.

Stanovich, K. E., & Cunningham, A. E. (1992). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of Educational Psychology, 85(2)*, 211-229.

Stanovich, K. E., & Cunningham, A. E. (1993). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition, 20(1),* 51-68.

Stedman, L. (1993). The condition of education: Why school reformers are on the right track. *Phi Delta Kappan, 75(3)*, 215-225.

Stedman, L. C. (1996). Respecting the evidence: The achievement crisis remains real. *Education Policy Analysis Archives, 4(7)*, 1-36. Retrieved July 31, 2008, from http://epaa.asu.edu/epaa/v4n7.

Sternberg, R., & Powell, J. S. (1983). Comprehending verbal comprehension. *American Psychologist, 38,* 878-893.

Sum, A., Kirsch, I., & Taggart, R. (2002). *The twin challenges of mediocrity and inequality: Literacy in the U.S. from an international perspective*. Princeton, NJ: Educational Testing Service.

Turnbull, W. W. (1985). *Student change, program change: Why the SAT scores kept falling (ETS RR No. 85-28)*. New York: College Entrance Examination Board.

West R. F. & Stanovich, K. E. (1991). The incidental acquisition of information from reading. *Psychological Science, 2*, 325-330.

Wolfe, M. B. W., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintch, W., & Landauer, T. K. (1998). *Discourse Processes, 25(2 & 3)*, 309-336.

Zipf. G. K. (1935). *The psycho-biology of languages*. Boston: Houghton-Mifflin.

Figure 8.1. Average SAT scores in 1962-63 and from 1966-67 to 2006-07. Data from Turnbull (1985) and National Center for Education Statistics (2007, Table 132).

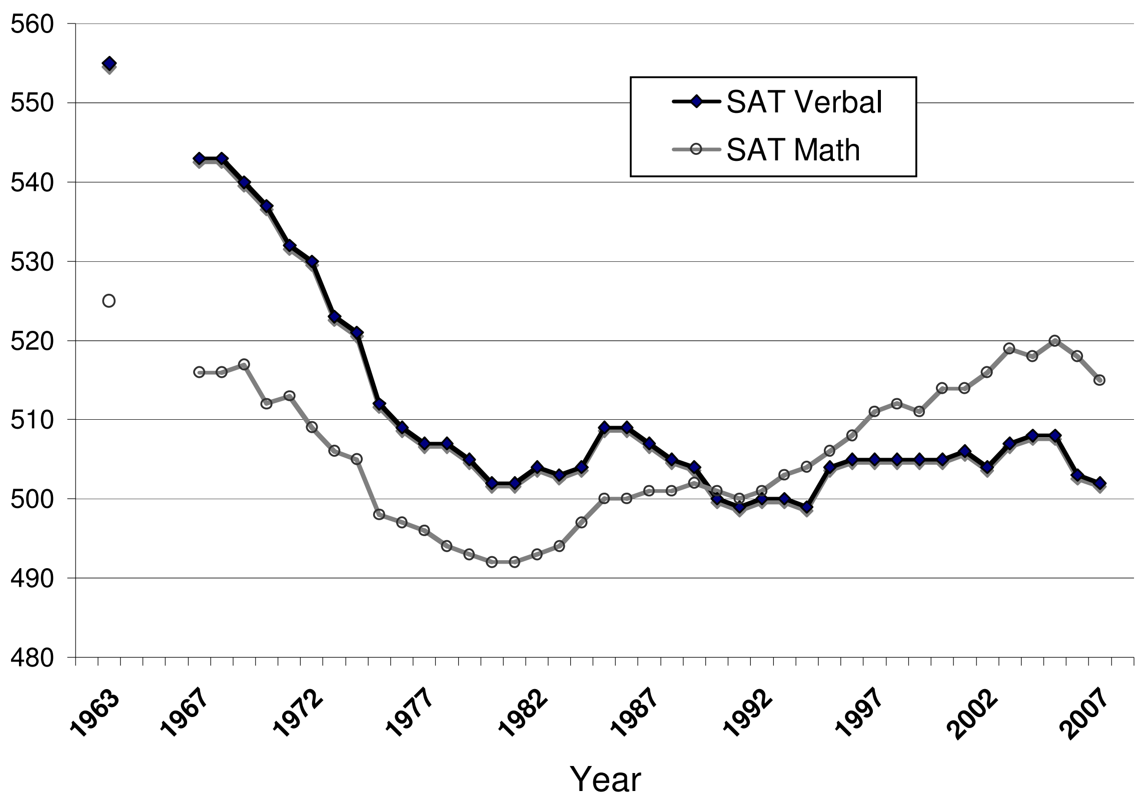

Figure 8.2. Frequency counts from the British National Corpus (2001) for adjectives, adverbs, verbs, and common nouns. For common nouns and verbs, counts are for base (uninflected) forms.