

# Technology for Developing Children's Language and Literacy:

Bringing Speech-Recognition to the Classroom

Prepared for  
the Joan Ganz Cooney Center

by

Marilyn Jager Adams  
Brown University  
Providence, Rhode Island

2011

**Technology for Developing Children's Language and Literacy:  
Bringing Speech-Recognition to the Classroom**

*Technology for Developing Children's Language and Literacy* by Marilyn Jager Adams and the Joan Ganz Cooney Center is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Suggested citation: Adams, M.J. (2011). *Technology for Developing Children's Language and Literacy: Bringing Speech Recognition to the Classroom*, New York, NY: The Joan Ganz Cooney Center at Sesame Workshop.

Marilyn Jager Adams  
139 Horseneck Road  
Westport, MA 02790  
phone: 508-636-5352  
email: [marilyn.adams@verizon.net](mailto:marilyn.adams@verizon.net)

The Joan Ganz Cooney Center at Sesame Workshop



# CONTENTS

1. Introduction.....	3
1.1. The Educational Challenge.....	3
1.2. The Reading Challenge.....	5
1.3. Overview of this Report.....	5
2. Basic Reading Fluency.....	7
2.1. Developing Basic Reading Fluency .....	7
2.1.1. Getting Children to Read in the Conventional Classroom .....	8
2.1.2. What’s so Good about Guided Oral Reading?.....	10
2.2. Summary .....	11
3. Technology for Promoting Basic Reading Fluency: Automatic Speech-Recognition Technologies .	13
3.1. Is Automatic Speech Recognition Up to the Challenge?.....	13
3.2. Detecting Word Recognition Difficulties.....	14
3.2.1. The Problem of Pronunciation Variability .....	14
3.2.2. The Nature and Frequency of Reading Dysfluencies .....	16
4. Using Dysfluencies to Monitor Reading Difficulty.....	19
4.1. Reconfiguring the Speech Recognizer to “Hear” Pauses.....	19
4.2. Distinguishing Symptomatic from Acceptable Pauses.....	20
4.3. Implementing the Pause Factor in a Speech Recognition-Based Reading Tutor.....	21
4.3.1. How Well Can Pauses Diagnose Difficulties: Clinicians vs. Computer.....	21
4.3.2. Testing the System’s Tutoring Performance.....	22
4.4. Designing for Impact and Usability.....	23
5. Beyond Basic Fluency Support .....	25
5.1. Prosody and Understanding.....	25
5.2. Automatic Speech-Recognition for Managing Interactive Learning.....	30
5.3. Speech-Recognition and Reading Assessment.....	31
6. Education, Literacy, and Technology: Concluding Comments .....	35
6.1. On the Importance of the Problem.....	35
6.2. On the Nature and Importance of the Solution.....	37
7. References .....	39



# 1 Introduction

## 1.1. The Educational Challenge

Throughout history, educational aspirations and opportunities have been driven by advances in technology. Reciprocally, at least for those countries that have prospered most with each technological surge, advances in technology have quantitatively and qualitatively increased educational access or delivery in response. With the invention of writing and then of the alphabet, with the invention of the printing press and the Renaissance it fueled, with the scientific and philosophical tools of the Enlightenment, with the productivity, infrastructure, and innovations of the Industrial Revolution—with each great technological shift, educational attainment in these countries has risen alongside.

In our own age, advances in computing technologies have hugely multiplied the demand for education and, especially, for literacy and for the knowledge, skills, and modes of thought that literacy enables. Advances in computing technologies have also hugely multiplied access to information though, again, principally for those who are educated or, minimally, who can read.

So far, however, modern technologies have done little to impact educational attainment or efficacy in the United States, where that includes reading acquisition. We know this because there is one sector of the educational system that has been significantly expanded through technology: Large-scale testing.

As technology has made large-scale testing ever easier and more frequent, the results have remained essentially the same. On international assessments, the performance of U.S. students has been mediocre, at best. On national assessments, the performance of the majority of our students is well below grade-level in history, civics, science and math. And no wonder, for the reading ability of the majority of our students is also well below grade level throughout the school years (See Section 6).

For decades now, the data from such assessments have been met with a flow of proclamations about who or what is to blame for the ever-deteriorating state of

our educational system. Writing in 1987, however, Stedman and Kaestle cautioned that the United States' big educational issue was not, in fact, that its schools were getting worse. Instead, they argued, when consideration is restricted to students of the same age, socioeconomic status, and amount of education, any recent declines in literacy levels are so small as to pale in comparison to the enormous rise in literacy that occurred over the course of the twentieth century. Even so, pointing out that the growth in U.S. literacy levels across the twentieth century was due principally to increases in educational attainment as schooling was opened to ever larger segments of the population, Stedman and Kaestle concluded:

Even if schools are doing about as well as they have in the past, they have never done well in educating minorities and the poor or in teaching higher-order skills. If increased education is the only reason the population has kept up with increasing literacy demands of our society, we have plenty to worry about. (1987, p. 42).

At the outset of the twentieth century, the United States had a strong international advantage in education due to what was broadly seen in Europe as its “wasteful” practice of offering free (government subsidized) education to commoners (Goldin & Katz, 2008). In the nineteenth century, public grammar schools were already widespread in the United States, and during the first half of the twentieth century, public education was rapidly expanded to high schools. In result, average educational attainment in the United States increased by 6.7 years over the course of the century.

Alongside, economic productivity in the United States grew about 2.5 percent per year over the twentieth century. Goldin and Katz (2008) estimate that increases in educational attainment alone directly boosted the United States' economic growth by an average of nearly 0.5 percent per year between 1915 and 2005, adding that the indirect contributions of educational growth to the economy were likely even greater.

Yet, beginning in the 1970s, increases in educational

attainment within the United States slowed such that high school and college completion rates have barely changed since. Between 1980 and 2000, the fraction of U.S. productivity growth attributable to education fell by a third and, absent some dramatic change in the situation, is estimated to fall by half again in the next two decades (Goldin & Katz, 2008).

Meanwhile, over the last third of the twentieth century, as educational attainment began to level out in the United States, it began to pick up in many other countries. In the 1950s, high-school enrollment in most European nations was less than 40 percent. Today, educational attainment of young adults in most European countries and several Asian countries equals or exceeds that of young adults in the U.S. (Provasnik, Gonzales, & Miler, 2009). Until recently, public high schools were rare in less developed countries. Today, they are growing rapidly in quantity and quality. Given the surge in educational growth in the rest of the world, the slowing of educational growth within the United States effectively amounts to an even steeper relative decline.

The economic ramifications of the United States' educational slowdown have been especially visible, for just as its educational growth began to plateau, growth in the technology sector began to take off. Just as the flow of new college graduates was abating, the demand for them was accelerating. Beginning in the late 1970s, there resulted a "wage premium" for college-educated workers (Autor, Katz, & Krueger, 1998; Machin & van Reenan, 1998). Yet, this trend was relatively short-lived.

By the mid-1990s, the U.S. labor market began to polarize. The market strengthened for those with advanced degrees. Alongside, the number (though not the pay) of jobs requiring minimal education increased, too. However, the demand for college-educated workers fell.

Analyzing this "convexation" of the U.S. labor market, Autor, Katz, and Kearney (2008) point, again, to technology as a cause. As technology grew in distribution and sophistication, as it increasingly became adapted for and adopted by different sectors of the economy, it increasingly absorbed the load of work that could be "routinized" (Autor, Levy, & Murnane, 2003). Done by people, this is the sort of work that requires attention and intelligence; it is the sort of work that once occupied a large segment of America's better-educated workforce.

At the same time, there arose an acute demand for workers with the knowledge and skills required to develop and deploy the new technological capabilities. Such knowledge and skills include not only those essential for good management, but down the ladder, many more. Among these are the abilities to understand, exploit, and maintain currency with new technologies and with their possibilities and limitations. Beyond that, however, they include the capacity to understand assignments and projects in terms of the goals that drive them as well as the constraints of time and resources that govern them. They include the inclination and capacity to detect and diagnose problems in the conception and execution of tasks and to consider the costs, benefits, and concomitants of alternative approaches or compromises. They include the capacity to understand the needs and perspectives of others, including customers and the public—present and future—as well as co-workers. They also include the capacity to write clearly, incisively, and as appropriately to the audience and task as well as the ability to speak clearly and to the point. As David Olson (e.g., 1994, 2009) has so forcefully argued, all of these sorts of knowledge, capabilities and modes of thought are integrally dependent on literacy, both culturally and individually.

The technology sector will continue to grow. How will it fulfill its need for highly skilled workers?

One broadly advocated option is for the United States to develop aggressive programs toward increasing the rate of college graduation beyond its resting level of 30 percent. Yet, neither would this seem a promising solution, in itself. On one hand, assessment data show that barely 30 percent of U.S. twelfth-graders are mastering high-school expectations. On the other, the data equally make clear that even among young Americans who have already completed college, too many have not learned what is needed. So, yes, we would like more students to pursue higher education, but only to the extent that they can benefit from higher education. Students cannot learn that for which they are unprepared. Relaxing entry qualifications and curricular demands so as to increase college completion statistics is a strategy that, in itself, can serve only to undermine the value of a college education, both on paper and in its real impact.

Another option, already hot, is for U.S. companies to look offshore for the skills they want and need. On the downside, of course, this is a strategy that

undermines our country's workforce, both individually and collectively, even as it must inexorably undermine our country's capacity to do anything to reverse the trend.

In any case, the day is past when the U.S. might ensure its position in the global economy through numbers alone. Relatively speaking, after all, the population the United States is less than 5 percent of the world population and shrinking.

The only real alternative for our country, and it is a pressing alternative, is to increase the quality and impact of our educational system. Core to that goal is taking seriously the challenge of ensuring that far more children develop the literacy levels on which education depends, in school and beyond.

## 1.2. The Reading Challenge

According to the National Assessment of Educational Progress (NAEP), as many as forty percent of fourth-graders in our nation's schools are unable to read with minimal fluency (Daane, Campbell, Grigg, Goodman, & Oranje, 2005). Further, the lower the students' fluency, the lower too is their reading comprehension. Average reading comprehension of those fourth-graders who were not yet able to read NAEP's test passage with minimal fluency fell below the "Basic" cut-off, indicating an inability to understand or learn from grade-level texts. At the other end of the distribution, only 10 percent of fourth-graders demonstrated an ability to read the passage "with phrasing that was consistent with the author's syntax and with some degree of expressiveness" (Daane et al., 2005, p. v), and only this group obtained reading comprehension scores that were at or above grade-level ("Proficient" on the NAEP).<sup>1</sup>

As recognized by the NAEP committee, there are two levels to reading fluency. The first depends on the ability to read the separate words of a text with sufficient ease and accuracy. The second depends on the ability to grasp or reconstruct the structure and force of the author's collective words.

To most, it is obvious that learning to recognize printed words involves skills and practices that are specific to the written domain. Yet, research shows this to be equally true of the vocabulary, syntax,

background knowledge, and modes of thought that characterize text. Moreover, because the knowledge and skills required for reading and understanding text are specific to text, their acquisition comes about only through experience in reading and understanding text.

But here is the Catch-22: A text read without fluency can barely be understood, and what has not been understood cannot be learned. It follows that unless and until children can read and understand texts on their own, they need support and instruction to help them through it. The value of providing such help is not merely one of ensuring that students will gain from the text at hand but, more importantly, that they will be better able to manage the next text on their own—after all, schoolbooks only become harder with time.

## 1.3. Overview of This Report

The thesis of this paper is that it is for lack of such help that reading fluency eludes so many of our school children. In support of this argument, Section 2 is focused on the nature of the support needed for gaining basic fluency and why its adequate provision is effectively impossible within the conventional classroom. The next three sections are devoted to the argument that speech recognition-based reading systems offer a technically viable escape from this dilemma. In Section 3, acknowledging skepticism about the feasibility of speech-recognition based reading systems for children, discussion is given to the most loudly voiced misgivings and to how each has been solved by researchers. Section 4 then overviews the special challenges—and demonstrated success—of such systems for promoting basic reading fluency in the classroom. In Section 5, focus is turned to the challenge of supporting more advanced dimensions of reading fluency and comprehension. Here again the argument is made that speech recognition technologies offer key capabilities for monitoring student performance and for providing both corrective and enriched learning opportunities that are beyond the human capacities of the conventional classroom. Section 5 closes with a discussion of the potential value of such technology for assessing children's reading growth. The "Concluding Remarks" in Section 6 begin with a reminder of the importance of improving the educational efficacy of our schools, and then recapitulate the costs and benefits of providing to our schools the resources they require to offer real literacy support to every student, where that centrally includes the technology for helping all to learn to read.

<sup>1</sup> These results essentially replicated the findings of NAEP's 1995 fluency study (Pinnell et al., 1995).





## 2 Basic Reading Fluency

Reading fluency, as the name suggests, is the ability to read aloud with the kind of ease, accuracy, rhythm, and intonation that signals ongoing command of the meaning and flow of the text. Reading with fluency depends, first, on the ability to recognize the separate words of a text quickly and accurately and, second, on having and applying the linguistic and background knowledge needed to interpret the words both separately and in their ensemble.

For beginning readers, sounding out or *decoding* even the simplest words can be an arduous task. Yet, the very process of decoding a word leaves an interconnected trace in memory of its spelling and pronunciation. As these traces strengthen and overlap, they gradually enable the reader to recognize (as distinct from figuring out) whole words and spelling patterns and to map them instantly to pronunciations. Thus, mature readers generally are able to read never-before-seen words with remarkably little effort—so little, in fact, that if the word is in the reader's speaking or listening vocabulary, she or he may never even notice that it has never before been seen.

The capacity to read novel words easily is called *decoding automaticity*. Decoding automaticity is owed to the fact that the reader's cumulative knowledge of spelling patterns and their mappings to speech correspondences provides a support structure by which nearly every word is partly recognized even on its first visual encounter. Although the development of decoding automaticity is significantly hastened by good phonics instruction, the orthography of written English is complex and quirky and its vocabulary is vast. Further, because it is the meanings of words that serve to best differentiate them from one another, the growth of children's sight word vocabulary depends heavily on experience and practice in reading meaningful text.

It is through mapping the spellings of words to their pronunciations that print becomes bound to the language centers of the brain. Provided that a word is read and understood in context, the word's activation will extend through its pronunciation to its meanings

and usage. Yet this, as it turns out, is a heavy proviso for young readers. First, active attention can be directed to only one mental activity at a time. Where active attention is required for the task of decoding a word, it is necessarily removed from the task of understanding the text. Second, on every dimension, be it vocabulary, syntax, background knowledge, or modes of thought, the comprehension requirements of written language are more demanding, more complex, and in many ways qualitatively different from those that characterize oral language situations (Biber, 2009; Chafe, 1985; Hayes, 1992; Hayes & Ahrens, 1988; Hayes, Wolfer, & Wolfe, 1996). Because the comprehension requirements of text are specific to text, their acquisition depends on reading and, more, on reading and understanding.

Gaining a basic level of reading fluency is a core challenge in the primary grades. At the outset of grade two, most children's reading is not yet fluent. When they read aloud, their voicing is often hesitant and choppy, and their comfort zone is generally limited to short texts built from short sentences, words that are easy and familiar, and ample repetition. By the end of second grade, on-pace students will have developed a basic level of reading fluency (Adams, 1990; Chall, 1983). Only now are they ready for longer stories, comic books, chapter books, and also for learning how to learn through reading, as will be required in the grades to come. To be sure, even the best readers' fluency is still far from mature at this point. They will continue to stumble when they encounter unfamiliar words, language, concepts, and modes of thought. However, the more they read—or, more accurately, the more they learn through reading—the less this will happen.

But again, as shown by the National Assessment of Educational Progress (NAEP), as many as forty percent of students in our nation's schools are unable to read with minimal fluency by the time they are in fourth grade (Daane et al., 2005).

## 2.1 Developing Basic Reading Fluency

According to the report of the National Reading Panel (2000), the single best practice for developing students' reading fluency is one-on-one *guided oral reading*. Guided oral reading occurs when a child reads and rereads texts aloud alongside a helpful adult. As documented in the Panel's report, guided oral reading sessions effectively accelerate students' growth in reading fluency as well as their word recognition, comprehension, and full-scale reading scores. In view of such findings, the National Reading Panel strongly urged all teachers to make more time for guided oral reading.

Such urging is well and good, but, realistically, schools are not able to offer much one-to-one reading support due to the very structure of the classroom. For a teacher to read just 10 minutes a day with each of 25 children in a classroom would require fully two-thirds of the school day—assuming the impossibilities of zero transition time, no interruptions, and that each of the 24 unattended children in the classroom behaves perfectly for the duration.

Consistent with the arithmetic of the situation, field research affirms that one-to-one reading opportunities are generally both rare and brief during school hours. This is so whether in the general education classroom (Durkin, 1978-79; McIntosh, Vaughn, Schumm, Haager, & Lee, 1993; Moody, Vaughn, & Schumm, 1997) or in special education settings (Vaughn, Levy, Coleman, & Bos, 2002; Vaughn, Moody, & Schumm, 1998). Further, and for the very reason that it is so urgently and intensively needed, the amount of individual guided reading is even less in poorer schools and with poorer readers: To give poor readers anything close to the amount of individual support that they need requires more time and attention than teachers and aides, together, have available (Allington, 1983; Allington & McGill-Franzen, 1989; Birman et al., 1987).

### 2.1.1. Getting Children to Read in the Conventional Classroom

Teachers are deeply aware that helping children learn to read well depends integrally on getting them to read enough. Nevertheless, classroom practices for engaging students in reading necessarily rely on strategies that reduce to manageable levels the per-child amount of teacher time and support required.

Many classrooms set time aside for students to read

by themselves. Classroom time for such *independent reading* comes under many different labels—e.g., SSR (Sustained Silent Reading), DEAR (Drop Everything and Read), SQUIRT (Sustained Quiet Uninterrupted Reading Time), and IRT (Independent Reading Time). The sheer number of different labels reflects the number of times that independent reading has been reinvented and re-energized as a hopeful means of solving the read-enough impasse.

But alas, at least for students who yet lack basic fluency, and despite many efforts to show otherwise, programs to increase independent, silent reading have generally failed to boost learning outcomes (Hairrell, Edmonds, Vaughn & Simmons 2010; National Reading Panel, 2000; Wilkinson, Wardrop, Anderson, 1988). Why does silent, independent reading fail to advance reading growth among these students? One possibility is that children tend not to apply themselves sufficiently when reading on their own. Another is that silent reading does not afford children the help needed to address and conquer difficulties. A third, consistent with the genesis of silent reading both historically (Manguel, 1996) and developmentally (Chall, 1983) and also with analyses of the neural substrates involved in reading (Dehaene, 2009), is that silent reading is enabled only through the rich establishment and near automatization of the pathways and processes involved in oral reading.

None of this is to dismiss the importance of silent or independent reading. After all, the core goal of teaching students to read in the first place is toward ensuring that each of them eventually can and will read silently and independently. The problem is that advancing students' reading ability to a level where independent reading becomes productive depends on additional support. Actually, the problem is worse than that: Although providing such support is arguably the very most important challenge of public schooling, it has also been its least attainable.

To concretize the difficulty of this challenge, it is worth reviewing the principal options available to classroom teachers.

*Round-robin reading* is perhaps the most common approach to offering oral reading support and instruction in the classroom (Ash, Kuhn, Walpole, 2009). In round-robin reading sessions, the children take turns reading or rereading successive fragments of the text while the others follow along in their own books. Efficient as this process may sound, it turns out that children tend to pay little attention to the text or the teacher when it is not their own turn to read

aloud or speak (Anderson, Mason, & Shirey, 1984; Anderson, Wilkinson, & Mason, 1991). In result, the amount of practice and instruction afforded any given child effectively reduces to whatever time she or he actively participates, which is necessarily precious little.

In *choral reading* sessions, by contrast, the teacher leads all of the students with whom she is working to read aloud in unison, stopping them intermittently to ask questions or prompt discussion. If the text is challenging, then the teacher will often read each segment aloud before asking students to read it back to her. In principle, choral reading sessions offer more oral reading opportunity to every child than do round-robin sessions but, in fact, just as with round-robin reading, children who choose not to participate can easily lie low. A watchful stroll around the back of someone else's classroom during choral reading inevitably turns up a number of children who are not really reading at all. Many are just moving their mouths and noncommittally vocalizing to the rhythm. Typically, several of the kids don't even have their books open to the right page. Very likely, of course, among the students who are off-text or off-task are those who are in greatest need of practice and guidance.

In *partner-reading* sessions, students sit in pairs and take turns reading to each other: One child reads aloud while the other assists, switching roles at designated junctures. The partner-reading dynamic accords each child more time to read aloud than in round-robin reading sessions and more responsibility to do so than in choral reading sessions. Given the mutual support and interaction that partner reading affords, children may also be expected to learn more and make more of the story than when reading alone. At least where partner-reading sessions are given proper training, supervision, and scheduling, research shows that they can effect significant gains in reading progress (Simmons, Fuchs, Fuchs, Mathes, & Hodge, 1995). Nevertheless, meta-analyses indicate that partner-reading sessions typically accelerate reading growth only where children are "partnered" with a significantly better reader and preferably an adult (Chard, Vaughn, & Tyler, 2002; Thierren, 2004), bringing the problem full circle again.

Given the sensitivity of reading growth to the quality of guidance provided, many teachers divide students into *ability groups* for reading instruction. For students who need special help with beginning reading skills, such as phonemic awareness and basic decoding, the value of such ability grouping is generally strong and

often dramatic (Cavanaugh, Kim, Wanzek, & Vaughn, 2004; Juel, 1990; Wanzek & Vaughn, 2007). On the other hand, research on ability-grouping with readers beyond the entry level is less clear: Aggregated across such studies, the impact of ability-grouping on reading growth is slightly positive but very small and highly variable (Kerckhoff, 1986, Lou et al., 1996).

Again, a major motivation for ability-grouping is to adjust the pace, focus, and materials of instruction as appropriate to students' needs. It would seem, then, that ability grouping should at least be helpful for weaker readers. Yet just the opposite seems true. Relative to whole-class instruction, ability-grouping tends to accelerate gains for children in the high-ability groups but to depress reading growth among those in the low-ability groups (Lou et al., 1996; Kerkhoff, 1986). Consistent with these findings but still more troubling, before-after reading studies repeatedly indicate that the best predictor of an individual student's reading growth is not that particular child's ability at study outset, but the average ability of the group in which she or he is taught (Anderson, Wilkinson, & Mason, 1991; Barr & Dreeben, 1983; Juel, 1990). The implication is that different strategies for within-class grouping—whether whole class, ability-groups, small heterogeneous groups, or partners—can alter the distribution of reading growth across students but can do little to affect the class average.

To be sure, evaluating the impact of ability-grouping as though it were a single procedure is crude, for there are many different approaches, structures, and realizations of how such sessions might be conducted, some very good, some not (Wasik & Slavin, 1993). Nevertheless, if the focus is on getting students to read more, not even the best small-group dynamic, with or without ability-grouping, offers a solution in and of itself. Although the teacher may feel like she can help and learn more about her students when she gathers them close to her in smaller groups, time is time: So long as there is just one teacher, dividing the class into groups cannot increase the average amount of attention she can lend to each child.

Although there are a myriad of other strategies for engaging children in reading in the classroom, most amount to variations on these basic alternatives, some more efficient, some less. No matter, in their impact on reading acquisition, all such teacher-diluting strategies pale in comparison to *individual guided oral reading* sessions. Research has demonstrated across many decades and tutoring methods that, given sufficient one-on-one support, virtually all healthy children

are able to make normal or accelerated progress in reading. Moreover, the rate of progress in such tutorial interventions is shown to be most strongly related to the amount of supported reading that takes place (Monroe, 1932; Pinnell et al., 1995; Slavin, Karweit, & Wasik, 1994; Slavin, Karweit, & Madden, 1989; Vellutino et al., 1996; Wasik & Slavin, 1993). Indeed, given the basics, evidence suggests that just a few minutes a week of such one-to-one reading is enough to make a substantial difference in a child's reading growth (Adams, 2006).

### 2.1.2. *What's So Good About Guided Oral Reading?*

Why is guided oral reading so helpful? The basic answer is that when text is challenging, reading is hard. It requires concentration, perseverance, and is rewarding only to the extent that it is successful. Until children can read comfortably on their own, they need someone or something to guide them across the difficulties, to support their on-going understanding, to help them to appreciate their own progress and accomplishment and, not least, just to keep them engaged and on task.

Among the specific benefits of guided oral reading is that the adult leads the child to lend due attention to visually new words. Indeed, most helpful in this vein may be the adult's very expectation that the child will earnestly try to read such words. Research with on-page second- and third-graders shows that only a single instance of correctly identifying a new word while reading is sufficient to create a trace of that word in memory that is remarkably complete and enduring. Just a few such encounters with a new word, and it becomes a *sight word*, instantly available forever (Cunningham, 2006; Cunningham, Perry, Stanovich, & Share, 2002; Share, 2004).

What about reading delayed or disabled students? Perhaps, hypothesized Share and Shalev (2004), these children require more encounters with a word to "own" it. Yet, on studying this hypothesis, Share and Shalev found that when such children did indeed identify the new words in their texts correctly, the words were invariably retained, just as they are for better-reading peers. Why then, asked Share and Shalev, is it that poorer readers characteristically acquire so many fewer new words through reading than their better-reading peers? The answer, they found, is that poor readers are less likely to identify new words correctly during reading, tending instead to give up or skip the words all together. Thus, guided oral reading offers a remedy, again, for sightword acquisition is hastened

when students, including struggling readers, are given immediate correction and required to try again (e.g., Ehri & Saltmarsh, 1995; Manis, 1985; Reitsma, 1983, 1989).

At the same time it should be borne in mind that with or without the help of a tutor, readers can focus attention on any individual word only by diverting attention from the text's larger meaning. This, in turn, raises two reasons why attention to word-reading accuracy must be managed with care. First, children are significantly more likely to retain useful knowledge of a new word where they understand the context in which it appears (Cunningham et al., 2002; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Nation & Snowling, 1998). Second, conquering new words is hardly the only reason for asking kids to read. Good tutors know this well. Where word struggles disrupt ongoing interpretation of a text, good tutors will reread and/or direct the child to reread the sentence, a procedure that serves to reinforce the recognition, meaning, and use of the troublesome word even as it recovers the meaning and thread of the text.

Nor can the importance of vocabulary and language assistance be overstated. Written text—where that includes text that has been written specifically for preschool and primary grade children—presents orders of magnitude more words than children can have learned through oral language (Adams, 2009, 2011; Hayes & Ahrens, 1988; Hayes et al., 1996). Furthermore, the meaning and message of written text generally depend on more complete and precise understanding of word meanings and usages than do oral language situations (Biber, 2009; Gardner, 2004).

Teachers are often taught that students should be expected to infer the meanings of new words while reading. In fact, this is not particularly good advice. Where the meaning of a word is essentially if crudely known to the reader, encountering it again and again is invaluable toward gaining control of its full meaning and usage (e.g., Landauer & Dumais 1997; Nagy, Anderson, & Herman 1987; Swanborn & de Glopper 1999). On the other hand, where a word is wholly unfamiliar, the likelihood that context will inform its meaning is slim (Nagy et al., 1987; Schatz & Baldwin, 1986). Summarizing his own extensive work in this area, Richard Anderson (1996) reports, "The overall likelihood [of learning the meaning of new words through reading] ranged from better than 1 in 10 when children were reading easy narratives to near zero when they were reading difficult expositions" (p. 61).



Once again, guided oral reading offers a good solution. Supplying the meaning of an unknown word in the course of reading, at the very moment when needed, is shown to be a singularly helpful means of anchoring it in the student's memory (e.g., Jenkins, Matlock, & Slocum, 1989; see also National Reading Panel, 2000).

Guided oral reading sessions also offer ideal opportunities for supporting children's larger understanding and appreciation of what they read. Always knowing exactly where the child is in the text, the listening adult can choose just the right moments to probe or support understanding. If the child has trouble parsing a sentence, the adult may expressively reread it and invite the child to paraphrase it and reread it her- or himself. If the child seems troubled by the content, the adult will provide background knowledge. When key events or information arise in a text, the adult may ask the child to pause to make predictions or otherwise take time to consider the possible significance of what has just been read. Perhaps the adult will ask the child to pause to lend explicit attention to the attitudes or dilemmas of the characters in a text. Perhaps the adult will invite the child to consider the impact or aesthetics of how the author has chosen to word or structure the text. Perhaps the adult will lead the child to lay bare the author's perspective and point or to examine the consistency of the author's larger argument. For a reading tutor, the reward of the job lies in making each reading as interesting and educative as possible.

In complement to such instructional opportunities, reading tutors and clinicians are also very good at managing and monitoring students' reading progress. They are good at judging the interests, needs, and capacities of their students, and they are good at judging the demands and difficulty of different texts (Klare, 1984). They are good at deciding what texts a child might profitably read, and they are good at deciding what texts or sections of text warrant rereading. They are good at making note of the orthographic, vocabulary, and learning challenges on which a student needs extra work, and they are good at selecting or devising activities in response. They are good at administering timed reading assessments so as to gauge growth in fluency. And they are good at communicating with their tutees both about their needs and, no less, about their progress and accomplishments.

## 2.2. Summary

But here lies the core dilemma. Such opportunities are effectively unavailable in schools, and relatively few children have the luxury of a private tutor. Nevertheless, if learning to read depends so heavily on this sort of help, then it must somehow be available to most children who do learn to read. Where might successful students find it?

The answer, of course, is in the home (e.g., Bradley, Corwyn, McAdoo, & Coll, 2001; Burkham, Ready, Lee, & LoGerfo, 2004; Sénéchal & LeFevre, 2002). In many homes, parents and caretakers regularly read to and with their children. Across as many sessions, they share many different books and reading selections. They provide instruction and engagement with print, with language, and with information both in the texts and behind them. They provide models of good reading and of different ways of responding to what is read. In all, if more leisurely and intimate than the typical tutorial, the sort of shared reading that often takes place between parents and children essentially amounts to a broad and continual program of guided oral reading, extended in time and dimensionality, and ever sensitive to the individual child's knowledge, interests, needs, and readiness to learn. Keep in mind, however, that researchers are able to measure the impact of such support only because there are so many homes in which it is lacking.

Because home educational and literacy practices correlate strongly with parental income and education, they present themselves as the single largest causal factor underlying the have-/have not achievement gap. Data from the Early Childhood Longitudinal Study show that the reading ability of children from disadvantaged homes is already six months behind that of their more advantaged peers on kindergarten entry; by third grade the difference has widened to 16 months (Aikens & Barbarin, 2008; see also Loeferfo, Nichols, & Reardon, 2006). To be sure, there are children who perform miserably in school despite the fact that their homes offer all manner of educational stimulation and support (Bhattacharya, 2010). Nevertheless, longitudinal study of children of poverty affirm that absent home support, the prospects of learning to read are slim (Entwisle & Alexander, 1996; Entwisle, Alexander, & Olson, 2005).

Textbooks, homework, tests, libraries, laws, newspapers, the World Wide Web—all remain inaccessible except as children learn to read. Thus it is that by teaching children to read, public education is supposed to afford full educational access to all children, regardless of

background. Yet, if the likelihood that children will indeed learn to read depends less on school than on the reading support and instruction they receive at home, then this is a false promise. Somehow, our schools must find a way to offer every student at least the minimal support that is required for learning to read.

# 3 Technology for Promoting Basic Reading Fluency: Automatic Speech-Recognition Technologies

Learning to read depends on application, discipline, and understanding that are barely attainable without individual guidance and support. Providing individual children the amount and kinds of support that are required is impossible within the structure or economics of the conventional classroom. Therefore, to the extent that children depend on public school for their formal education, they must be expected to fare relatively poorly unless and until public schools can provide the kind of educational support on which its educational demands depend.

Could schools escape from this dilemma by hiring additional reading tutors? The answer is no, and the arithmetic of the number of (good) people as well as the amount of money that would be needed to hire them is left to the reader.

The purpose of this section is to argue that there is a viable, cost-effective solution to this impasse. Automatic speech recognition can be harnessed so as to give the computer “ears.” Speech-recognition based reading tutors offer a cost-effective and scalable means of ensuring ample, individual reading and learning support to every student.

## 3.1 Is Automatic Speech Recognition Up to the Challenge?

Today, people around the world, in dozens of languages, depend on automatic speech recognition for telephone call-routing and directory assistance. It is widely used for dictation and information capture in the defense, health care, and legal sectors. It is used for captioning live television so we can watch our favorite games in noisy sports bars and by unnamed agencies for transcribing suspicious communications. It is used by people to talk to their computers and mobile devices, for example, while browsing the Web, creating their own voice commands, and managing their bookmarks. People use it to issue commands to their cell phones and, in reverse, for asking their cell phones to transcribe their voicemail and send written copy to

their e-mail. People use automatic speech recognition to talk to their televisions, their music players, their cars, and their navigation systems. And, of course, speech recognition is very hot in the gaming industry.

In other words, automatic speech recognition is a technology that is mature and even commonplace in industry after industry with the salient exception of where it is needed most: in education. Whatever the economic or social value of the applications mentioned above, most pale in comparison to the potential of speech recognition as it could and should be used to help people learn to read and read to learn.

So why is it that speech recognition has yet to penetrate the schoolhouse?

Somehow, the “word on the street” is that speech recognition technology is not feasible for young children. Skeptics claim that young children’s syntax is too immature and unpredictable. Their classrooms are too noisy. Their voices are too high. Their elocution is too idiosyncratic and fraught with regional and other-language accents. Let’s take each of these in turn.

First, yes, the spontaneous language of young children is marked by unconventional wording and syntax. If the goal were one of recognizing their spontaneous speech, this would indeed be a complication. However, given instead the goal of monitoring and supporting their reading, it is not. When reading, the syntax and wording of what the children are to say is prescribed by the print on the page. The speech recognizer’s challenges are thus reduced to deciding whether the child reads what is written and to discerning whether and where the child experiences difficulty in so doing.

Turning to the second issue, yes, classrooms are very noisy places. This, too, is a nonstarter. Speech recognition systems have long depended on distinguishing speech sounds from non-speech sounds and on compensating or correcting for background noise. Similarly, one can program the system to excise snuffles, coughs, sneezes, and, with special deference to the grade 5-and-up crowd, burps from the speech stream. Of course, some of the noise in a classroom

is from the vocalizations of others, but today, directional, speech-quality microphones are wholly affordable and, thanks to the gaming industry, are available with child-sized headsets. Besides which, like a human tutor, the software system is always allowed to tell the student that it couldn't hear and to request that the student reread troublesome segments as needed. The system may also be programmed to alert the teacher of persistent audio trouble as may arise from, for example, bad sound cards, flakey USB connections, or broken headsets.

With respect to the third issue, yes, it is true that the original work on automatic speech recognition was focused on grown men or, more specifically, adult English-speaking men and, even today, few speech recognition systems work well with young children. Crudely, speech recognition programs work by imagining the vocal tract as a long pipe where the identity of a phoneme depends on where in the pipe it is made. Of course, grown men's vocal tracts vary considerably in size and shape, but given standard statistical smoothing functions, the model estimate of the size of the pipe need not be very exact. The problem with young children is that their vocal tracts are *much* smaller than those of grown men. Young children are off the scale, as it were, and because of this the basic, adult-male recognizer works poorly (see Hagen, Pellom, & Cole, 2003).

On the other hand, adjusting the recognition algorithms for young children is no big deal. At worst, it requires collecting or licensing speech samples from children for use in determining a more suitable vocal tract estimate. Alternatively, in the 1990s at BBN Technologies, we mathematically "shrank" the adult-male model's vocal tract size by adding a simple, guesstimated coefficient and, interestingly, that seemed to work nearly as well with young readers as our later efforts based on empirically derived settings. Also useful to developers, research suggests that, at least prior to adolescence, using a single cross-age model works just as well as using different models for children of different ages or sizes (Arcy, Wong, & Russel, 2004; Hagen et al., 2003).

Determining the proper vocal tract settings for women (which, importantly, includes most teachers of young children) and for adolescents is slightly more difficult for two reasons. The first reason is that the physical sizes of people in these two groups vary so widely that even with normalization, a

significant subset will remain out of range no matter where the distribution is centered. The second reason is that the responses of neither women nor teenagers prove trustworthy when asked the simple question of whether their voice sounds more like a man, a woman or a child. No matter, the answer can be determined through a simple five-minute enrollment exercise that the student completes on initial sign-in to the system (Beattie, 2010).

In contrast to the other three issues, the fourth issue, pronunciation variability, has been truly nettlesome for speech-recognition applications, whether for children or adults. In recognition, transcription, and speech-understanding systems, pronunciation variability complicates decisions about what the speaker has intended to say. In reading support applications, it undermines the speech-recognizer's ability to decide whether a student has correctly or incorrectly grasped what she or he was supposed to say.

## 3.2 Detecting Word Recognition Difficulties

### 3.2.1. The Problem of Pronunciation Variability

With respect to the utility and acceptability of an automated reading tutor, the window of grace is very narrow. To the extent that the machine fails to recognize difficulties, it can neither assess nor assist the student as required. To the extent that the machine flags words with which the child did not actually have difficulty, its assessment data are equally invalid. Worse still, false interventions prove extremely annoying to students; if students dislike the system, they will not apply themselves.

Although young children are famous for such lexical distortions as /bisgetty/ for spaghetti, /alligator/ for escalator, and /allabody/ for everybody, these sorts of errors aren't really a problem. Such errors are not very frequent and, in any case, are the sorts of misperceptions that correct themselves as children learn to read and write (Adams, 1990; Bissex, 1980).

It is, instead, the pronunciation variants that arise even when children have properly perceived or decoded a word that cause trouble for automatic speech-recognition applications. Some of these—for example, /hēō/ for hill or /fwog/ for frog—are rooted



in poor phonological or articulatory refinement. Some are rooted in regional accents, such as /shok/ for shark, /dwog/ for dog, or /māyan/ for man. Some are rooted in dialect or other-language phonology, such as /mēster/ for mister, /wid/ or /wif/ for with, or /fahdoh/ for pharaoh. On top of these sorts of variations, comfortable human speech and fluent reading are marked by a number of phonological compromises such as syllable deletion (e.g., /varīty/ for variety), reduced vowels (e.g., /tuh/ for to or /yih/ for you), phoneme deletions (e.g., /ri/ for right and /wuhda/ for would have), and phoneme substitutions (e.g., /wuhjoo/ for would you) (c.f., Jurafsky et al., 2001). Note that, except perhaps for the first, none of these categories of phonological variation are common among adults as well as young children.

Because such pronunciation variability is rooted in the speaking habits of humans, it is not diminished by overtraining of the speech recognizer (Jurafsky et al., 2001). Speech scientists have therefore turned to adaptations of the speech-recognition network as means of coping with them.

One means of adjusting for such variation has been to require individual users to “train” the recognizer for her or his particular speech patterns. This is the approach often used with personal dictation systems, for example. To train the recognizer, the user is asked to read sets of prescribed passages aloud, continuing until the software has succeeded in attaining stable, accurate recognition performance. For any given system, both how much training material must be read and how well the training works varies from person to person. Regardless, because such recognition training relies on a considerable amount of fluent, accurate reading by the user on enrollment, it is not an ideal solution for young children or for poorer readers in general.

For systems that do not require such individual training, i.e., “speaker-independent” systems, the most common strategy for coping with pronunciation variability has been to augment the recognition network with families of acceptable pronunciations for the phonemes and words of interest. Unfortunately, except for systems with very small vocabularies, neither has this tactic proved adequate, for there is a trade-off between forgiveness and accuracy of recognition (or, in signal detection terms, between sensitivity and specificity). In recognition and dictation systems, increased tolerance for alternate pronunciations

of one word (one sound, one speaker) is inevitably coupled with increased misrecognition of others (c.f., Riley et al., 1999). Similarly, in a reading application, increased tolerance for alternate pronunciations must be coupled with decreased sensitivity to reading errors.

In view of this, Bell and colleagues (2003) have argued that conquering such phonological variation necessarily depends on using other dimensions of the language situation in conjunction with the phonological signal.

In most speaker-independent applications, this is done by turning the challenge for the speech recognizer into a multiple-choice task. Often the system is set up to constrain what the human is allowed to say to some enumerable set. An automated telephone call center might, for example, be designed to permit only numerals as responses (as, e.g., when requesting a caller's frequent flyer number). Alternatively, the system may list permissible answers for the caller (e.g., “Would you like to book a flight, change an existing reservation, or something else?”) or ask direct questions for which only a limited set of answers is possible (e.g., “What is your city of origin?”).

One might imagine that an analogous multiple-choice strategy would be useful for managing the problem of pronunciation variability in the reading situation. After all, correctly read words should, one-to-one, match the string of words in the text. The problem, however, is with words that are *not* read or recognized correctly. For the tutor to be helpful, the question is not whether the word that the student has voiced is similar to the word that should have been read: Given an alphabetic writing system such as English, mistaken readings of a word are nearly always phonologically similar to the word intended. Thus, the critical question for a reading tutor is, instead, whether what is voiced by the child suggests a failure or difficulty in recognizing or understanding the word intended.

In this spirit, a number of developers have endeavored to improve error detection in speech-recognition based reading systems by coupling target words to sets of foils chosen to represent likely spelling-to-sound errors. So far, and whether the best-bet errors have been generated on principle or, empirically, through analyses of children's reading corpora, results of these efforts have been disappointing (e.g., Fogarty et al., 2001; Mostow et al., 2002; Price et

al., 2009).

Again, following Bell and colleagues (2003), the question becomes one of what other dimensions of the language situation, in addition to letter-sound misrenderings, are indicative of reading difficulty. Surely there must be such information, for reading teachers and specialists agree quite reliably as to when children have misread words or are otherwise struggling with a textual difficulty (Cucchiari, Strik, & Boves, 2000; Harris & Sipay, 1990). These people are neither trained phoneticians nor mind readers. What is it that they hear?

Here is a thought experiment for the reader. Imagine that you are sitting in your office. You hear a young child's voice from down the hall, maybe in somebody else's office. Nobody tells you that this is the voice of a young child. They don't need to; it's obvious to your ear. Also obvious to your ear is that the child is not talking, but reading. In addition, you can tell whether the reading is easy for the child or hard for the child, and you can even hear exactly when the child is experiencing special difficulty. You can hear all of this without seeing or knowing the child, without seeing or knowing what she or he is reading, and without even being able to hear the words well if at all. What is it that you hear?

It was from this thought experiment that a team at BBN Technologies embarked on developing a reading tutor in the early 1990s. The answer, ultimately verified through analyses of timed transcriptions, is that the predominant sign of reading difficulty is reading dysfluency.

### *3.2.2. The Nature and Frequency of Reading Dysfluencies*

Recall that reading fluency emerges as the recognition of printed words becomes connected to the language-processing centers in the brain. For purposes of text understanding there are two advantages of these connections. The first is that the print on the page is mapped almost instantly and automatically to language. The second, due to the fact that the flow of information over established pathways in the brain is bidirectional (Bear, Connors, Paradiso, 2001; Dehaene, 2009; Gazzaniga, 2008), is that the language centers serve reciprocally to facilitate recognition and interpretation of each word as it arrives. It is this back and forth fluid—or fluent—handshaking of information that underlies

reading fluency and that allows the reader to focus active attention on the meaning of what is read rather than the processes of reading it.

Yet, when the flow is disrupted at any point, so too are fluency and understanding. When a word is unfamiliar, it cannot be mapped to language; there is a hiatus. When a word is misread or misunderstood, it misdirects the interpretive process; again, whether at that moment or later, when ensuing words become irreconcilable, the flow comes to a halt. When the language is too complex, the thread of the text is lost; the reader is reduced to word-by-word or phrase-by-phrase reading and comprehension fails. When active attention is directed to the process of decoding a word, it is necessarily removed from the process of understanding; the task of building the thread of the text is put on hold, at best.

A case in point is offered from the Center for Speech and Language Processing at the University of Colorado. The team had enhanced their reading-tutor software by adapting it for young children's voices. Finding, to their surprise, that young readers rarely pause at the ends of sentences, they also relaxed their active text window to span sentence boundaries. They implemented several layers of heuristics for predicting which word would be read next and for responding with alacrity when the child read some other nearby word instead. And, on top of all that, they required individual training of the system by each of their young readers. Even so, the word-recognition error rate of the system remained too high. (Hagen, Pellom, & Cole, 2003).

In an effort to identify what else might be done to improve the system's recognition accuracy, the Colorado team undertook an analysis of words that had been correctly and incorrectly "recognized" by their system when listening to young readers (Lee, Hagen, Romanyshyn, Martin, & Pellom, 2004). The database was collected from 106 children, including 17 in third grade, 28 in fourth grade, and 61 in fifth grade, with each reading one of 25 different stories that were roughly 1000 words in length.

Each of the words read by the children was marked if it had been produced correctly, fluently, and without interference. These words accounted for 92 percent of all of the words that the system had transcribed from the reading sessions, and it had recognized all but 5.7 percent of them correctly. All of the remaining 8 percent of the words that the system had transcribed were associated with some sort of

input flaw, and fully 31.5 percent of these words were “recognized” incorrectly. In order to determine which sorts of input flaws were most troublesome to the system, the research team divided them into categories.

When the students’ readings were marred by background noise, the system’s error rate was 15 percent. However, since background intrusions were rare, accompanying only 0.3 percent of all words, they accounted for just slightly more than 1 percent of the recognition errors on the full set of flawed words. Similarly, 58 percent of non-speech sounds, such as coughs and sniffles, were heard as words by the system, accounting for another 6 percent of its recognition errors with flawed input.

Where the children had mispronounced or misread words, the system’s recognition output was in error more than 40 percent of the time. However, the analysis also showed that mispronounced and misread words accounted for less than 1 percent of all of the words to which the system responded and only 12 percent of all of its false recognitions of flawed input.

The remainder of the flawed events, amounting to fully 80 percent of them in all, were due to dysfluent reading behaviors of the children, including repetitions of words, parts of words, or strings of words; pauses and hesitations; interjections such as “um”; letter- or syllable-wise decoding; and stretching of syllables or phonemes during word attack. Grouping these dysfluencies into categories revealed three particularly troublesome sets of problems.

The first set of problems centered on “breath noises,” which the system too often mistook for words. Though this may seem an amateur problem to many speech scientists, it should be pointed out that many children breathe *very* heavily when laboring to read. In addition, many seem to have runny noses for the duration of the school year. Managing this set of problem required better acoustic models of children’s breath noises.

The second set of problems was associated with the fact that the system had been programmed to count “recognized” words as having been read. In consequence, when a word or a string of words were repeated, the system was without immediate candidates with which to compare it. Because of this, repetitions were sometimes scored as incorrectly

read words. Alternatively, the system would map repeated words to same or similar words elsewhere in the text, thereby losing its alignment with the child’s actual position in the text which engendered still more errors. To manage this problem, the Colorado team revised the system so as to permit repetitions.

The third big set of problems centered on vocalizations tied to belabored word recognition. In particular, when the system heard the child enunciate the beginning of an expected word, it would often accept that fragment as a correct rendering of the whole word. This, too, produced two kinds of problems. The first, of course, is that a word that is only partially read is not correctly read: unless the system can discern partial from complete readings, it cannot correct the former. Second, having accepted the first part of the vocalization as the whole word, the system was left without any good immediate candidates for the word’s ensuing parts as they arrived. Motivated by these data, the Colorado team augmented the recognition network for its reading tutor with a sub-word recognition framework.

Operating in direct competition with the whole word candidates, the sub-word framework helps the recognition process in two ways. First, when a child vocalizes an initial fragment of a word, the corresponding sub-word candidate wins the recognition contest, beating all other candidates, including the target word itself. In this way, the sub-word recognition framework serves to protect the system from falsely accepting a partially articulated word as a complete and correct rendering of the whole word it expected. Second, through the sub-word framework, when the system has recognized an initial fragment of a word, it is biased to listen for the balance of the word. This serves to protect the system from falsely recognizing ensuing phonemes or syllables as other words even as it provides a means for discerning whether the word in focus has indeed been read fully and correctly, if effortfully.

At the level of dynamics, such sub-word trees serve much the same function as the sorts of mispronunciation foils explored by Mostow et al. (2002) and Price et al. (2007). Both approaches are intended to refine the system’s ability to discriminate whether or not the child has read correctly by setting up competition for the target word that is phonologically similar and that represents errors that the children are somehow expected to make.

Even so, the differences between the two approaches are equally important. First, the mispronunciation foils are referenced against canonical or “received” letter-sound correspondences and, thus, easily run amuck of differences in accents and variability in diction. In contrast, for the sub-word trees, colorings of accent and diction affect the target and its foils equally. Second, whether based on phonics principles or on analyses of pronunciation errors in children’s readings, generalization of mispronunciation errors to new words is risky both because the applicability and realization of phonics principles ranges widely within English orthography and because realization of phonemes varies widely with context as well as the language background and speech habits of the reader. In contrast, because the elements of the sub-word trees are based on the essential left-to-right, spelling-to-sound dynamic of the reading process and are phonologically competed against the word itself, they can be automatically generated for all text words (albeit with careful checking by a human) during the software-build process.

On implementing their sub-word framework, the Colorado team found that their system’s false alarm rate for partial words fell to 5 percent (Hagen, Pellom, & Cole, 2007) as compared to the prior 49.6 percent error rate reported by Lee et al. (2004). Beyond issues of system accuracy, such sub-word recognition frameworks the pedagogical potential of such sub-word frameworks is also significant. They provide the information necessary for discouraging students from visually glossing word endings, a common short-cut among weaker readers that interferes with sight word and vocabulary growth (Adams, 1990). They also provide the information necessary for discouraging students from failing to articulate or attend to word endings, a common speech habit within African American Vernacular English that interferes with reading acquisition (Charity, Scarborough, & Griffin, 2004).

Unfortunately, when the University of Colorado team compiled their data (Lee et al., 2004), they chose to tag each word with one and only one type of flaw. Thus, for example, if a word was tagged as having been read incorrectly, information as to whether it was also marked by pausing or hesitation was not tallied. In consequence, their compilation of flaws does not inform the full distribution of the children’s dysfluent behaviors; nor does it permit estimates of the relative frequencies of the counted occurrences of each type of difficulty.

Even so, several core points emerge from the work of the Colorado team. First, when common types of dysfluencies are included in the system’s language model, its ability to stay aligned with the children’s readings and to distinguish correctly from incorrectly read words increases markedly. Second, frank misreadings and mispronunciations of words while reading are relatively rare among young readers. Instead, the great majority of children’s “flaws” while reading correspond to dysfluencies—pauses, repetitions, and belabored efforts to decode. We are reminded of the thought experiment presented earlier. What we hear as we listen to the child down the hall is the special rhythm and cadence of fluent reading—and of its disruption.

To be useful, it is essential that a speech-recognition based reading tutor have the capacity to accurately monitor a child’s progress through a text despite the occurrence of errors and dysfluencies. In addition, of course, such a system needs means of evaluating divergences from the text as-written and of deciding whether and how to respond to them.



# 4 Using Dysfluencies to Monitor Reading Difficulty

The relation between children's reading fluency and their overall reading ability has long been appreciated by educators (e.g., Gray, 1919). In educational settings, reading fluency is conventionally assessed by asking the children, individually, to read a brief, grade-level text aloud as the teacher marks all errors and times the reading with a stopwatch. The value of this measure rests on the fact that the children's reading rate, as gauged in correctly read words per minute, varies directly with their overall reading ability as younger and poorer readers take significantly longer to read a given passage than do good readers. Across many studies, grade-school children's performance on such assessments generally correlates upwards of 0.80 with full-scale assessments of overall reading proficiency (for reviews, see Good & Jefferson, 1998; Marston, 1989). In addition, children's oral reading fluency scores align closely with global, subjective ratings of children's reading fluency by teachers and reading/language specialists (e.g., Cowie, Douglas-Cowie, & Wichman, 2002; Cucchiarini et al., 2000).

What are the factors that underlie the lower fluency scores of younger and poorer readers? In fact, children's rate of articulation increases with development, not reaching adult levels of maturity until age 12 (Lee, Potamianos, & Narayanan, 1999). In addition and independent of maturation, word articulation time is slower for poorer readers than for better readers (Cucchiarini et al., 2000; Lee et al., 1999). The number of reading errors made by a child also impacts the fluency score because each is subtracted from the total number of correctly read words. Nevertheless and by a large margin, the single greatest contributions to the slower reading times of younger and poorer readers are aberrant pauses and other sorts of time-outs within sentences (Clay & Imlach, 1971; Cowie et al., 2002; Cucchiarini et al., 2000; Goldman-Eisler, 1968; Mostow & Aist, 1997; Pinnell et al., 1995).

In other words, the reading patterns documented by the Colorado team are consistent with a massive

body of psychometric data on children's oral reading behaviors. Moreover, both are consistent with the larger portrait of reading comprehension and its development. The seamless flow of mental activity that transforms print to understanding depends on instantaneous recognition of the words of the text along with ready recognition of the language and meaning that the words are intended to carry. This flow is necessarily disrupted whenever the reader encounters a word or phrase that eludes, misleads, or requires active attention for its apprehension or understanding.

Again, such dysfluencies in a child's oral reading are shown to be strong predictors of clinicians' ratings of a child's reading ease or difficulty. They are also major determinants of clinicians' decisions about when a child needs help while reading (Harris & Sipay, 1990; Betts, 1946). For reading clinicians, dysfluencies are used in complement to the child's vocalizations in assessing needs and progress. For human listeners, in other words, dysfluencies provide that other dimension of the language situation that is so critically needed to cope with the variability of the phonological signal (Bell et al, 2003).

## 4.1 Reconfiguring the Speech Recognizer to "Hear" Pauses

Could dysfluencies be used by speech-recognition based reading tutors in a similar manner? The argument in this section, is that yes, they can, and to great effect. Designing a speech-recognition based reading tutor to do so, however, depends on conquering two component challenges. First, the system must be re-configured so as to register dysfluencies so that they can be evaluated. Second, there is the puzzle of how to evaluate the dysfluencies. Although reading difficulties are almost always accompanied by dysfluencies, not all dysfluencies are signs of difficulty. The system must somehow be designed to distinguish between those struggles

that warrant attention or assistance and those that do not.

For a good speech engineer, the first challenge—that of re-designing the system so as to register and time dysfluencies—is not hugely difficult. After all, computational speech-recognition depends on millisecond processing of the temporal parameters of the speech stream. On the other hand, every world-class speech engineer with whom I've worked has found it a hugely irritating request. After all, speech engineers and scientists are experts in phonology. They have devoted their lives and minds to thinking about speech in terms of phonology. Thus, there arises a disciplinary-based skepticism as to whether pauses or stammerings could possibly be informative. On top of that, there is a technological impediment. That is, in deference to the memory limitations of aural, speech-recognition engines are typically designed to discard silence, stammering, sighing, muttering, false starts, and repetitions from the user record. Moreover, their automatic excision is itself the product of much work and ingenuity.

Thus, in early work toward developing a speech-recognition based reading tutor at BBN Technologies, the first challenge was to persuade the engineers to redesign the recognizer to maintain timed records of such erstwhile detritus. Happily, having done so, they were rewarded with exactly the sort of evidence they needed to validate their effort. The speech recognizer, in itself, often rejected unconventionally pronounced words such as /hēō/ for hill or /seester/ for sister even when, save pronunciation, the child had given every sign that the words had been recognized and understood with fluency and confidence. On the other hand, when the children struggled or failed to recognize a word, their readings were almost always marked by pauses. Often, the pauses were coupled with vocalizations, as in the case of false starts, repetitions, and belabored decoding. Sometimes, they were wholly silent. Either way, the question now became one of discerning the likelihood with which such events were symptoms of difficulty.

## 4.2 Distinguishing Symptomatic from Acceptable Pauses

Not all pauses are signs of reading difficulty. Pauses also arise during fully fluent, mature reading. Good readers are strongly inclined to pause between

paragraphs and also just before and after quoted dialogue. Not infrequently, whether to breathe, to mark the break, or to think, good readers also pause at boundaries between sentences (Goldman-Eisler, 1968). Dismissing these sorts of pauses is straightforward since the boundaries of sentences, quotes, and paragraphs are explicitly demarcated in the text.

On the other hand, good readers also insert brief pauses to align meter with message. For instance, good readers insert brief pauses in their speech stream so as to move important words to stress beats for special emphasis. Here is an example, where bolding indicates stress, bolding plus italics indicates emphatic stress, and the / (slash) indicates a brief pause:

“He’s **talking** about the big **lion** that we **heard** about.”

“He’s **talking** about the / ***big*** lion that we **heard** about.”

“He’s **talking** about / ***the*** big lion that we **heard** about.”

Data collected at BBN suggested that wherever a non-syntactic pause exceeded the duration of a stress beat, it was safe to infer that it was a symptom of difficulty as opposed to a showing of expressiveness (Adams & Schwartz, 1998).

In the vast majority of cases, symptomatic pauses were found to occur in the immediate vicinity of a troublesome word (i.e., immediately before, after, or during). This was excellent news, for following Bell and colleagues' advice (2003), it offered a straightforward feature of the language situation that could be used in conjunction with the phonological signal to locate the source of reading difficulties.

Occasionally, however, symptomatic dysfluencies are more distal from their cause. For example, seeing an unknown word ahead, readers sometimes elongate, repeat, or pause unduly between leading but known words so as to buy time and protect the phrase structure of the sentence (e.g., “It waasss // // // // iit // wwaasss // it was unusual for the whales to come ....” Clearly the problem word in this example was neither it nor was, but unusual. Given additionally such behaviors as false starts, repetitions, repairs and word-skipping, it was clear to the BBN team that more was needed. A Markovian dysfluency grammar, encompassing common dysfluencies as

well as the correct pathway through the text, was built to support the recognizer's ability to keep track, with each utterance, of where in the text the child was and what she or he might be expected say next. That was as far as things progressed at BBN Technologies.

#### 4.3 Implementing the Pause Factor in a Speech Recognition-Based Reading Tutor

A year or so later the project was born again at a now-defunct company. Most of the early technical employees were from BBN Technologies, and the goal of the company was to make a speech-recognition based reading tutor that would be usable and useful in classrooms. The core recognizer, Sphinx III, was licensed from Carnegie Mellon University, and a children's voice model was developed from licensed data. The speech recognizer was augmented to deal with non-vocal and background noise, a sub-word recognition framework was set up in competition with the whole word recognition algorithms, and a dysfluency grammar was implemented. Further, instead of ignoring sentence boundaries, as the University of Colorado team had done, the user interface was set up to force the children to read sentence by sentence or, for very long or complex sentences, phrase by phrase. This partitioning of sentences was helpful in governing the dysfluency grammar as well as in managing the buffering and shipment of voice data from the children. Just as importantly, cognitive research shows that better readers actually do pause mentally (though not necessarily vocally) between sentences (Just & Carpenter, 1987), and that helping poorer readers to pause at sentence boundaries improves their comprehension (Kleiman, 1975).

Despite these and numerous other refinements, the system too often persisted both in intervening with corrections when it should not have and also in failing to offer help when it seemed clearly needed. The only obvious option left seemed to be that of requiring the children to individually train the system, as the Colorado team had done. But again, this is not an ideal requirement for young and poor readers.

##### 4.3.1. How Well Can Pauses Diagnose Difficulties: Clinicians vs. Computer

The diagnostics were then augmented with a pause factor. The pause duration was statistically estimated from data collected from 197 second- through sixth-grade children, each of whom had read a set of five or six 100-word passages. The passages in each set ranged in difficulty from well-below to well-above the child's grade level with the goal of obtaining reading data that ranged from easy and fluent to difficult and dysfluent for children at each age level.

"Ground truth" was established by asking two experienced reading specialists to evaluate each of the children's records. Very high frequency grammatical words, such as to, of, is, for, and the, were excluded from the analysis since data showed these words to be swallowed or substituted as often by good as by poor readers. Otherwise, the reading specialists indicated for every word in each child's record whether they believed it had been (1) read acceptably, (2) warranted immediate intervention, or (3) deserved post-reading review. No mention of pausing or timing was made to the reading specialists. Their assignment was simply to listen and re-listen to each record until they came to full agreement with each other on every word. Ultimately, the two reading specialists reached agreement on nearly 98 percent of the words.

A subset of the rated reading records were then divided into matched sets: A Training Set and a Test Set. The training set was used for exploring parameters that might estimate the specialists' ratings. The Test Set was set aside to permit clean corroboration of any estimates that were derived.

Within the two sets, the records were balanced for gender and represented a range reading ability and passage difficulties for each grade-level. Although a number of parameters and combinations of parameters were evaluated, the best predictor of the reading specialists' ratings was the presence of (filled and unfilled) non-syntactic pauses. The records in the Training Set were used to derive a statistical estimate of the pause duration (as measured by the time between offset of word<sub>n-1</sub> to onset of word<sub>n</sub>) that, across all children, best discriminated words that the reading specialists judged to warrant immediate intervention. This pause value was then added to the system in complement to its phonological algorithms for detecting reading troubles.

Table 1 shows the system's resulting intervention performance as compared to the (human) reading specialists. For both the Training Set and the Test Set of recordings, the system's decisions as to whether to accept or intervene agreed with the reading specialists on close to 95 percent of the words across grades. Results were comparable for English language learners and native-English speakers.

To lend perspective to these numbers, the same research program found that agreement between even the most reliable reading clinicians in marking students' reading-miscues was less than 90% on first-listens—which, of course, is the only listen afforded to humans in real time.

**Table 1.**  
**Intervention Performance: System vs. Reading Specialists**

Children	Training Set				Test Set			
	No. Text Words	% FN	% FP	Total % Correct	No. Text Words	% FN	% FP	Total % Correct
Grade 2	1130	2.4	2.9	94.7	1040	2.2	2.3	95.5
Grade 3	5540	2.1	1.7	96.2	5078	2.6	2.0	95.4
Grade 4	1130	1.5	2.3	96.2	1040	1.5	2.0	96.5
Grade 5	1017	1.7	2.5	95.8	624	1.6	0.9	97.5
Grade 6 (all)	9975	1.5	1.6	96.9	11177	1.6	1.4	97.0
Grade 6 ELL	3355	2.0	2.1	96.0	2289	1.7	1.1	97.2

**Total % correct** corresponds to how often the system's intervention decisions agreed with those of human reading specialists. False negatives (**FN**) correspond to instances when the system rejected a word that the specialists had accepted; false positives (**FP**) correspond to instances when the system accepted a word that the specialists had rejected. Among grade-six students, 31 percent (45% Spanish; 45% Portuguese) were English Language Learners (**ELL**). (From Adams, 2006).

#### 4.3.2. Testing the System's Tutoring Performance

With its intervention algorithms thus tamed, the next question was whether reading with the software would in fact accelerate children's reading growth. To find out, the impact of this software was evaluated with 410 mainstream children in the second through fifth grades of two matched schools within the same middle-class district. The research plan followed a quasi-experimental design: In one school, children in all grade 2 and grade 3 classrooms were assigned to read with the speech-recognition based tutor while those in grades 4 and 5 participated as controls; in the second school, children in the grades 4 and 5 classrooms were to read with the software while those in grades 2 and 3 participated as controls. Software use was scheduled for 30-minute sessions once or twice a week over 17 weeks. The actual number of sessions averaged 15, 20, 24, and 23 for the second- through fifth-graders,

respectively, where the actual amount of reading time averaged eight minutes per session or between two and three hours total across all sessions.

In other words, this was not a lot of reading time. But it still made a big difference. Assessments of the children at the end of the school year showed significantly greater fluency gains at every grade for those who had read with the software than those who had not with effect sizes ranging from 0.53 for second-graders to 0.26 for fifth graders. The inverse relation of fluency gains to grades related to the fact that there were far more second- than fifth-graders who yet lacked basic fluency in this middle-class school district. Relative to expected fluency growth per grade-level as documented by national norms, the effect size for those children who used the software was 0.90 standard deviations (Adams, 2006).



#### 4.4 Designing for Impact and Usability

Even the best of readers occasionally stumble and often compromise or distort the pronunciations of words when reading aloud. Thus, in addition to discerning when and where dysfluencies arise, a speech-recognition based reading tutor must decide what to do about such events. Again, the trade-off is that failing to correct an error amounts to a lost learning opportunity, but offering corrections inappropriately or too often is irritating to the students and disruptive of their understanding.

As an example, when good readers have misparsed or misunderstood where a sentence is going, they are inclined to backtrack, spontaneously rereading a phrase, word, or sentence so as to reprocess it. Because such repairs necessarily divert from the word-by-word organization of the text, a phonologically-centered reading tutor might be obliged to intervene. For developing readers, however, permitting the system to intervene on such occasions would be a mistake, for it would effectively stifle their occurrence. To the contrary, such repairs are to be encouraged among young readers as they are a sign of active thinking and understanding. Thus, another advantage of using pause parameters to monitor difficulties is that it enabled the system to refrain from interrupting the children whenever repairs and rereads were executed with alacrity.

Similarly, it is valuable to imbue the system with a sense of which words are most likely to pose difficulties and why. Although the high-frequency grammatical words (e.g., the, have, for, to) are often produced poorly by readers, they are unlikely to be visually unfamiliar and therefore unlikely to warrant correction except among very beginners. Instead, because of the lock-in nature of word acquisition, the words that are likely to cause difficulty are those that are new, whether visually or semantically. But again, what can be expected to be challenging depends on level: Where on-pace second-graders might be flummoxed by ambiguous spellings such as bead versus dead or bowl versus howl, on-pace fourth-graders are yet likely to balk when first encountering long (e.g., photosynthesis, environmental) and irregular words (e.g., colonel, ukulele, trough). Where readers struggle with words that are orthographically easy relative to their reading level (e.g., silt, meek, tamp), it is a good bet that the problem is vocabulary or usage rather than decoding. Vocabulary knowledge is also the prime

suspect where children repeatedly render a word incorrectly despite having previously been corrected.

Moreover, and regardless of level, what is easy for one child may be troubling for another. Clinicians report that one child may have trouble with certain vowel spellings, another may balk on even the most familiar words when a suffix is added, and still another may panic at virtually any long word. In fact, there is very little hard research data on how children's word recognition skills grow beyond first grade. Developers should consider programming the speech-recognition based reading tutor to collate the words that each child reads correctly and incorrectly over time. Beyond affording invaluable information for research on reading development, such data should be seminal in informing the system's own diagnostic, intervention, and instructional options.

Again, too, children tend to find interventions during reading aversive unless truly needed. Given the importance of providing help always but only when needed and the uncertainty of that need, developers will find it wise to set higher criteria for intervening during reading than for including words in review lists or activities.

There are many issues such as these that cannot be managed by the speech-recognition technology by itself, but instead must be relegated to sensitive design of the user interactions. The coda, therefore, is that in addition to good speech engineering, part of the art of building a good automated reading tutor depends on disciplined research and analysis and on pedagogical expertise. Tagging words according to their recognition difficulty depends on a good sense of the language's orthography and how it generally develops, and similarly for tagging words that are likely to be perplexing to children in meaning or usage. Writing clear and useful comprehension questions is a remarkably difficult task and so, too, is the job of creating glossary entries that will usefully clarify the meanings of words as they arise in different contexts. Because readability formulae remain poor indices of the complexity or difficulty of texts, pedagogical sensibilities are also needed for structuring libraries.

In addition, the value of any classroom system depends no more on how well it works with the children than on how useful and useable it is for their teachers. Beyond robustness, considerations for teachers extend to:

- » Painless management of student rosters including:
  - compatibility with the electronic enrollment files used by schools;
  - easy procedures for adding, removing, and editing student entries;
  - ready ways to sort data by student characteristics as required by state and district reporting protocols.
- » Means of exporting data, records, and content, whether for
  - special statistical examination;
  - lesson planning, building student portfolios; or
  - sharing with other specialists or parents.
- » Assistance for instructional support and lesson planning, including:
  - for creating student groups;
  - for setting and monitoring assignments; and
  - for a host of decisions about which data to summarize or make summarizable and how, so that teachers can manage progress and design needed instructional guidance.

These considerations also include such vital but regularly overlooked essentials as ensuring that teachers have their own accounts for exploring and experimenting with the content and behavior of the software as “students.” This seems so basic. How could a teacher possibly be comfortable turning her students loose on software with which she has no working familiarity? Yet, it is with stunning frequency that school software licenses include no provision for teachers’ user-accounts.

In designing a speech-recognition-based reading tutor, a close partnership between engineers, teachers, and other experts on reading development is indispensable.

# 5 Beyond Basic Fluency Support

For on-pace students, basic reading fluency should be attained by the end of second grade. But again, as documented by the NAEP (Daane et al., 2005), basic reading fluency continues to elude nearly half of fourth-grade students, nationwide.

The tutoring system described in Section 3.3 was designed to help children with reading difficulties in the course of reading, to redirect the children's attention to missed words and belabored sentences for review and rereading, and to maintain records of their progress and fluency over time. Among human tutors, these are core practices toward promoting basic reading fluency.

The evidence that the system, though used for relatively little time and even in its infancy, did indeed accelerate the children's fluency growth should be taken seriously, for unless and until basic reading fluency is established, reading for understanding and learning remain out of reach. Made available in all schools, speech-recognition based reading systems could provide ample, cost-effective, individual guided-oral reading sessions for *all* children—as needed by all children in order to gain basic fluency.

Even so, basic fluency support is hardly the only potential educational benefit of speech-recognition based reading technology. For one thing, basic reading fluency is only a starting point: Even among good readers, fluency growth does not begin to plateau until adolescence. And beyond fluency, just think about it. Imagine that every book had ears, a readiness to interact richly that was bolstered by all manner of activities and resources, plus a single-minded commitment to productively engaging the mind of every child who opened it. The educational possibilities are limitless.

This section explores three more potential applications for such technology: (1) helping students to acquire the linguistic structures and comprehension monitoring skills on which advanced reading fluency depends, (2) providing interactive

text-based support of content area learning, and (3) compiling information on students' needs and progress in ways that are more efficient, more informative, and less disruptive than conventional assessment options.

Although each of these three applications is arguably much needed in our classrooms, all three are offered as examples. The goal is one of provoking you, the reader, to continue contemplating how much and in how many ways such technology could benefit our educational progress.

## 5.1. Prosody and Understanding

According to the National Reading Panel, the vocal signature of fluency “is the ability to read a text quickly, accurately, and with proper expression” (2000, p. 3-5). Thus, in the measurement of reading fluency for the NAEP, students were evaluated not just on the speed and accuracy of their oral reading but also “for phrasing, adherence to the author's syntax, and expressiveness” (Daane et al., 2005, p. 5). But alas, the latter dimensions of fluency are precluded unless and until children are able to read the words of a text with adequate speed and accuracy (e.g., Laberge & Samuels, 1974; Perfetti, 1985) which, again, the NAEP showed nearly half of U.S. fourth-graders unable to do. In view of this situation, the importance of basic reading fluency gained national attention such that many schools now require timed oral reading assessments of all elementary students at least twice a year.

Unfortunately, as Jay Samuels (2006b) has cautioned, the emphasis on timed oral reading assessments may have become an over-emphasis in too many classrooms. The danger of over-emphasizing timed oral reading assessments is that it risks promoting rapid word recognition to an end in itself—in assessment, in the lessons and curricula that grow around the assessments, and in the students' habits and minds. Students can acquire new language, new thoughts, and new insights from print only as they

learn to monitor and critique their understanding of the texts they read (Chall, 1983). Doing so depends on what Pikulski and Chard (2005) have termed “advanced fluency.” As Samuels reminds us, “The most important characteristic of the fluent reader is the ability to decode and to comprehend the text at the same time” (2006a, p. 9).

In keeping with this, good tutors do more than listen for word-recognition difficulties. They also listen for whether children sound like they *understand* what they are reading. This they discern from the prosody—the patterns of rhythm, stress, and intonation of the child’s voice. As in the case of reading speed and accuracy, agreement among human listeners is quite reliable when asked to judge the appropriateness of students’ prosody (Cowie et al., 2002; Daane et al., 2005; Pinnell et al., 1995). Might it be possible to program a computer to evaluate children’s prosody as well?

English is what is known as a stress-timed language. Thus, the prosody of fluent oral reading is regularly if elastically metered by its stressed syllables, more or less as in a poem: *Once* upon a time, there was a *storyteller* who *lived* on the *edge* of the *village*. In essence, the reading is realized in waves, with each wave cresting at its stressed syllable. For good readers, the waves correspond roughly to phrases, modulated such that the syntactically and semantically most important words are located at their stressed crests. Less important words and syllables are unstressed as they are paced to fit the time available in between stressed syllables, even where that means speeding them and squashing them together. Relatively speaking, stressed syllables are both higher in pitch and longer in duration. In addition, final words of a phrase and, more so, of a clause or sentence, are prosodically marked as the duration of their vowels is stretched and the pitch of the voice is lowered. In complement, with the beginning of a new phrase, clause, or sentence, pitch and rhythm are reset such that the prosody of oral reading reflects sentence structure.

The question is how do young readers learn to do this? The tendency among speakers to stress the most important words of their utterances is universal and so, too, is the tendency among listeners to focus on those words that the speaker has stressed (Bolinger, 1983). But the written form of a word is the same regardless of how its author may have intended its importance or imagined its stress.

Nor does children’s oral language experience prepare them with models that they can usefully draw on when learning to read. First, the syntax of spontaneous speech tends to be ill-formed by written standards. “Sentences in spoken language are often difficult to identify,” wrote Chafe (1985, p. 11), continuing that the basic, natural unit of speech consists of a clause that typically has a single intonational contour, is composed of no more than seven words, a single idea unit, and lasts for no more than two seconds. Second, in spontaneous speech, importance trumps syntax in determining stress in any case (Shattuck-Hufnagel & Turk, 1996). Third, oral language exploits many other prosodic devices that variously serve to indicate the affect, importance, referent, function, or proper interpretation of the speaker’s words; typically, none of these is marked in written text.

History teaches us that the formalization of English syntax began to evolve only in the seventeenth century as English became a written language. Moreover, its standardization was owed principally to the authors of the English Enlightenment as it was developed and refined *specifically* in service of aiding interpretation and compensating for the ambiguities of the prosody-less medium of print (Halliday, 1985; Olson, 1994).

Thus, in written language, formal syntax is the principal means by which authors convey the intended focus and the structure of their language. It is through formal syntax that authors specify how the meanings of their words are to be interrelated and understood collectively. In oral language, most of this information is provided by prosodics and, indeed, oral language cannot be understood when its prosodics are disturbed (Huggins, 1978). In written language, by contrast, it is just the opposite: It is the syntax of the sentence that prescribes the prosodics.

It follows that to read a sentence aloud with appropriate “phrasing, adherence to the author’s syntax, and expressiveness” the reader must grasp the structure of the sentence and assign syntactic roles to its words (Chafe, 1988). And more: The reader must do so in the fleet moment between gleaning the words from the page and delivering them orally.

The quandary for the young reader is that the syntax of written language was invented for and is generally used only *in* formal written language (Biber, 2009; Miller & Weinert, 1998; Olson, 1994; Tomasello, 1998). Necessarily, then, *the syntax of written*

*language can be learned only through experience with print.*

Thus, meta-analyses show that growth in the ability to understand “book language” is one of the strongest outcomes of reading aloud with preschoolers (Bus, Van Ijzendoorn, & Pellegrini, 1995; Scarborough & Dobrich, 1994). Numerous studies have documented that both receptive and productive command of the syntactic structures that characterize written text develop only gradually across the school years and that children's sensitivity to the syntactic structures of written language predict their reading comprehension above and beyond their decoding abilities (e.g., Bohannon, 1975; Bowey, 1986; Levin & Buckler-Addis, 1979; Willows & Ryan, 1986). Moreover, use of written language structures in spontaneous speech is strongly related to higher education and, whether as enabler or consequence, seems linked to the requirement of higher education to read syntactically complex texts (Miller & Weinert, 1998).

Whether the measurement is based on the prosodic contours (i.e., timing and pitch) of their oral readings or by asking them to mark syntactic boundaries with a pencil, mature readers are found to conform closely to each other in their perception of the syntactic structure of text (Chafe, 1988; Goldman-Eisler, 1968; Klatt, 1987; Koriat, Greenberg, & Kreiner, 2002; Schwanenflugel, Hamilton, Kuhn, Wisenbaker, & Stahl, 2004). In contrast, younger and poorer readers cannot begin to demarcate the syntactic units of written language, and the prosodic contours of their oral reading often diverge substantially from the adult model. Children's ability to generate or interpret the syntactic structure of written language develops gradually across the school years in hand with their reading growth (e.g., Cutler & Swinney, 1987; Dowhower, 1991; Koriat, Greenberg, & Kriener, 2002; Schreiber, 1987, 1991).

Children who read with more adult-like prosodic contours demonstrate superior reading comprehension (Schwanenflugel, Hamilton, Kuhn, Wisenbaker, & Stahl, 2004). Indeed, Miller and Schwanenflugel (2008) have shown that the production of appropriate prosody by first-graders when reading the simple sentences of their first-grade books predicts their reading comprehension two years later in third grade. At the extreme, young children engage in what is known as “word by word” reading, lending equal stress to each word and

showing little respect for phrasing or even sentence boundaries (Clay & Imlach, 1971). Even students who show themselves able to render mature prosody with easier texts often lose it when texts involve longer and more complex sentences (Benjamin & Schwanenflugel, 2010; Klatt, 1987).

Because children's acquisition of syntax is strongly correlated with their linguistic experiences, its developmental course is characterized by large individual differences and a strong effect of socio-economic status (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002; Vasilyeva, Materfall, & Huttenlocher, 2008). Yet, Huttenlocher and colleagues have also shown that, given varied exposure to any given syntactic structure, young children are remarkably quick in learning to understand and use the structure in new sentences on their own. In one study with preschoolers, for example, the researchers read ten brief (30 sentence) stories that used a preponderance (60 percent) of passive sentences. Each story was read twice over a period of two weeks. In result, the preschoolers' ability to comprehend passive sentences as well as their inclination to produce them on their own, grew significantly (Vasilyeva, Huttenlocher, & Waterfall, 2006). Another study documented that preschoolers' comprehension and use of multi-clause sentences increased in direct proportion to the frequency with which their teachers happened to use such structures in their own daily speech (Huttenlocher et al., 2002).

Children seem similarly quick to internalize new syntactic structures encountered when reading on their own. But, again, this depends on whether or not they have grasped the structure and meaning of what they have read. Toward helping them do so, Dowhower (1991) summarizes several approaches that seem beneficial. The first is the practice of having the children reread passages aloud until they can read them fluently; meaningful gains in fluency are evident with up to four readings (Therrien, 2004). The second is that of physically demarcating the syntactic structure of the text *for* the children by, for example, inserting extra space or line breaks at phrasal and clausal boundaries (Frase & Schwartz, 1979; Kirby, & Gordon, 1988; O'Shea & Sindelar, 1983). Given the helpfulness of such demarcations, it is curious that breaking lines at phrasal and clausal boundaries is common practice in trade books for beginning readers but not in their school textbooks. The third is that of modeling the appropriate



phrasing and intonation of difficult sentences for the children (Chomsky, 1976). Best of all, of course, are guided oral reading sessions wherein the tutor can lead the child to focus on challenging sentences by providing corrective feedback, modeling their proper reading, requesting their rereading, examining their words and wordings, and eliciting paraphrases and discussion of their import (Therrien, 2004).

If a speech-recognition based reading tutor were designed to monitor the prosody of its tutees, it could assist in each of these ways. Although educational research on reading prosody has most often been focused on pitch contours (e.g., Cowie et al., 2002; Schwanenflugel et al., 2004), timing is arguably a better metric for such work. First, whereas there is a voluntary dimension to pitch modulation, the assignment of timing appears to be an early and obligatory part of the speech generation process (Klatt, 1987). Second, even while highly correlated with pitch, timing metrics can additionally capture information about pauses and about relative word stress. Patterns of relative word duration have been shown to correspond reliably to surface structure phrasal chunks (e.g., as marked by psycholinguistically naïve subjects) and, conversely, the alteration of relative word duration results in consensually unacceptable prosody (Aaronson & Scarborough, 1976; Huggins, 1978; Gee & Grosjean, 1983). In addition, temporal parameters are continually computed as a part of the speech recognition process and, as such, are technologically available for diagnostic exploitation with less extra development.

Figures 1 and 2 show the timing contours for five sixth-grade students, each reading the sentences, "Life in ancient Egypt was governed by the Nile. The Nile was predictable." In each figure, vocalization durations for words or word-parts are graphed in gray while pause durations are graphed in black.

The students shown in Figure 1 are strong readers, with average reading fluency greater than 150 words correct per minute across passages. As shown, sometimes these students do pause at sentence-final periods, and sometimes they don't. As evident from the graphs, two of the students treat "ancient Egypt" as adjective-noun pair, accenting the noun; the other two students instead treat it as a compound (accenting the first word, like New Yorkers do with *rye*'bread). Otherwise, the word-timing contours of all four students are virtually identical: Grammatical

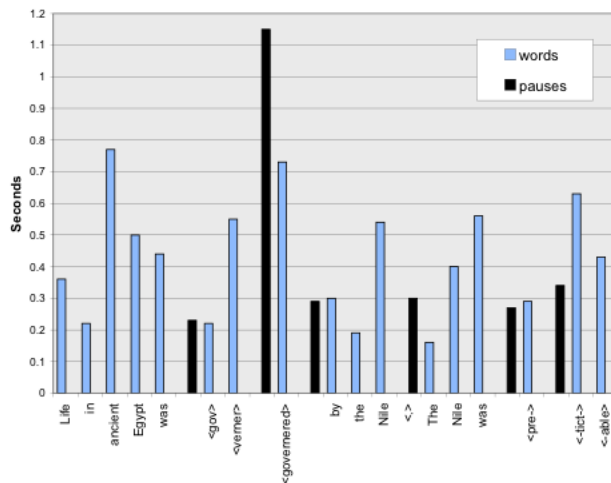
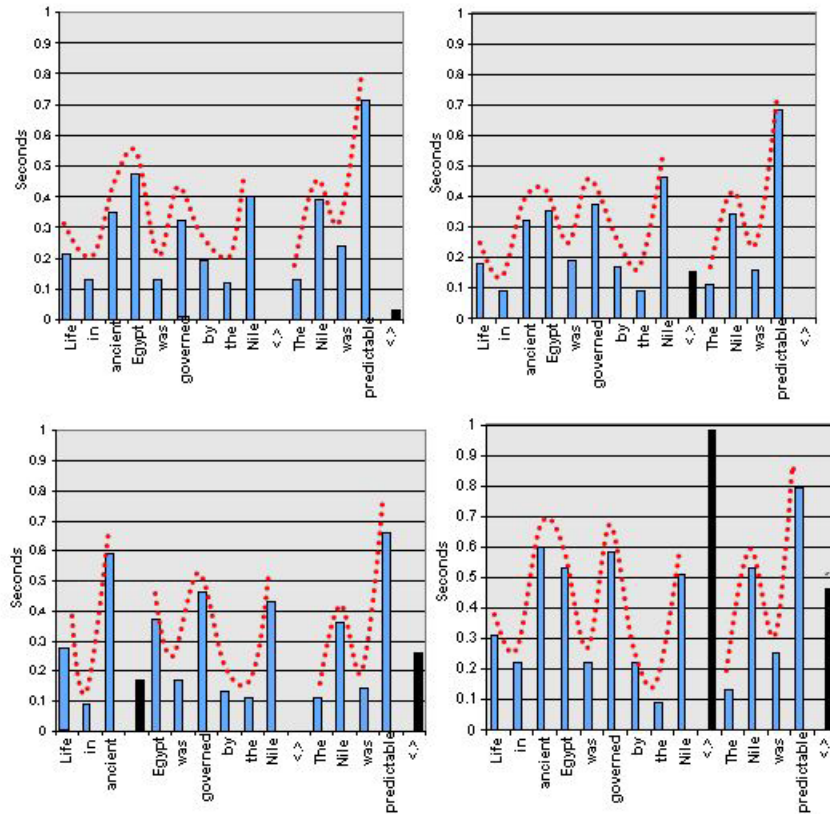
words are brief, and content words are stretched in accordance with their syntactic importance and elongated in sentence-final position. Note that the second student maintains this contour despite a slip in the middle of the first sentence.

In contrast, the student displayed in Figure 2 is a relatively weak reader with average reading fluency of less than 100 correct words per minute over passages. As shown, the student has difficulty with the text words governed and predictable. Both words are coupled with nonsyntactic pausing and neither is well-rendered despite the extra effort. Most especially, note the differences in the prosodic contour of this student's reading with those of the good readers. The relative durations of this student's words are reliably governed by neither their importance nor their syntactic status.

The importance of prosody—or, more precisely, of the dependence of comprehension on the reader's recognition of the syntactic structure and intended focus of what is read—prompts another suggestion for developers. Out of the box, speech recognition kits are designed to index texts as a single string of words, first to last, from the beginning of the passage to its end. In building reading applications, developers would instead do well to implement hierarchical indexing of pages, paragraphs, sentences and possibly of clauses and phrases, too. As the system matures, they will find such indexing very useful in ways ranging from diagnostics, record-keeping, adaptive querying, as well in conferring greater flexibility for highlighting the text for review and for positioning multimedia options.

### Figure 1 Good Fluency, Good Prosodics

Durations of successive words (blue bars) and pauses (black bars) as a function of text words and punctuation. The four students shown in Figure 1 are strong sixth-grade readers with fluency scores upwards of 150 words per minute. The dotted curves trace the prosodic contour of each reading as given by word durations.



### Figure 2: Poor Fluency, Disrupted Prosodics

Durations of successive words (blue bars) and pauses (black bars) as a function of text words and punctuation. The student shown in Figure 2 is a weak, sixth-grade reader with fluency scores beneath 100 words per minute.

## 5.2. Automatic Speech-Recognition for Managing Interactive Learning

A potentially invaluable asset of speech-recognition based reading applications derives from the fact that, by virtue of the speech-recognition layer, the computer continually knows exactly where in the text a student is. With proper design, this feature offers many different kinds of opportunities for enriching and extending students' reading experiences.

As an example, let us consider what we will here call a "Read and Think" application that was developed for the specific purpose of encouraging students to direct due attention to meaning, message and vocabulary in the course of content-area reading. The key to this application was the insertion of "invisible triggers" into the text map. Whenever the student's reading reached one of the invisible triggers, a pop-up question would appear on the screen. (See Adams, 2008).

Of course, it is commonplace for paper texts to be complemented with questions designed to structure and guide students' understanding. Often the questions are provided at the end of the chapter or section, but sometimes they are printed in call-outs or in the margins beside the text to which they pertain. Regardless, students often fail to appreciate the reading comprehension support such questions are intended to provide. Still more dysfunctionally, students often find ways to use such questions to avoid rather than to enhance their reading. How many parents have gotten wrinkles listening to their child explain that the teacher didn't *tell* them to read the chapter, but only to answer the questions at the end? Similarly, using eye-movement technologies, Paulson and Henry (2002) have shown how students are inclined to undermine the standard passage-plus-question reading assessment format by first reading the questions and then groping around the passage for the answer.

In the Read and Think design, by contrast, the students could not see a question or even know where one might arise except by reading the text to which it pertained. Nor were they allowed to skip questions. Because the questions were focused on terms or issues that had to be understood by the reader for the next section of the text to make sense, the software required students to answer each question correctly before moving on. The students could always do so by looking back at what they had just read. However, kids are very impatient people,

and re-reading something that they just read already is something many are loathe to do. But what is the option? Not being able to tell when a question might pop up, our students quickly learned that the most efficient way to beat the system was to think *while* they were reading, just in case. As one student remarked about this software, "I really like it because, see, it gives you something to do when you're reading."

To evaluate the impact of this software, it was trialed at the Sumter Correctional Institute in Bushnell, Florida, with adult intermediate readers (i.e., adults with reading levels equivalent to grade 3-7 students). Although intermediate adult readers currently make up more than 60 percent of the adult basic education (ABE) population nationally (National Center for the Study of Adult Learning and Literacy, 2006), research documents that the overall impact of ABE coursework on their literacy skills is small at best (see reviews by Kruidenier, 2002; Sheehan-Holt & Smith, 2002; Sticht & Armstrong, 1994; Venezky, Bristow, & Sabatini, 1994). Consistently, although all participants in this study had completed at least two and as many as six ABE courses, none had shown significant gains in their reading ability in result. They were stuck. Their reading scores were not budging.

In contrast, given 25 one-hour sessions with the Read and Think software, the men's reading levels grew dramatically. Their reading achievement grew by 2.3 grade-equivalent levels, from 5.27 to 7.54, as measured by the Woodcock Reading Mastery Test (WRMT) and by 1.7 grade-equivalent-levels, from 5.8 to 7.5, as measured by the Test of Adult Basic Education (TABE) (Smith & Adams, 2007).

Such results with adult literacy students are unprecedented. Clearly adults *can* learn to read given proper support. That, by itself, is a finding of great social importance. High school diplomas or GED certificates are increasingly required even for such jobs as driving trucks and packing grocery bags. Yet, people may take the GED exam only if they demonstrate the ability to read at a level equivalent to tenth grade.

The results of reading initiatives for at-risk school children, grades three and up, are equally as dismal as those for adult intermediates. Even in at-risk schools with strong primary-grade programs, such intermediate-level initiatives generally fail to maintain, much less to accelerate, students' reading



growth (Adams, 2008). Beyond phonics, the literacy progress of intermediate students—be they children or adults—depends on guided practice in reading, understanding, thinking about, and learning from increasingly complex texts. Effective, affordable reading support for intermediate students, whether children or adults, is crucial toward helping them move up and on with their lives. Again, speech-recognition based reading tutors potentially offer a key part of the solution.

Beyond general literacy support, one may also think about specific kinds of learning and thinking that such systems might assist. The fact that speech-recognition based reading tutors continually know exactly where in the text a student is, enables a wealth of different kinds of interactive teaching and learning dynamics. That is, the invisible trigger points described above are nothing more than calls to the computer and, as such, could be exploited in countless ways.

In reading a passage about the Lewis and Clark expedition, for example, the triggers might call up an interactive map, asking the student to show the expedition's starting point. Or, such interactive graphics might be deployed to ask the students to arrange and rearrange tokens in the course of reading a logic or math problem. Other texts might be clarified through sound files, for example, when reading about musical instruments, whale songs, or bird calls. Or perhaps the computer would ask students to verify interactively what they have just read about the pitch effects of the length, tension, or density of a vibrating string. Indeed, many scientific texts could be significantly enhanced through complementary video or animation, e.g., how the rotation of the earth results in day and night, the progress of a tsunami from formation to landfall, the growth of a chicken embryo from fertilization to hatching, or the workings of such culture-changing inventions as using gears to translate circular to up-down movements. In addition, of course, educators would be immeasurably grateful were such dynamics designed to present writing or note-taking prompts to students in course of their reading.

### 5.3. Speech-Recognition and Reading Assessment

Finally, there is the issue of assessment. At every level—from individuals to programs to systems—the possibility of managing and improving literacy instruction depends critically on good assessment. Yet, if there is one core issue on which virtually all literacy experts agree, it is that sensitive, informative assessments of reading are wanting (e.g., Committee on the Prevention of Reading Difficulties 1998; National Reading Panel, 2000; Paris & Stahl, 2005).

It happens that among tests of reading ability, oral reading fluency assessments stand out in terms of both validity and reliability. Again, oral reading fluency assessments consist of having students individually read a brief (typically 50–300 words in length, depending on grade), level-appropriate passage. The student's score is the number of words that she or he reads correctly per minute. In terms of validity, there is the frequent caution that oral reading passages be followed with comprehension questions so as to discourage mindless “NASCAR” reading. Given that precaution, oral reading fluency tests carry strong face validity, as the student's task is to read and understand connected text. They also evidence strong concurrent validity, yielding correlations with formal reading assessments such as the Stanford, the California, the Iowa, the Gates-McGinitie, and the Metropolitan that are generally upwards of 0.80. In addition, the reliability of oral reading fluency assessments (as measured by testing and re-testing a group of students a week or so apart and looking at the correlation between the two sets of scores) tends to hover around 0.90 when the same passage is read in both sessions and is nearly as high when different passages are read (for reviews, see Good & Jefferson, 1998; Marston, 1989).

In short, oral fluency assessments are relatively easy to administer, statistically reliable, and correlate strongly with measures of word recognition, comprehension, and full scale reading tests. This being so, many districts require that they be administered two or three times per year to all students in grades 2 through 5 and more frequently to students evidencing reading difficulty. Unfortunately, this is where their chief weaknesses arise. In such mandated administrations of oral fluency tests, teachers are given two objectives. The

first is to identify any children whose reading fluency is below-level as defined by norms for grade and time of year (e.g., Hasbrouck & Tindal, 2006). The second is to evaluate whether or not each individual student's oral reading fluency score has increased since the last testing occasion and, in particular, whether the slope of that change projects that the child's fluency score will be at or above grade-level by the end of the year. The problem is that, despite their impressively high reliability and validity statistics, the predictive sensitivity of oral reading assessments is relatively poor (Carlisle, Shilling, Scoot, & Zeng, 2004; Jenkins, Hudson, & Johnson, 2007; Roehrig et al., 2008; Stage & Jacobsen, 2001).

The test-retest reliability statistic measures the stability of the children's performance relative to each other, but it is indifferent to their absolute scores. That is, a test-retest correlation of 0.90 indicates that it is a pretty good bet that the rankings of scores across students will be similar across test sessions or passages: The student with the highest score on the first test is likely to obtain one of the highest scores on the retest, and so on. The problem is that children's fluency scores bounce around considerably from one test session to the next for reasons having nothing to do with their reading ability. This is true of the average scores for the group as a whole and still more so for the particular score earned by any given individual. Yet, decisions made about the help any individual student will or will not receive are based on the specific scores and score-changes from one testing occasion to the next for that one, particular, individual child.

Let us consider the error of these estimates in turn, starting with the children's fluency scores. Large-scale norms indicate that the average oral reading fluency of students in the winter of first through fifth grade are roughly 28, 72, 94, 112, and 128 words per minute, respectively (Hasbrouck & Tindal, 2006; Christ & Silberglitt, 2007). However, for any given child's fluency score, the Standard Error of Measurement has been found to range from 8 words per minute for first graders to 12 words per minute for fifth graders (Christ & Silberglitt, 2007; Howe & Shinn, 2002; Poncy, Skinner, & Axtell, 2005). That means, for example, that given a first-grader who attained a smack-average score of 28 correct words per minute on the winter oral fluency assessment, the teacher could be 95 percent confident that the student's "real" fluency score is someplace between 12 and 44 words per minute—scores that would respectively place the child deeply at-risk or well above-level.

Because estimates of oral reading growth rates are derived from the children's raw reading rates, the problem is similar but worse. In the primary grades, the actual rate of improvement of oral reading fluency is expected to be between one and two words per week (Deno, Fuchs, Marston, & Shin, 2001; Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993). Where children's improvement is slower than that, they are to receive intensive remedial assistance until their growth rate picks up. Again, however, due to measurement error, children's actual rate of growth is difficult to estimate from their performance on oral fluency tests. Collecting data from published research studies, Christ (2006) has estimated that the average error of the estimated growth slope is approximately 10 words per minute per week. Thus, if the results from a fluency assessment indicate that the child's fluency has increased 1.5 words per week, then the teacher could be 95 percent confident that the actual change per week in the child's fluency was someplace between a gain of 21 words per minute and a loss of 18 words per minute.

Clearly these are not useful numbers for purposes of setting instructional plans, much less for justifying educational designations. But in our schools, that is how teachers are told to use the children's fluency scores.

What is the source of this measurement error? Poncy et al. (2005) estimate that approximately half of it is owed to differences in the actual difficulty of the passages that children are asked to read during testing. As captured in the group test-retest reliability statistics, the ranked oral reading scores of a given group of children are generally very similar across passages. Nevertheless, the mean performance of the group can differ substantially across passages. Such differences in passage difficulty are shown to be poorly estimated by readability formulae and are pronounced even within equivalently "leveled" passage sets presented by developers (Ardoyn, William, Christ, Klubnik, & Wellborn, 2010; Francis et al., 2008). Other sources of measurement error derive from the happenstances of the testing situation and include such issues as procedural fastidiousness by the test administrator, disruptions that arise during the course of the test, and whether the child is having a good day or bad day. Consistently, the error is substantially larger when either passage calibration or test-administration vagaries are less tightly controlled (Christ, 2006).

As is to be expected, such "measurement error" averages out over multiple test occasions and passages.

Thus, Shinn (2002) recommends that fluency-based educational decisions be withheld until a minimum of ten data points, two per week, have been collected. Based on rigorous, large-scale statistical analyses, Christ (2006) concludes that measurement error comfortably asymptotes only after fifteen weeks of bi-weekly data collection, totaling 30 data points.

Practically speaking, it hardly matters whether the correct number of data points per estimate per child is 10 or 30. Given 25 or so students per class, both numbers far exceed the teacher's capacity. Remember: Besides testing kids, the teacher must also teach; and besides sitting around waiting to be tested, the kids must also be taught.

Thus, schools generally require no more than three oral reading scores per child per year, one in the fall, one in the winter, and one in the spring. This is a sad dilemma. The schools know that, in aggregate, such data ought to be useful, and they badly need for them to be useful. Yet the logistical demands of collecting the data ensure that they cannot be useful or, at least, not where they are most needed: In informing educational guidance for the individual students from whom they are collected.

Here again, then, is a priority need for a speech-recognition based reading application. If the children were regularly reading with the speech-recognizer, it would automatically collect their fluency for each different reading of each different passage on each different day. In doing so, moreover, it would neither disrupt nor displace learning time. It would instead collect these data unobtrusively, in the background, while at the same time enabling every child in the classroom to be productively engaged in learning through guided oral reading.

The software would continually collect and aggregate the children's fluency data with the sort of single-minded procedural homogeneity and measurement precision of which humans are incapable. It would automatically discard outliers and generate regression lines; it would objectively measure the relative difficulty of the different passages, sections of passages, and even words; and it would cogently summarize and report all such data to the children's teachers. Just as importantly, while hosting readings with the children, the system would free the teacher for the human dimensions of instructional support, whether to work with individuals on their special needs or to observe the children's reading clinically by watching over their shoulders as they work.

Arguably, such speech-recognition based reading software is just as direly needed in support of formal tests of reading achievement. Whether norm-referenced or criterion-referenced, formal reading tests generally follow a single measurement strategy. Passages are presented and the children's reading ability is inferred from the success with which they respond to associated comprehension probes. The nature of the comprehension probes may variously be multiple-choice questions, fill-in-the-blank as in the Cloze format, or prompts for short or extended written responses. People can and do argue heatedly over the advantages and disadvantages of these different kinds of comprehension probes. However, no one could possibly imagine any of them as a direct, valid, or diagnostic measure of the student's reading ability *per se*.

If such formal reading tests were instead presented on a computer enabled with speech-recognition reading software, it would be possible to separate children's ease or difficulty in reading the test passages from the accuracy, astuteness, well-formedness or penmanship of their responses to the comprehension probes.

For that matter, it is past the time to seriously consider the following suggestion from the National Research Council's Committee on the Foundations of Assessment, now ten years old:

One can imagine a future in which the audit function of external assessments would be significantly reduced or even unnecessary because the information needed to assess students at the levels of description appropriate for various external assessment purposes could be derived from the data streams generated by students in and out of their classrooms. Technology could offer ways of creating over time a complex stream of data about how students think and reason while engaged in important learning activities. Information for assessment purposes could be extracted from this stream and used to serve both classroom and external assessment needs, including providing individual feedback to students for reflection about their metacognitive habits.

A metaphor for this shift exists in the world of retail outlets, ranging from small businesses to supermarkets to department stores. No longer do these businesses have to close down once or twice a year to take inventory of their stock. Rather, with the advent of automated checkout and barcodes for all items, these businesses have access to a continuous stream of information that can be used to monitor inventory and the flow of items. Not only can business continue without interruption, but the information obtained is far richer, enabling stores to monitor trends and aggregate the data into various kinds of summaries. (Committee on the Foundations of Assessment, 2001, p. 284).



# 6 Education, Literacy, and Technology: Concluding Comments

## 6.1. On the Importance of the Problem

As set forth in the Massachusetts School Law of 1642, public education in America was first instituted precisely to make sure that every citizen could read. Now, nearly 400 years later, Americans as a group still cannot read, at least not very well. National assessment results document that this is true not just among school children in the United States but also among adults (Sum, Kirsch, & Taggart 2002; Kutner et al., 2007), and even among college-bound high school students (ACT Inc, 2006).

Some analyses suggest that the literacy skills of even our best and brightest are waning (for review, see Adams, 2009, 2011). However, the data from every source and on every subgroup indicate that it is the *unevenness* of our students' literacy outcomes that is at the core of our nation's poor literacy performance. In particular, reading performance is far weaker among students who come from families that are poor or less educated; that includes students for whom English is a second language, most saliently those of Hispanic background.

According to the NAEP, for example, the percentage of fourth-graders reading at grade level is:

- » 14 percent for blacks, 17 percent for Hispanics, and 43 percent for whites;
- » 17 percent for those eligible for subsidized lunch and 44 percent for those of higher household income;
- » 12 percent for those whose parents did not finish high school and 42 percent for those whose parents graduated from college;
- » 7 percent for students who are classified as English Language Learners and 35 percent who are not.

Conversely, the percentage of students whose reading performance fell below the basic-level cut-off is:

- » 54 percent for blacks, 50 percent for Hispanics, and 22 percent for whites;
- » 50 percent for those eligible for subsidized

lunch and 21 percent for those of higher household income;

- » 44 percent for those whose parents did not finish high school and 17 percent for those whose parents graduated from college;
- » 70 percent for students who are classified as English Language Learners and 31 percent who are not.

The NAEP shows these disparities in achievement levels to be essentially identical across grade-levels and testing domains and essentially constant over time (National Center for Education Statistics, 2007, 2008).

International assessments also highlight the unevenness of our students' literacy achievement. The performance of U.S. elementary students is slightly stronger than that of U.S. secondary students in these comparisons. Nevertheless, and regardless of age, the reading scores of students from the United States tend to cluster disproportionately within the very highest and the very lowest proficiency bands across participating countries (Lemke et al., 2002; Sen, Partelow, & Miller, 2005).

If such reading disparities are disturbing enough in themselves, it must still be borne in mind that they are only the top layer of a failure that is far more pervasive and debilitating.

First, it is not *only* reading that is at stake, for students' entire education depends heavily on reading. Thus, to the extent that students struggle with reading, they must also be handicapped in other subjects. Consistently, while the NAEP shows that only 33% of fourth-graders and 32% of eighth-graders are at or above grade-level in reading (National Center for Education Statistics, 2009b), it also documents that, nationwide, the percentage of fourth- and eighth-graders whose learning is at or above grade-level is only 29% and 32% in science, 34% and 22% in civics, 18% and 17% in U.S. history, and 39% and 34% in math (Grigg, Lauko, & Brockway, 2006; Lee & Weiss, 2007; Lutkus & Weiss, 2007; National Center for Education Statistics, 2009a, , respectively). Because no student is to take more than one of the NAEP



assessments, one might argue that these statistics are loose evidence of the co-morbidity of reading and other-subject difficulties. However, the Program for International Student Assessment (PISA), which does test the same students across domains, documents that of U.S. students who scored within the lowest of its six proficiency bins in mathematics literacy, 62 percent also scored within its lowest proficiency bin in reading; of those U.S. students who scored within the lowest of its six proficiency bin in reading, 82 percent scored within its lowest proficiency bin in mathematics (Miller, Sen, & Malley, 2007).

Second, because of the history of the NAEP and the way its statistics are compiled, its results make learning difficulties most visible for schools in neighborhoods of poverty. Yet, learning difficulties are by no means limited to such schools. To wit, the achievement gap *within* schools is typically wider than the average achievement difference *between* schools (Wong & Alkins, 2001; Wong & Lee, 1998). While the percentage is greater in at-risk neighborhoods, there are some children in every classroom who depend on school to learn to read. Furthermore, most children in most classrooms depend on school for *some* core aspect of their education. And all children in every classroom could and should be learning significantly more than whatever they bring to the school from home.

If the performance of our schools is not appreciably worse than in decades gone by, neither is it appreciably better. In any given domain, outcomes may fluctuate from time to time, rising or falling a little, due to any number of factors. The overall picture is one of a system that is mature, one that on average performs just about as well as it long *has* performed—because that is just about as well as it *can* perform given the resources and constraints within which it strives to operate.

The problem, as so passionately argued by Dustin Heuston (2011), is the delivery system. The output or work of any system, he reminds us, is a function of its potential output times its efficiency. For a school system, the work or output that matters is the learning of each and every one of its individual students. Thus, the output of the one-teacher-per-class model is inherently limited to what one teacher can accomplish per child. The productivity of whole-class instructional delivery rests on whether every child in the class has the same strengths and needs, the same confusions and interests, and the same infallible inclination to attend to and understand what the teacher is sharing. Failing any of these conditions—and, from one child or one lesson to the next, these conditions must always fail—

instructional effectiveness depends on the teacher's capacity to discover which students are in need of what kinds of extra support and then to help each child accordingly. Yet this, Heuston argues, is exactly what teachers in the conventional classroom cannot do. They do not have time.

Regardless of how we select, train, exhort, bribe, or extort our teachers, it remains the case that no teacher in any conventional classroom has the time or resources to discover much less to lend substantial guidance and support to the special needs or interests of each of the individual students in her or his charge.

We have tried everything over and over, renamed the same reforms, and then tried them again... No matter how we have tried or how much we have spent on each new reform, the gains have been disappointingly small and frustratingly local and temporary. Still worse, our educational budgets have become too rich for us to afford: economically, we are being forced to pare these budgets back even when, educationally, it is so obvious that far more is needed. (Heuston, 2011, pp. 1-2).

What America needs urgently, argues Heuston, is an educational delivery system that is effective and affordable, and his thesis is that computing technologies offer just that. Well-designed software can deliver individualized interactive lessons even as it interactively assists, assesses, and manages the child's progress alongside. If, as classroom studies indicate, one teacher can manage only one minute per day of individual attention to each of her or his students (Conant, 1973, cited in Heuston, 2011; Goodlad, 1977), then engaging students with well-designed instructional software for just 15 minutes would raise the productivity of each student's school day (week, year, career) by an order of magnitude.

The present paper has focused on technology to support reading and, more specifically, to promote reading fluency and comprehension. It is not that this is the only overdue use of technology in our classrooms. To the contrary, virtually every domain of school learning, including reading basics, could be taught faster and better with the help of well-devised instructional technology.

Nevertheless, better support for reading fluency and comprehension is the clear priority. First, more so than any other scholastic challenge, learning to read with fluency and comprehension depends heavily—perhaps inseparably—on individually guided practice and support. Second, so much of what people want or need to learn—in school or out, and with or without

the help of technology—depends on the ability to read as well as the knowledge, language and modes of thought that grow through having read.

## 6.2 On the Nature and Importance of the Solution

In this paper, it is argued that the principal reason that adequate reading ability eludes so many of our school children is that they fail to gain basic reading fluency. Without basic fluency, the process of reading requires intense concentration even as it is both error-prone and painfully slow. Moreover, the very arduousness of the process undermines the comprehension and learning that one might hope would attend the effort.

The brain sciences have shown that the emergence of basic reading fluency is coupled with the establishment of a special region of the cortex that is devoted to printed word recognition. This region, the Visual Word Form Area, is located in the left hemisphere of the brain and is tightly linked to the language centers of the brain. The Visual Word Form Area specializes in processing the orthographic properties of printed words—from individual letters, to pairs of letters, to syllables, to whole words—as distinct from their raw visual properties. In on-pace readers, the Visual Word Form Area becomes established toward the end of second grade. In contrast, the responsiveness to print of this area of the brain has been found to be weak or aberrant in developmental dyslexics and illiterates. (Dehaene, 2009; McCandliss, Cohen, & Dehaene, 2003; Schlaggar & McCandless, 2007).

Gaining basic fluency, in other words, is a big deal. Moreover, as excruciatingly difficult as reading can be for those who lack basic fluency, it is only *through* reading that basic fluency can be gained. Fluency depends on phonics, for it builds on the traces between spellings and pronunciations that are left each time a word is decoded. But it grows through reading as each of those traces is strengthened, refined, and more richly connected to language and meaning. Only through the practice of reading—that is, only as the outcome of encountering, re-encountering, and assimilating more and more printed words—can children gain the decoding automaticity on which basic fluency depends, and only through understanding the words and wordings encountered while reading can they tie them to the centers of language and thought in the brain.

Again, it is extremely difficult for children who yet lack basic fluency to read much on their own. It is for this reason that texts written for beginning readers

are short, conceptually simple, and contain much repetition. It is also for this reason that attaining basic fluency depends so heavily on children's having someone or something to guide them, whether to assist with textual difficulties, to support their on-going understanding, to help them to appreciate their own progress and accomplishment, or just to keep them engaged and on task.

A chronic impasse in the educational mission of our schools is their inability to provide anything close to the amount of individual guided oral reading on which the acquisition of basic fluency depends. In consequence of this impasse, the education that children can gain from school is strongly limited by the education that they bring to school from home. Yet, the reason for public schooling—the very hope and promise on which it was founded—is that of ensuring equal educational opportunity to all children, regardless of what their homes can offer.

Among children who develop reading on-pace, basic fluency emerges toward the end of second grade. In contrast, the NAEP (Daane et al., 2005; Pinnell et al., 1995) shows that 40 percent of U.S. fourth graders still lack basic fluency. What is the likelihood that a child who has failed to conquer basic fluency by fourth grade will ever do so? The results of the U.S. Department of Education's National Assessment of Adult Literacy (NAAL) similarly indicate that more than 40 percent of U.S. adults are unable to comprehend texts of moderate everyday difficulty (Kutner et al., 2007; also see Sum, Kirsch, & Taggart, 2002). Is that a coincidence?

The thesis of this paper is that our schools' incapacity to help students learn to read is an insuperably important problem, and that technology exists that can go a long way toward fixing it. Speech recognition-based reading tutors stand as a cost-effective and scalable means of ensuring ample, individual guided oral reading and learning support to every student in our schools. As reviewed, across past efforts to develop effective speech-recognition based reading tutors, major technological issues have been solved, and the technology's usability and efficacy has been demonstrated in classrooms with both children and adults.

Again, absent basic fluency, reading to learn is precluded. For this reason, assisting students' acquisition of basic fluency is the most urgent need for such technologies in our schools. But it is not the only one. As argued, speech-recognition based reading technology also offer strong benefits toward developing the linguistic sensitivity that is the hallmark of advanced fluency, for

supporting young readers' metacognitive control and enriching their content-area reading, and as a platform for more efficient, informative, and timely assessment regimen.

With respect to assessments, such a system could continually collect, collate, and analyze the children's performance. Moreover, it could do so in the background, in the course of instructional reading activities, rather than in displacement of them. Beyond any pencil and paper assessment, it could afford separate evaluation of the students' difficulty in reading the passages from their facility in responding to comprehension probes. It could give teachers continual access to student performance data for instructional use, and it could aggregate data from multiple readings, yielding more valid and reliable estimates of students' needs and growth. It also could collect a wealth of data still needed for the field's larger understanding of reading development. The nature of the hurdles and progress between beginning and advanced reading is still mostly a mystery to reading scientists. Similarly, sound approaches for measuring text complexity still belong to the future.

With respect to supporting and enriching text-based learning, such a system could provide access to countless books and texts, nominating titles at the right level and on the right topic for each child and each lesson. Better than any paper dictionary, it could provide definitions and meaning support for challenging vocabulary specific to the contexts in which they appear. Better than any workbook, it could select and populate learning activities in response to children's individual needs, extending and adapting the exercises as needed and without constraints of page space. Beyond any "read-now, lab-later" approach, such a system could be developed specifically for purposes of marrying the information and argument offered by the children's science texts to the kinds of interactive and supported hands-on, minds-on activities that are so valuable in complement. Such a system could similarly provide these sorts of support and instruction for second-language and foreign-language learning.

Most importantly, such a system could provide to every child the individualized guidance and support on which the acquisition of both basic and advanced fluency depends. Listening to each child read aloud, it could discern difficulties, provide real-time help, and make records of progress and difficulties in the background. By keeping every child engaged during reading time, it would ensure that each was spending time really reading even while keeping track of what,

how much, and how well each read. At the same time, it could provide ever-updated information to every teacher about the growth and needs of each child, about the successes and weaknesses of each learning goal, and about the special challenges of each text. And beyond alerting teachers to which kids and which challenges were in need of special attention, such a system would free time for teachers to provide such attention.

Again, unless and until children are able to engage productively with texts on their own, they need individual help both in learning to read and in reading to learn. In the history of the United States, public schooling was intended specifically to offer literacy to children whose homes could not provide such help. Nevertheless, providing young readers with the kinds and amount of individual support that we know they need is beyond the human capacity of the conventional classroom. For this reason, more than any other, the development of speech-recognition based reading technologies for our classrooms must be high on our country's educational agenda.

The maturity of automatic speech recognition technology is attested by its very ubiquity across other sectors of the economy, often in applications that are far more complicated than those in quest for our schools. Further—and in contrast to most of the software applications and dynamics one might dream of—neither is content development an issue: So much of what we would like our students to read, do, and think about already exists in print. Speech-recognition based reading and learning applications are well within reach were we to redirect just a fraction of the time, genius, and creativity that is now focused on developing ever more seductive ways for us to play games, watch ads, and otherwise waste our time with our mobile devices and computers.

The great social promise of technology, after all, has been centered on its potential for multiplying productivity. Nowhere is there a greater need to multiply our country's productivity than in our educational sector. Our future welfare, both individually and collectively, depends inseparably on the instructional productivity of all of our schools and the learning productivity of all of our students.



# 7

## References

- Aaronson, D., & Scarborough, H. S. (1976). Performance theories for sentence coding: Some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 2(1), 56-70.
- ACT, Inc. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading*. Iowa City, IA: Author.
- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: The MIT Press.
- 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, R., & D. Reinking (Eds.), *Handbook of literacy and technology, Vol. 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.), *Educating the other America: Top experts tackle poverty, literacy, and achievement in our schools* (pp. 277-300). Baltimore, MD: Paul H. Brookes.
- Adams, M. J. (2009). The challenge of advanced texts: The interdependence of reading and learning. In E. H. Hiebert (Ed.), *Reading more, reading better* (pp. 163-189). New York: Guilford.
- Adams, M. J. (Winter, 2010-2011). Advancing our students' language and literacy: The challenge of complex texts. *American Educator*, 34(4), 3-11, 53.
- Adams, M. J., & Schwartz, R. M. (1998). *Adding speech recognition to electronic storyware for beginning readers*. Cambridge, MA: BBN Technologies.
- Aikens, N.L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology*, 100(2), 235-251.
- Allington, R. L. (1983). The reading instruction provided readers of different reading abilities. *The Elementary School Journal*, 83, 548-559.
- Allington, R. L., & McGill-Franzen, A. (1989). School response to reading failure: Instruction for Chapter 1 and special education in grades two, four, and eight. *The Elementary School Journal*, 89(5), 529-542.
- Anderson, R. C. (1996). Research foundations to support wide reading. In V. Greaney (Ed.), *Promoting reading in developing countries* (pp. 55-77). Newark, DE: International Reading Association.
- Anderson, R. C., Mason, J., & Shirey, L. (1984). The reading group: An experimental investigation of a labyrinth. *Reading Research Quarterly*, 20(1), 6-38.
- Anderson, R. C., Wilkinson, I. A. G., & Mason, J. A. (1991). A microanalysis of the small-group, guided reading lesson: Effects of an emphasis on global story meaning. *Reading Research Quarterly*, 26(4), 417-441.
- Arcy, S., Wong, L., & Russel, M. (2004). Recognition of read and spontaneous children's speech using two new corpora. *Proceedings of the 7th International Conference on Spoken Language Processing*. Denver, CO: International Speech Communication Association, Jeju Island, Korea.
- Ardoin, S. P., William, J. C., Christ, T. J., Klubnik, C., & Wellborn, C. (2010). Examining readability estimates' predictions of students' oral reading rate: Spache, Lexile, and Forcast. *School Psychology Review*, 39(2), 277-285.
- Ash, G. E., Kuhn, M. R., & Walpole, S. (2009). Analyzing "inconsistencies" in practice:

- Teachers' continued use of round robin reading. *Reading & Writing Quarterly*, 25(1), 87-103.
- Autor, D. H., Katz, L. F., & Krueger, A. B. (1998). Computing inequality: Have computers changed the labor market? *Quarterly Journal of Economics*, 113(4), 1169-1213.
- Autor, D. H., Lawrence F., Katz, L. F., & Kearney, M. S. (2008). Trends in U.S. wage inequality: Revising the revisionists. *Review of Economics and Statistics*, 90(2), 300-323.
- Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4), 1279-1333.
- Barr, R., & Dreeben, R., (1983). *How schools work*. Chicago: University of Chicago Press.
- Bhattacharya, A. (2010). Children and adolescents from poverty and reading development: Research review. *Reading & Writing Quarterly*, 26(2), 115-139.
- Bear, M. F., Connors, B. W., & Paradiso, M. A. (2001). *Neuroscience: Exploring the brain*. Baltimore, MD: Lippincott Williams, & Wilkins.
- Beattie, V. (2010). Scientific learning reading assistant: CMU Sphinx technology in a commercial educational software application. *CMU Sphinx workshop 2010*, Dallas, TX.
- Bell, A, Jurafsky, D., Fosler-Lussier, E, Girand, C, Gregory, M, Gildea, D. (2003). Effects of dysfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, 113, 1001-1024.
- Benjamin, R. G., & Schwanenflugel, P. J. (2010, in press). Text complexity and oral reading prosody in young readers. *Reading Research Quarterly*, 45(4), 388-404.
- Betts, E. A. (1946). *Foundations of reading instruction*. New York, NY: American Book Co.
- Biber, D. (2009). Are there linguistic consequences of literacy? Comparing the potentials of language use in speech and writing. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 75-91). Cambridge, UK: Cambridge University Press.
- Birman, B. F., Orland, M. E., Jung, R. K., Anson, R. J., Garcia, G. N., Moore, M. T., Funkhouser, J. E., Morrison, D. R., Turnbull, B. J., & Reiser, E. R. (1987). *The current operation of the Chapter 1 program: Final report from the national assessment of Chapter 1*. Washington, D.C.: Government Printing Office.
- Bissex, G. (1980). *GNYS AT WRK: A child learns to write and read*. Cambridge, MA: Harvard University Press.
- Bohannon, J. N. (1975). The relationship between syntax discrimination and sentence imitation in children. *Child Development*, 46(2), 444-451.
- Bolinger, D. L (1983). Intonation and gesture. *American Speech*, 59(2), 156-174.
- Bowey, J. A. (1986). Syntactic awareness and verbal performance from preschool to fifth grade. *Journal of Psycholinguistic Research*, 15(4), 285-308.
- Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & Coll, C. G. (2001). The home environments of children in the United States Part I: Variations by age, ethnicity, and poverty status. *Child Development*, 72, 1844-1867.
- Burkam, D. T., Ready, D. D., Lee, V., & LoGerfo, L. F. (2004). Social-class differences in summer learning between kindergarten and first grade: Model specification and estimation. *Sociology of Education*, 77, 1-31.
- Bus, A. G., Van Ijzendoorn, M. H.; & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research*, 65(5), 1-21.
- Carlisle, J. F., Schilling, S. G., Scott, S. E., &

- Zeng, J. (2004). *Do fluency measures predict reading achievement? Results from the 2002-2003 school year in Michigan's Reading First Schools* (Technical Report #1, Evaluation of Reading First in Michigan). Ann Arbor: University of Michigan.
- Cavanaugh, C. L., Kim, A-H., Wanzek, J., & Vaughn, S. (2004). Kindergarten reading interventions for at-risk students: Twenty years of research. *Learning Disabilities: A Contemporary Journal*, 2(1), 9-21
- Chafe, W. (1985). Linguistic differences produced by differences between speaking and writing. In D. R. Olson, N. Torrance, & A. Hildyard (Eds.), *Literacy, language, and learning: The nature and consequences of reading and writing* (pp. 105-123). Cambridge, UK: Cambridge University Press.
- Chafe, W. (1988). Punctuation and the prosody of written language. *Written Communication*, 5(4), 396 - 426.
- Chall, J. S. (1983). *Stages of reading development*. New York: McGraw-Hill.
- Chard, D. J., Vaughn, S., & Tyler, B. J. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities*, 35(5), 386-406.
- Charity, A. H., Scarborough, H. S., & Griffin, D. M. (2004). Familiarity with school English in African American children and its relation to early reading achievement. *Child Development*, 75(5), 1340 - 1356.
- Chomsky, C. (1976). After decoding: What? *Language Arts*, 53(4), 288-296.
- Christ, T. J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review*, 35(1), 128-133.
- Christ, T. J., & Silberglitt, B. (2007). Estimates of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review*, 36(1), 130-146.
- Clay, M. & Imlach, R. (1971). Juncture, pitch, and stress as reading behavior variables. *Journal of Verbal Learning and Verbal Behavior*, 10(2), 133-139.
- Committee on the Foundations of Assessment. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, D.C.: National Academy Press.
- Committee on the Prevention of Reading Difficulties. (1998). *Preventing reading difficulties in young children*. Washington, D.C.: National Academy Press.
- Conant, Eaton H. (1973). *Teacher and paraprofessional work productivity*. Lexington, Mass.: D. C. Heath and Company.
- Cowie, R., Douglas-Cowie, E., & Wichmann, A. (2002). Prosodic characteristics of skilled reading: Fluency and expressiveness in 8-10-year old readers. *Language and Speech*, 45(1), 47-82.
- Cucchiari, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America*, 107(2), 989-999.
- 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.
- Cunningham, A. E., Perry, K. E., Stanovich, K. E., & Share, D. L. (2002). Orthographic learning during reading: Examining the role of self-teaching. *Journal of Experimental Child Psychology*, 82, 733-740.
- Cutler, A., & Swinney, D. A. (1987). Prosody and the development of comprehension. *Journal of Child Language*, 14, 145-167.
- Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading* (NCES 2006-469). Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of

- Education.
- Dehaene, S. (2009). *Reading in the brain: The science and evolution of a human invention*. New York: Viking.
- Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review*, 30, 507-524.
- Dowhower, S. L. (1987). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. *Reading Research Quarterly*, 22(3), 289-406.
- Dowhower, S. L. (1991). Speaking of prosody: Fluency's unattended bedfellow. *Theory into Practice*, 30(3), 165-175.
- Durkin, D. (1978-79). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 14(4), 481-533.
- Ehri, L. C., & Saltmarsh, I. (1995). Beginning readers outperform older disabled readers in learning to read words by sight. *Reading and Writing: An Interdisciplinary Journal*, 7, 295-326.
- Entwisle, D. R., & Alexander, K. L. (1996). Family type and children's growth in reading and math over the primary grades. *Journal of Marriage and the Family*, 58, 341-355.
- Entwisle, D. R., Alexander, K. L., and Olson, L. S. (2005). First grade and educational attainment by age 22: A new story. *American Journal of Sociology*, 110(5), 1458-1502.
- Fogarty, J., Dabbish, L., Steck, D. M., & Mostow, J. (2001). Mining a database of reading mistakes: For what should an automated reading tutor listen? *Proceedings of the tenth international conference on artificial intelligence in education (AI-ED) 2001*, San Antonio, TX.
- Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology*, 46(3), 315-342.
- Frase, L. T., & Schwartz, B. J. (1979). Typographical cues that facilitate comprehension. *Journal of Educational Psychology*, 71(2), 197-206.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect. *School Psychology Review*, 22, 27-48.
- Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal measures of reading comprehension. *Remedial and Special Education*, 9(2), 20-28.
- Gardner, D. (2004). Vocabulary input through extensive reading: A comparison of words found in children's narrative and expository reading materials. *Applied Linguistics*, 25(1), 1-37.
- Gazzaniga, M. (2008). *Cognitive Neuroscience: The biology of the mind*. New York: W. W. Norton & Company.
- Gee, J., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411-458.
- Goldin, C. D., & Katz, L. F. (2008). *The race between education and technology*. Cambridge, MA: Harvard University Press.
- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.
- Good, R. H. I., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61-88). New York: Guilford.
- Goodlad, J. I. What goes on in our schools? *Educational Researcher*, 6(3), 3-6
- Gray, W. S. (1919). Reading in the elementary schools of Indianapolis, II. *The Elementary School Journal*, 19(6), 419-444.
- Grigg, W., Lauko, M., & Brockway, D. (2006). *The nation's report card: Science 2005* (NCES 2006-466). U.S. Department of Education, National Center for Education Statistics. Washington, DC:



- U.S. Government Printing Office.
- Hagen, A., Pellom, B., & Cole, R. (2007). Highly accurate children's speech recognition for interactive reading tutors using subword units. *Speech Communication, 49*(12), 861-873.
- Hagen, A., Pellom, B., & Cole, R. (2003). Children's speech recognition with application to interactive books and tutors. *IEEE automatic speech recognition and understanding (ASRU) Workshop*, St. Thomas.
- Hairrell, A., Edmonds, M., Vaughn, S., & Simmons, D. (2010). Independent silent reading for struggling readers: Pitfalls and potential. In E. H. Hiebert & D. R. Reutzel (Eds.), *Revisiting silent reading: New directions for teachers and researchers*, (pp. 275-289), Newark, DE: International Reading Association.
- Halliday, M. A. K. (1985). *Spoken and written language*. Oxford, UK: Oxford University Press.
- Harris, A. J., & Sipay, E. R. (1990). *How to increase reading ability*. New York: Longman.
- Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher, 59*(7), 636-644.
- 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.
- Heuston, D. H. (2011). The third source: A message of hope for education. Salt Lake City, UT: The Waterford Institute.
- Howe, K. B., & Shinn, M. M. (2002). *Standard reading assessment passages (RAPs) for use in general outcome measurement: A manual describing development and technical features*. Retrieved August from <http://www.aimsweb.com>
- Huggins, A. W. F. (1978). Speech timing and intelligibility. In J. Requin (Ed.), *Attention and performance VII* (pp. 279-298). Hillsdale, NJ: Erlbaum Assoc.
- Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology, 45*(3), 337-374.
- Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*(4), 719-729.
- Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36*(4), 582-600.
- Jenkins, J. R., Matlock, B., & Slocum, T. A. (1989). Two approaches to vocabulary instruction: The teaching of individual word meanings and practice in deriving word meaning from context. *Reading Research Quarterly, 24*(2), 215-235.
- Juel, C. (1990). Effects of group reading assignment of reading development in first and second grade. *Journal of Literacy Research, 22*(3), 233-254.
- Jurafsky, D, Ward, W, Jianping, Z, Herold, K, Xiuyang, Y, and Sen, Z. (2001). What kind of pronunciation variation is hard for triphones to model? *Proceedings of ICASSP-01*, 577-580, Salt Lake City, Utah.
- Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn and Bacon.
- Kerckhoff, A. C., (1986). Effects of ability grouping in British secondary schools. *American Sociological Review, 51*(6), 842-858.
- Kirby, J. R., & Gordon, C. J. (1988). Text segmenting and comprehension: Effects of reading and information processing abilities. *British Journal of Educational Psychology, 58*, 287-300.



- Klare, G. R. (1984). Readability. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 681-744). New York: Longman.
- Klatt, D. (1987). Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America*, 82(3), 737-793.
- Kleiman, G. M. (1975). Speech recoding in reading. *Journal of Verbal Learning and Verbal Behavior*, 14(3), 323-339.
- Koriat, A., Greenberg, S. N. Kreiner, H. (2002). The extraction of structure during reading: Evidence from reading prosody. *Memory & Cognition*, 30(2), 270-280.
- Kruidenier, J. (2002). *Research-based principles for Adult Basic Education reading instruction* (National Institute for Literacy Contract No. ED-01-PO-1037). Portsmouth, NH: RMC Research Corporation.
- 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, D.C.: National Center for Educational Statistics, Institute of Education Sciences, U.S. Department of Education.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6(2), 293-323.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211-240.
- Lee, K., Hagen, A., Romanyshyn, N., Martin, S., & Pellom, B. (2004). Analysis and detection of reading miscues for interactive literacy tutors. *Proceedings of the 20<sup>th</sup> international conference on computational linguistics*. Morristown, NJ: Association for Computational Linguistics.
- Lee, S., Potamiano, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, 105(3), 1455-1468.
- Lee, J., & Weiss, A.R. (2007). *The Nation's report card: U.S. History 2006* (NCES-2007-474). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.
- Lemke, M., Calsyn, C., Lippman, L., Jocelyn, L., Kastberg, D., Liu, Y. Y., Roey, S., Williams, T., Kruger, T., and Bairu, G. (2001). *Outcomes of Learning: Results From the 2000 Program for International Student Assessment of 15-Year-Olds in Reading, Mathematics, and Science Literacy* (NCES 2002-115). U.S. Department of Education. Washington, D.C.: National Center for Education Statistics.
- Levin, H., & Buckler-Addis, A. (1979). *The eye-voice span*. Cambridge, MA: The MIT Press.
- LoGerfo, L., Nichols, A., & Reardon, S. F. (2006). *Achievement gains in elementary and high school*. Washington, DC: Urban Institute.
- Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, 66(4), 423-458.
- Lutkus, A.D., & Weiss, A.R. (2007). *The Nation's report card: Civics 2006* (NCES-2007-476). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.
- Machin, S., & van Reenen, J. (1998). Technology and changes in skill structure: Evidence from seven OECD countries. *Quarterly Journal of Economics*, 113(4), 1215-44.
- Manguel, A. (1996). *A history of reading*. New York: Viking.
- Manis, F. R. (1985). Acquisition of word identification skills in normal and disabled readers. *Journal of Educational Psychology*, 77, 78-90.
- Marston, D. (1989). A curriculum-based measurement approach to assessing

- academic performance: What is it and why do it. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford.
- McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, 7(7), 293-299.
- McIntosh, R., Vaughn, S., Schumm, J., Haager, D., & Lee, O. (1993). Observations of students with learning disabilities in general education classroom. *Exceptional Children*, 60, 249-261.
- Miller, D. C., Sen, A., and Malley, L. B. (2007). *Comparative indicators of education in the United States and other G-8 countries: 2006* (NCES 2007-006). Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Miller, J., & Schwanenflugel, P. J. (2008). A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading Research Quarterly*, 43(4), 336-354.
- Miller, J., & Weinert, R. (1998). *Spontaneous spoken language: Syntax and discourse*. Oxford, UK: Oxford University Press.
- Monroe, M. (1932). *Children who cannot read*. Chicago: University of Chicago Press.
- Moody, S. W., Vaughn, S., & Schumm, J. S. (1997). Instructional grouping for reading: Teachers' views. *Remedial and Special Education*, 18, 347-356.
- Mostow, J., & Aist, G. (1997). The sounds of silence: Toward automated evaluation of student learning in a reading tutor that listens. *Proceedings of the fourteenth national conference on artificial intelligence (AAAI-97)* (pp. 355-361). Providence, RI: American Association for Artificial Intelligence.
- Mostow, J., Beck, J., Winter, S., Wang, S., & Tobin, B. (2002). Predicting oral reading miscues. *7th International conference on spoken language processing*. Denver, CO: International Speech Communication Association.
- Nagy, W. E., Anderson, R. C., & Herman, P. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24, 237-270.
- Nation, K., & Snowling, M.J. (1998). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development*, 69(4), 996-1011.
- National Center for Education Statistics. (2007). *The Nation's report card: Reading 2007*. Retrieved from [http://nationsreportcard.gov/reading\\_2007/](http://nationsreportcard.gov/reading_2007/)
- National Center for Education Statistics. (2008). *The Nation's report card: Long term trends 2008*. Retrieved from [http://nationsreportcard.gov/ltr\\_2008/](http://nationsreportcard.gov/ltr_2008/)
- National Center for Education Statistics. (2009a). *The Nation's report card: Mathematics 2009* (NCES 2010-451). Washington, D.C.: Institute of Education Sciences, U.S. Department of Education.
- National Center for Education Statistics. (2009b). *The Nation's report card: Reading 2009* (NCES 2010-458). Washington, D.C.: Institute of Education Sciences, U.S. Department of Education.
- National Reading Panel. (2000). *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* (NIH Publication No. 00-4754). Washington, D.C.: U.S. Government Printing Office.
- Olson, D. R. (1994). *The world on paper*. Cambridge, UK: Cambridge University Press.
- Olson, D. R. (2009). A theory of reading/writing: From literacy to literature. *Writing Systems Research*, 1(1), 51-64.
- O'Shea, L., & Sindelar, P. (1983). The effects of segmenting written discourse on the reading comprehension of low and high performance readers. *Reading Research Quarterly*, 18(3), 458-465.

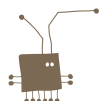
- Paris, S. G., & Stahl, S. A. (Eds.) (2005). *Children's reading comprehension and assessment*. Mahwah, NJ: Erlbaum Associates.
- Paulson, E. J., & Henry, J. (2002). Does the degree of reading power assessment reflect the reading process? An eye-movement examination. *Journal of Adolescent and Adult Literacy*, 46(3), 234-244.
- Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.
- Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, 58(6), 510-521.
- Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B., & Beatty, A. S. (1995). *Listening to children read aloud*. Washington, D.C.: National Center for Educational Statistics, Office of Educational Research and Improvement, U.S. Department of Education.
- Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2005). An investigation of the reliability and standard error of measurement of words read correctly per minute. *Journal of Psychoeducational Assessment*, 23(4), 326-338.
- Price, P., Tepperman, J., Iseli, M., Duong, T., Black, M., Wang, S., Boscardin, C. K., Heritage, M., Pearson, P. D., Narayanan, S., & Alwan, A. (2009). Assessment of emerging reading skills in young native speakers and language learners. *Speech Communication*, 51, 968-984.
- Provasnik, S., Gonzales, P., & Miller, D. (2009). *U.S. performance across international assessments of student achievement: Special supplement to the condition of education 2009* (NCES 2009-083). Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Reitsma, P. (1983). Printed word learning in beginning readers. *Journal of Experimental Child Psychology*, 36, 321-339.
- Reitsma, P. (1989). Orthographic memory and learning to read. In P. G. Aaron & R. M. Joshi (Eds.), *Reading and writing disorders in different orthographic systems* (pp. 51-73). Dordrecht, NE/Norwell, MA: Kluwer Academic.
- Riley, M. D., Byrne, W., Finke, M., Khudanpur, S., Ljolje, A., McDonough, J., Nock, H., Saraclar, M., Wooters, C., & Zavaliagkos, G. (1999). Stochastic pronunciation modeling from hand-labelled phonetic corpora. *Speech Communication*, 29, 209-224.
- Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the dibels oral reading fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology*, 46(3), 343-366.
- Samuels, S. J. (2006a). Reading fluency: Its past, present, and future. In T. Rasinski, C. Blachowicz, & K. Lems (Eds.), *Teaching reading fluency: Meeting the needs of all readers* (pp. 7-20). New York: Guilford.
- Samuels, S. J. (2006b). Toward a model of reading fluency. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about fluency instruction* (pp. 24- 46). Newark, DE: International Reading Association.
- Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review*, 14(3), 245-302.
- Schatz, E. K., & Baldwin, R. S. (1986). Context clues are unreliable predictors of word meanings. *Reading Research Quarterly*, 21(4), 439-445.
- Schlaggar, B. L., & McCandless, B. D. (2007). Development of neural systems for reading. *Annual Review of Neuroscience*, 30, 475-503.
- Schreiber, P. A. (1987). Prosody and structure in children's processing. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending oral and written language* (pp. 243-270). New York: Academic Press.
- Schreiber, P. A. (1991). Understanding prosody's role in reading acquisition. *Theory into Practice*, 30(3), 158-164.
- Schwanenflugel, P. J., Hamilton, C. E., Kuhn, M. R., Wisenbaker, J. M., & Stahl, S.

- A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology*, 96(1), 119-129.
- Sen, A., Partelow, L. A., and Miller, D. C. (2005). *Comparative indicators of education in the United States and other G8 countries: 2004* (NCES 2005-021). Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Sénéchal, M., & LeFevre, J. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development*, 73(2), 445-460.
- Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology*, 87, 267-298.
- Share, D., & Shalev, C. (2004). Self-teaching in normal and disabled readers. *Reading and Writing: An Interdisciplinary Journal*, 17, 769-800.
- Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25, 193-247.
- Sheehan-Holt, J., & Smith, M. C. (2002). Does basic skills education affect adults' literacy proficiencies and reading practices? *Reading Research Quarterly*, 35(2), 226-243.
- Shinn, M. R. (2002). Best practices in using curriculum-based measurement in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 671-698). Bethesda, MD: National Association of School Psychologists.
- Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, 21, 459-479.
- Simmons, D. C., Fuchs, L. S., Fuchs, D., Mathes, P., & Hodge, J. P. (1995). Effects of explicit teaching and peer tutoring on the reading achievement of learning-disabled and low-performing students in regular classrooms. *The Elementary School Journal*, 95, 387-408.
- Slavin, R. E., Karweit, N. L., & Wasik, B. A. (1994). *Preventing early school failure: Research, policy, and practice*. Boston: Allyn & Bacon.
- Slavin, R. E., Karweit, N. L., & Madden, N. A. (1989). *Effective programs for students at risk*. Boston: Allyn & Bacon.
- Smith, R. D., & Adams, M. J. (2007). Efficacy of computer-mediated guided oral reading for developing reading fluency and comprehension among ABE-level prison inmates. Unpublished manuscript.
- Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review*, 30(3), 407-419.
- Stedman, L. C., & Kaestle, C. F. (1987). Literacy and reading performance in the United States, from 1880 to the present. *Reading Research Quarterly*, 22(1), 8-46.
- Sticht, T., & Armstrong, W. (1994). *Adult literacy in the United States: A compendium of quantitative data and interpretive results*. Washington, D.C.: National Institute for Literacy. (ERIC Document Reproduction Service No. ED 371 241).
- 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.
- Swanborn, M. S. L., and de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, 69(3), 261-285.
- Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education*, 25(4), 252-261.
- Tomasello, M. (Ed.) (1998). *The new psychology of language: Cognitive and functional*



- approaches to language structure*. New York: Taylor & Francis.
- Vasilyeva, M., Huttenlocher, J., & Waterfall, H. (2006). Effects of language intervention on syntactic skill levels in preschooler. *Developmental Psychology*, 42(1), 164-174.
- Vasilyeva, M., Waterfall, H., & Huttenlocher, J. (2008). Emergence of syntax: Commonalities and differences across children. *Developmental Science*, 11(1), 94-97.
- Vaughn, S., Levy, S., Coleman, M., & Bos, C. S. (2002). Reading instruction for students with LD and EBD: A synthesis of observation studies. *Journal of Special Education*, 36, 2-15.
- Vaughn, S., Moody, S., & Schumm, J. S. (1998). Broken promises: Reading instruction in the resource room. *Exceptional Children*, 64, 211-226.
- Vellutino, F., Scanlon, D. M., Sipay, E., Small, S., Pratt, A., Chen, R., & Denckla, M. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, 88, 601-838.
- Venezky, R. L., Bristow, P. S., & Sabatini, J. P. (1994). Measuring change in adult literacy programs: Enduring issues and answers. *Educational Assessment*, 2, 101-131.
- Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, 36(4), 541-561.
- Wasik, B. A., & Slavin, R. E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly*, 28, 178-200.
- Wilkinson, I., Wardrop, J. L., & Anderson, R. C. (1988). Silent reading reconsidered: Reinterpreting reading instruction and its effects. *American Educational Research Journal*, 25(1), 127-144.
- Willows, D. M., & Ryan, E. B. (1986). The development of grammatical sensitivity and its relationship to early reading achievement. *Reading Research Quarterly*, 21(3), 253-266.
- Wong, K. K., & Alkins, K. (2001). Toward systemic reform in high-poverty school: Title I schoolwide programs in two large districts. In D. L. Taylor and E. A. Kemper (Eds.), *Readings on Equal Education*, Vol. 17 (pp. 83-118). Brooklyn, NY: AMS Press.
- Wong, K. K., & Lee, J. Y. (1998). Interstate variation in the achievement gap among racial and social groups: Considering the effects of school resources and classroom practices. *Advances in Educational Policy*, 4, 119-142.





advancing  
children's learning  
in a digital age

the Joan Ganz Cooney Center  
at Sesame Workshop

The Joan Ganz Cooney Center  
One Lincoln Plaza  
New York, NY 10023  
p: (212)595-3456  
[joanganzcooneycenter.org](http://joanganzcooneycenter.org)