In R. Fisher, M. Lewis, & G. Brooks (Eds.), Raising standards in literacy. London: Falmer Press.

Textbooks and Model Programs:

Reading Reform in the United States

Elfrieda H. Hiebert

University of Michigan

Efforts to reform reading education in the United States share similar goals with the reform efforts in the United Kingdom. The digital age is a time when the demands of literacy have increased, for information is more copious and more accessible than at any prior period in history. Judging the importance of one source over another, choosing what to remember from a source, and using this information in projects and in communication are among the critical literacy proficiencies of the digital age. In both nations, educators and policymakers are engaged in reform efforts to ensure students' attainment of these literacy proficiencies.

Although the underlying motivation for reading reform is the same in both nations, reading reform efforts in the United States and in the United Kingdom—England in particular—have manifested themselves in different ways. In the United Kingdom, the National Literacy Strategy states a clear goal for 11-year-olds on a particular assessment and prescribes the Literacy Hour as a means for achieving that goal (Fisher & Singleton, 2000). In contrast, while a U.S. goal has been and continues to be that "every child can

read by the third grade" (President Bush, 2001), no tasks, texts, or assessments accompany presentations of the goal. At the individual state level, considerable time and investment has gone into establishing standards. However, none of these standards have included tasks or texts that definitively show what grade-level proficiency looks like at different levels (Stotsky, 2000). While almost every state has a statewide reading assessment, the connection between the tasks and texts of the assessments and the state standards often goes unarticulated.

While the preeminence of states and local educational agencies (LEAs) in guiding their own educational policies continues to be part of political rhetoric, two reading reform efforts can be regarded as national in scope. The origins and foci of these reforms differ, but each is likely to influence the practices of individual teachers and their students in states, counties, and towns across the nation. The first reform involves the revision of reading textbooks by America's two largest states, Texas and California. These two states' policies influence the content of textbooks in less populous and less centralized states. The second reform consists of model school instructional programs that are implemented with the help of external consultants. Over the past decade, the U.S. Department of Education has provided funds for model programs to states for distribution to LEAs. Through such federal programs, the practices of teachers and their students around the country can be influenced.

Evaluating the efficacy of reading reform efforts, which is the goal of this chapter, requires perspective on the definition of the literacy problem in the United States. Consequently, we first review the results of the National Assessment of Educational Progress (NAEP). Although U.S. education is designated as the responsibility of states rather than the federal government, the NAEP results are presented as "the nation's report card" (U.S. Department of Education, 2001). With the NAEP results as a background, the content of U.S. reading reform activities and efficacy of the these efforts at attaining the goals of the reform movement are reviewed. This review points out obstacles to achieving the goal of higher literacy levels in the digital age. The third, and final, section of the chapter is devoted to proposals for removing those obstacles.

How Are American Students Reading?

The 2000 NAEP report on fourth graders' reading achievement (Donahue, Finnegan, Lutkus, Allen, & Campbell, 2001) is part of a series of congressionally mandated assessments of different subject areas at grades 4, 8, and 12 that began in 1969 and that are conducted every several years. Since 1992, fourth graders' literacy proficiency has been classified into four levels: (a) advanced, (b) proficient, (c) basic, and (d) below basic. In 2000, 37% of fourth graders were below basic, 31% were basic, 24% were proficient, and 8% were advanced. The distribution has not changed substantially from 1992, when the classification system was first used.

When the NAEP results were compared across states for the first time in 1994 (Campbell, Donahue, Reese, & Phillips, 1996), their impact was particularly strong. Of particular interest was the performance of California's students. In 1987, California mandated that all selections in the reading textbooks of all elementary grade levels purchased with state funds needed to be "authentic" literature (California English/Language Arts Committee, 1987) or texts sold from a trade book division, as opposed to texts written specifically for a textbook program. California's low standing in the 1994 state-by-state comparison of the NAEP was taken to be an indictment of literature-based instruction or whole language. While this consideration was not taken into account in subsequent interpretations of the data, California's 1994 classrooms had a higher percentage of linguistically and culturally diverse students and recent immigrants than classrooms in any other state. Many reading reform efforts, the reports of blue-ribbon panels of researchers on reading education (National Reading Panel, 2000; Snow, Burns, & Griffin, 1998), changes in and emphases on reading assessments, and the debates about phonics and whole language can be traced to California's performance on the 1994 NAEP.

By the time the 1994 NAEP results were available, California schools had purchased textbooks for another seven-year cycle. However, the state of Texas, which in 1990 had enacted a mandate similar to California's for authentic literature (Texas Education Agency, 1990), moved to reverse this policy in 1997 (Texas Education Agency, 1997). Hoffman, in the next chapter, describes Texas's mandates regarding decodable

texts for first graders. For those textbooks that will be purchased in the fall of 2002, California also has mandated decodable texts (California English/Language Arts Committee, 1999).

In the flurry of activity that followed the 1994 NAEP results, little attention was paid to a special study that examined the meaning of failure to attain a proficient level on the NAEP (Pinnell, Pikulski, Wixson, Campbell, Gough, & Beatty, 1995). A subsample of fourth graders was asked to read aloud a passage to which they had read and responded silently. The findings showed the majority of students (including those who scored below basic) were very accurate in their oral reading—that is, they could "say the words." There was no significant difference in accuracy of word recognition between those who scored "basic" or higher and those who did not. What differentiated the students most was the rate at which they read: The students who scored below basic read significantly slower than those students who were rated basic or above.

All but a very small percentage of an American age cohort can read the words in texts (Pinnell et al., 1995). For a sizable portion, this reading is slow. Plodding reading makes sophisticated interpretations of text difficult. The kind of instruction that supports automatic word recognition and sophisticated interpretations of text is likely to be quite different than the kind of instruction that supports acquisition of the alphabetic principle. As the following descriptions of the reform efforts show, however, policymakers have been focused on the latter rather than the former.

Reading Reform at the State Level: Textbooks

Whereas education may be the prerogative of LEAs, national companies provide the textbooks for instruction and the tests for establishing whether instruction was successful. Furthermore, when the largest states are also the ones with centralized responsibility for education, such states wield considerable influence over textbook publishers. This is precisely the scenario that exists in the United States. Its largest two states—California and Texas—account for approximately 22% or more of the nation's children and have centralized textbook selection procedures. Only one other large state—Florida—adopts textbooks centrally, but its guidelines have not been as prescriptive and its adoption of textbooks occurs concurrently with California's. While textbook publishers have long recognized the profitability of the California and Texas markets, it was not until the late 1980s that these two states began to use textbook content as their primary means of reading reform. In the 18 smaller states that adopt textbooks statewide and the 29 states where individual school districts or schools select their own textbooks, educators have little leverage that would allow them to obtain textbook programs that are compatible with their regional policies.

What Is the Reform?

Textbook programs for reading in the United States are called basal reading programs, although their content includes writing, speaking, and listening. A great many instructional schemes for reading/language arts are published in the United States such as

sets of small books, typically in paperback booklets numbering 8–16 pages. However, the production of the comprehensive literacy programs that are the focus of this chapter is limited to a small group of companies, currently six or seven. These comprehensive literacy programs are used by approximately 85% of American elementary classrooms (kindergarten through grade 5 or 6; Baumann, Hoffman, Duffy-Hester, & Ro, 2000) and consume the lion's share of funds for educational materials.

Between 1997 and 2000, the annual figure for purchase of basal reading programs was approximately $1.36 billion dollars (Association of American Publishers, 2001). While the market is lucrative, publishers' investment in basal reading programs can be risky (Chall & Squire, 1991). These programs have gotten so gargantuan that the initial investment is large—approximately $50 million dollars to produce a new program. Further, the return on the investment can be slow. The typical life of a textbook program in schools is six to eight years, leaving the unsuccessful publishers six to eight years before their product can be considered again in that context.

Publishers' large investments and the consequences they suffer have been exacerbated by California's and Texas's increased efforts to prescribe the contents of basal reading programs. The swing to authentic literature and then to decodable text in Texas and California mean that textbook publishers need to develop a new copyright every three years (the time between the California and Texas textbook adoptions) if they hope to remain competitive in these two states.

Five programs were officially approved for purchase with state funds in Texas for the textbook cycle that began in the fall of 2000. The programs contain many components. For grade one alone, approximately 30 different types of materials are offered in the catalog of the most widely purchased program in Texas (Farr et al., 2001). Fourteen guidebooks for teachers accompany the first-grade program. While first-grade programs are the most prolific in components, subsequent grade levels have numerous student texts and teacher guidebooks as well. The components of a single basal textbook program can easily fill an entire bookcase.

What Is the Evidence for the Efficacy of This Reform?

The textbooks are massive interventions, as policymakers have recognized. But are the efforts accomplishing the goals set out by policymakers? Have more children attained proficient reading levels in California and Texas as a result of these efforts? While states such as Texas and California have massive assessment programs the results of which can be viewed internationally on the World Wide Web, achievement of students in districts or schools as a function of textbook programs has yet to be reported. When the effects of reading primarily from different kinds of texts such as those that emphasize phonetically regular words and those that emphasize high-frequency words have been compared, text type has frequently been confounded with activities and form of teacher support. Insights into the effects of texts on student reading achievement come from studies of children reading passages from different textbook programs and from summa-

ries of the textbook program characteristics that are associated with beginning reading acquisition.

Comparisons of textbook programs. A major emphasis of the large-scale studies that began in the 1960s (e.g., the U.S. Department of Education's First-Grade Studies [Bond & Dykstra, 1967]) was the examination of the effectiveness of programs that contain different kinds of texts. But findings from these studies have been either inconsistent or inconsequential when comprehension, and not only word recognition, is measured and when student achievement is considered in subsequent grades (Lohnes & Gray, 1972). In the First-Grade Studies, particularly, variation across classrooms within a method was considerable. Further, methods often differed substantially in activities such as the amount of writing and spelling and in teacher support as well as in types of text. Recently, the problems inherent in equating a type of text with a program are apparent in an examination of different approaches to phonics instruction by Foorman, Francis, Fletcher, Schatschneider, and Mehta (1998). In addition to different approaches to phonics, the three approaches examined in the Foorman et al. (1998) study involved different types of texts and different instructional activities that have been shown to influence students' reading achievement regardless of text type, such as writing and spelling.

It was not until Barr and Dreeben (1983) used an alternative paradigm that the influence of text on beginning reading acquisition began to be understood. Barr and Dreeben examined the relationship of student achievement to the phonic structure and number

of different words that students read in their texts. Barr and Dreeben reported that what the texts covered was the variable most closely associated with first-grade learning, accounting for 83% and 71% of the variance in basal and phonics learning respectively and for 50% of the variance in reading achievement overall. They also reported that the difficulty of the materials and the amount of time teachers devoted to reading instruction covaried. Teachers with more difficult reading programs allocated more time to reading instruction. The number of phonics concepts covered during first grade was less responsive to the group mean aptitude than to the number of phonics concepts contained in the first-grade materials and the time allocated to phonics activities.

Juel and Roper/Schneider (1985) also examined students' reading achievement as a function of text characteristics. In this case, however, students received the same phonics lessons but they read texts from two different basal reading programs, which differed significantly only in the characteristics of the preprimers (one with decodable words and the other with high-frequency words). The two factors that accounted for end of grade 1 reading performance were students' initial scores on the Metropolitan Readiness Text and the basal series. Students who read from the decodable preprimers were more likely to learn letter-sound correspondences early and to use decoding knowledge when encountering unfamiliar words.

Both the Barr and Dreeben and Juel and Roper/Schneider studies were conducted when texts for beginning readers followed particular rules on pace and repetition. The texts have changed substantially since the early 1980s when those studies were con-

ducted. When Hoffman and his colleagues (1994) compared the textbooks with authentic literature that were adopted in Texas in 1993 with those that had been used in the period prior to this switch, they found that the number of unique words in first-grade texts between 1987 and 1993 had increased by almost 50%. Further, vocabulary control had been replaced with predictable syntactic patterns that encouraged children to use picture support, rhyme, and repeated patterns and phrases to decode text. Since this shift, there have been no reports with research designs similar to those of Barr and Dreeben (1983) and of Juel and Roper/Schneider (1985) examining children's reading development as a function of textbook coverage. Further, since Hoffman et al.'s (1994) analysis, yet another type of textbook program—this time, decodable texts--has been produced in response to Texas's mandate for decodable texts (Texas Education Agency, 1997). The research on the efficacy of authentic literature-based and decodable texts is limited to descriptions of children's reading of texts, in the case of the former, and to descriptions of text characteristics, in the case of the latter.

Children's reading of recent textbook programs. As has been described, the 1993 copyrights consisted of literature, rather than the specially written stories emphasizing particular vocabulary that characterized the texts of the mid-1980s. In two separate analyses of children's ability to read the literature-based texts, a sizable percentage of end-of-year first graders were unable to read even the first levels of these texts fluently: 45% in a sample in the Midwest (Hiebert, Liu, Levin, Huxley, & Chung, 1995) and approximately

35% in a sample in Texas (Hoffman, Roser, Patterson, Salas, & Pennington, 2000). On a fairly typical measure of word recognition—the high-frequency words from an informal reading inventory—these two groups of students read 8.4 (Hiebert et al., 1995) and 10.4 (Hoffman et al., 2000) words on average. A student reading at a first-grade level would score 60 on this measure.

In an instructional study of children's word learning in the predictable texts that served as authentic literature in the first-grade components of the past decade's textbook programs (Hoffman et al., 1994), Johnston (2000) reported that the highest readers remembered 30 of the 160 unique words in the predictable texts at the end of three weeks, the middle readers 15, and the lowest readers 6. Texts that have high numbers of unique words, as is the case with the texts in literature-based programs, do not facilitate the reading acquisition of any but the highest students.

Analyses of the features of textbook programs. Effects of different kinds of texts on students' reading achievement are difficult to capture when one type of text quickly replaces another kind. One technique for considering text demands in a continually changing context is to apply a task framework to texts. Task analyses describe the linguistic and conceptual knowledge needed to perform a task—in this case, to independently read a text. Hiebert (2001) has applied a task framework to the texts from Texas-adopted programs as well as those from the textbooks of previous generations.

The texts intended for the first third or half of grade1 have more phonetically regular words, as Texas mandated (Hiebert, 2001). On other features, however, the texts share the characteristics of authentic literature (Hiebert, 2001). Specifically, the texts continue to have high numbers of unique or different words per every 100 running words of text. The variables of pace (how many new words are presented in lessons) and repetition (how often words appear in instructional materials) that Barr and Dreeben (1983) identified as critical to students' reading achievement have not been part of the mandates. Comparing a cluster of the first texts from the 1980s (prior to the California mandate for literature) and the early 1990s to those of the 2000 copyright reveals that the number of unique words per 100 running words of text were 5 (1983), 29 (1993), and 21 (2000). Whereas children were asked to learn 72 new words in the first five weeks of first grade in 1983, children need to be able to read 211 different words during the same period with the 2000 books.

Millions of dollars and thousands of hours of teachers' time are devoted to obtaining and using new textbook programs. There is no empirical evidence that the texts that have been mandated over the last 15 years by California and Texas make it easier for children to learn to read or for teachers to teach their students to read.

Reading Reform at the Federal Level: Reform Models

Beginning with the Elementary and Secondary Education Act (ESEA) in 1965, the federal government has increasingly provided states with funds for intervening or providing more intensive support for reading instruction, especially for children from

low-income families. Currently, the Title I program aimed at students who qualify as poor (based on their eligibility for free or reduced school lunches) provides states with $8.4 billion dollars annually. Until recently, however, there have been no mandates regarding the instructional methodology to be used with these funds. This situation is in transition, as the existence of the Comprehensive School Reform Demonstration (CSRD) program shows.

What Is the Reform?

In 1998, Congress implemented the CSRD program to permit LEAs and schools to better use federal, state, and local funds in low-income schools. The centerpiece of the CSRD program is the use of "well-researched and well-documented models for school wide change that are supported by expert trainers and facilitators" (Northwest Regional Educational Laboratory, 2001). Annually, $145 million is available to state education agencies to make grants to districts for implementation of research-based models in individual schools. Schools can select models that are not listed in the legislation with the following stipulation: "[If] a school can demonstrate that the model selected will help the school implement a comprehensive program, it is acceptable." The legislation specifies nine components that constitute a coherent, well-designed comprehensive school reform program:

- Effective, research-based methods and strategies
- Comprehensive design with aligned components

- Professional development
- Measurable goals and benchmarks
- Support within the school
- Parental and community involvement
- External technical support and assistance
- Evaluation strategies, and
- Coordination of resources.

The congressional legislation implementing CSRD listed 17 comprehensive school reform models as demonstrating the intent of the program. One of the regional educational laboratories—the Northwest Regional Educational Laboratory (NWREL) in Portland, Oregon, an agency funded by the U.S. Department of Education—was given the responsibility for developing a catalog of models. As of spring 2001, there were 11[1] reading-specific models, 35 overall school models, and 10 models that fall into the "other" category, half of which include an emphasis on reading.

These model programs efforts are, for the most part, separate from the mainstream textbook programs in which the states and LEAs invest a major portion of their funds for reading instruction. According to NWREL, efforts that promote published materials (i.e., textbook programs) are no longer eligible for inclusion on the CSRD list (Buehler, 2001).

[1] Thirteen models are listed on the website of models. However, 2 of these models have been withdrawn at the request of the developers.

But this policy appears to have been a recent one. Included in the original congressional legislation was a list of model programs that demonstrated the aim of the funding; among the schoolwide demonstration programs, the Direct Instruction model promotes its own texts and programs entitled Reading Mastery and Success For All publish their own texts as well. Among the reading-specific models, two of the efforts are directly aligned with a published reading program: Breakthrough to Literacy with the publisher of the Wright Group little books, and Early Intervention in Reading with Houghton Mifflin's intervention readers. A third reading-specific model, Reading Recovery, encourages particular sets of guided reading little books. Further, the leader of Reading Recovery efforts in North America, Gay Pinnell, is an author of Scholastic's Literacy Place reading program, which is one of the five that was adopted for use in the last Texas reading adoption. In none of these programs, however, have the effects of the texts been analyzed separately. As will become evident in the next section, there is little evidence to substantiate the effectiveness of the numerous components that are part of the schoolwide and reading-specific projects, even among those programs offered by Congress as demonstrative of the legislation's intent.

What Is the Evidence for the Efficacy of This Reform?

In that the first component of the CSRD program is "well-researched and well-documented models for school wide change that are supported by expert trainers and facilitators," as mentioned above, and the fourth component is the presence of "measurable

goals and benchmarks," two forms of data would be anticipated: (a) data providing the basis for a project's identification as a demonstration project, and (b) ongoing data on students' achievements from the implementation of the model funded with CSRD funds. In neither case is there a substantial amount of data. Data have been very sparse for the original models as well as those that have been subsequently added to the program. A review of the data that program developers provided the federal government is examined first, followed by a discussion of the nature of ongoing data.

Herman and her colleagues (1999), who summarized the evidence from the 17 schoolwide models identified as demonstration programs in the congressional legislation, concluded that only 3 had any proof that their implementation made a positive difference in student achievement. Data on schoolwide models will not be reconsidered but, instead, a closer examination is given to the reading-specific models, which are the particular focus of this chapter. The reading-specific models are presented in Table 1.

Insert Table 1 about here

As the data in columns 2 and 3 of Table 1 indicate, convincing proof that students in economically challenged neighborhoods leave schools with higher levels of literacy is not available for any of the reading/language arts models.

Most of the 11 reading/language arts reform models are specific to a particular age level. Three focus on beginning reading instruction only (Breakthrough to Literacy, Literacy Collaborative, and Reading Recovery), while data on another three programs

were gathered when beginning reading was their focus (Early Intervention in Reading, First Steps, and Exemplary Center for Reading Instruction). Although early interventions are effective, they require changes in the reading experiences of subsequent grades if gains in literacy are to be maintained (Hiebert & Taylor, 2000).

Neither is it clear that instructional models that foster particular literacy proficiencies in high school are transferable to the elementary school. Two of the reading/language arts reform models were originally programs for high school or even college students: the National Writing Project and the Junior Great Books programs. Evidence that accommodations have been made to ensure that beginning readers and progressing readers in the middle grades receive the kind of guidance that develops the competencies of proficient reading, especially of informational text, has not been provided.

None of the reading/language arts models provide convincing data that implementation of these efforts will be the source of changes in reading profiles of students in economically challenged schools. While particular models may enhance particular dimensions of students' reading proficiency, transformation of students' reading achievement depends on instruction that attends to different tasks at different times. An intervention that emphasizes writing is unlikely to be effective in fostering the literacy fundamentals of beginning readers. Likewise, an intervention that emphasizes word recognition or fluency with simple, narrative texts is unlikely to foster strategies for comprehending and remembering complex, informational texts.

If there was no initial evidence, one might still expect that evidence would accumulate as projects were funded and moved forward. But evidence for the model programs' effectiveness has not been updated on the U.S. Department of Education's designated website or on the websites of the individual models. Further, the effects of the CSRD program as a whole have not been reported. The one evaluation that exists attends to the problems of implementing the models on a large-scale (Doherty, 2000). According to this evaluation, which was conducted by U.S. Department of Education personnel, comprehensive school reform involves more changes than any one model or strategy can address alone. As the program completes its third year of implementation, there is no evidence that gains have been realized in student reading achievement or even that the particular models have had prior success in sustaining gains in reading achievement.

While rhetoric for research-based reading instruction has increased since the implementation of the CSRD legislation, the burden of proof has been placed on schools rather than on the developers of the intervention models. According to the personnel associated with the identification of proven models (Buehler, 2001), schools need to show how their plan is research-based. That is, rather than expecting the model developers to provide data on the effectiveness of their practices, school personnel must piece together research studies to validate the particular set of practices that they have chosen. In high-poverty schools, a high percentage of teacher turnover takes place annually, if not mid-year (Darling-Hammond, 1997). It is unlikely that the programs such schools are able to

design and implement under these circumstances will be sufficient to meet the needs of the children they serve. Yet at the same time, findings from two recent blue-ribbon panels provide detail on the content of programs that can support higher levels of reading achievement. The needs of students, the consistency of at least a particular set of findings related to alphabetics, fluency, and comprehension (National Reading Panel, 2000; Snow et al., 1998), and the level of funding are all high. A question that is left unanswered is why the very program developers whose efforts are promoted by federal initiatives such as the CSRD are not asked to provide data that their models produce high levels of reading achievement. At the very least, model developers should be expected to show how their programs incorporate the practices that have been identified by the two recent panels of national reading experts as proven to support higher reading achievement.

Next Steps

Interpretations of American schoolchildren's literacy levels vary. However, even the most optimistic conclusion about these literacy levels needs to be accompanied by the caveat that what was sufficient for previous generations is not sufficient for the citizens of the digital age. Both of the reform efforts recounted here have lacked clarity about the underlying literacy processes or goals that the efforts are promoting. The two reform efforts also lack clarity regarding the instructional paths to these higher levels of literacy.

Such clarity is needed, and care should be taken that the goals of literacy and in-

structional paths promoted in textbook programs and intervention models converge. Similar goals and paths need to underlie the textbook programs to which states devote their reform efforts and the intervention models promoted by the national government. Suggestions for a shared vision of literacy proficiency and a means for supporting attainment of this proficiency by more students follow.

Clear Descriptions of Critical Reading Goals

There are three difficulties with the current descriptions of reading within the reading reform efforts of the state frameworks that mandate textbook changes and of the federal initiatives that support the model programs of the CSRD: (a) definitions of reading are vague—at the level of "reading as meaning," (b) definitions are generic rather than specific to different developmental levels, and (c) the grounding of definitions in tasks and texts has been limited. These difficulties persist despite the existence of numerous projects aimed at setting standards. A federally funded program has supported states in devising their own standards—a task that almost all 50 states have completed for reading/language arts. Further, professional organizations such as the International Reading Association and the National Council of Teachers of English (IRA/NCTE, 1996) have identified standards. But these efforts have uniformly shown little of the specificity, differentiation for different developmental levels, and elucidation of the relevant tasks and texts that administrators and teachers require to work toward common and critical goals (Stotsky, 2000.

The level of information that the public—including the nation's teachers and their students—is given about reading accomplishments is typified in the following statement about fourth graders' performances on the NAEP: "Overall, reading scale scores for the nation's fourth graders have not changed, with scores of 217 in both 1992 and 2000…. Thirty-two percent were at or above Proficient." (Phillips, 2001). From a report such as this one, it is impossible to tell what it is that fourth graders across the nation can read and what they cannot read. There needs to be some common ground in formative assessments; we cannot rely simply on the summative assessment of the NAEP. However, as a congressionally funded activity, the NAEP does have real potential to become more focused and responsive to the needs of children and their teachers.

To illustrate the nature of descriptions that are needed if teachers are to support their students in attaining proficient reading, an analysis of the sample text and questions from the 2000 fourth-grade NAEP assessment was conducted. An excerpt from the 2000 fourth-grade NAEP illustrates the texts that are used.

A Brick to Cuddle Up To

by Barbara Cole

> Imagine shivering on a cold winter's night. The tip of your nose tingles in the frosty air. Finally, you climb into bed and find the toasty treat you have been waiting for—your very own hot brick.

A first question to consider is whether the passage is an appropriate one for fourth graders. Since the demise of readability formulas in American reading education (Anderson, Hiebert, Scott, & Wilkinson, 1985), there have been no agreed-upon systems for designating text difficulty. Hiebert (in press) has developed a scheme for describing word recognition demands. The percentage of unique words that fall outside a particular curriculum of high-frequency words and phonetically decodable words is determined. The resulting figure is called the critical word factor (CWF)—the number of unique words per 100 running words that are beyond a particular curriculum. When this passage was assessed against a curriculum of the 1,000 most frequent words and all vowel patterns in single-syllable words plus simple morphological derivatives of these two groups, the critical word factor is revealed to be 5 unique words per 100. A level of 5 critical or difficult unique words per 100 running words of text is an acceptable number for the instructional to independent level of reading.

A second issue concerns the evidence that is gathered of students' comprehension of the passage. An elaborate scheme has been developed for the question types on the NAEP. Four types of questions are used: developing an initial understanding, developing an interpretation, developing a critical stance, and giving a personal reflection and response. When students' responses to 7 questions on the 2000 NAEP (Donahue et al., 2001) are considered, the issue appears less to be one of question type than of required response format. Many below-basic and basic students respond correctly to a critical

stance question. The difficulty for "basic" and "below-basic" students comes in writing responses to answers. When confronted with the first written response, which requires a one or two sentence answer, 38% of "basic" students provided satisfactory responses. But a portion of the "basic" and "below-basic" group continues to perform satisfactorily, even on the last item of the test (which requires an open-ended response).

A critical new direction in the demands of responses was taken on the 1992 NAEP. Following the authentic assessment movement of the early 1990s, open-ended responses became a prominent part of the 1992 NAEP. In the 2000 NAEP, 60% of the questions required written responses. In contrast to the amount of time that was spent categorizing question types, little time has been devoted to developing a scheme that distinguishes between the demands of different response formats. When the demands of the response modes are considered, strategies that teachers might take to ensure greater success for their students become apparent. Considerable periods of class time should be spent writing responses to texts and integrating background knowledge into these responses.

Rather than attributing low levels of interpretation to American schoolchildren, NAEP developers need to describe and justify what makes this text an exemplar of fourth-grade reading. They must also explain that it is entirely possible and even probable based on the findings of the NAEP special study (Pinnell et al., 1995) that children can read third-grade passages or fourth-grade passages that have been designated by other

means. Presentations of the NAEP data also need to indicate that it may well be that students are coming up short not in their reading but in their ability to write elaborated responses to questions.

Armed with information on the critical competencies, teachers could then proceed with their work. Suggestions regarding adapting the two existing reform efforts—textbooks and model instructional programs—follow.

Clear Descriptions of Instructional Programs For the Elementary Grades

The emphasis on proven models within the congressional legislation could be a positive impetus for school change, particularly if the characteristics of these models were reinforced in the teacher guidebooks that accompany textbook programs. Instructional programs can provide teachers with the tools to support students. In all likelihood, the focus provided by the model instructional programs will be better than the diffuseness of the instructional schemes promoted by textbook programs. The use of programs that bring in support that is external to the school is also likely to keep teachers focused and committed to working through snags as they arise. Such tools and a support network can be particularly critical in the current American context in which teachers in the most highly challenged schools over the next decade are likely to be underprepared (Darling-Hammond, 1997).

Rather than implementing a program that attends to one grade level or one aspect of literacy, reading reform efforts need to implement the features of effective reading

programs at all grade levels. For example, Adler and Fisher (2001) report that the reading programs in schools where potentially struggling readers are doing well share components such as a focus on student outcomes, multiple reading programs in every classroom, shared responsibility for student success across teachers, including specialists, strong leadership at school and classroom levels, and aveteran, knowledgeable staff, Such characteristics represent a level of specificity in the goals and instructional strategies of reading that most, if not all, of the current models lack. Only Success for All has addressed the elementary grades in a comprehensive fashion. In other cases, models are either focused on beginning reading (e.g., Reading Recovery and Early Intervention in Reading) or they are focused on upper levels (e.g., National Writing Project, Junior Great Books). For children who come to school without conventional literacy, an early intervention alone is unlikely to be sufficient. Attention needs to be paid to literacy throughout the developmental spectrum.

Evidence that underscores the need for attention to a developmental spectrum comes from research on the "leveling" effect of early interventions and the effects of generalizing strategies designed for older students to novices (Hiebert & Taylor, 2000). Unless instruction in subsequent grades builds on students' higher literacy levels as a result of an intervention, its effects will wane over time (Hiebert & Taylor, 2000). Further, efforts to transplant effective strategies with older students in responding to literature (Junior Great Books) or in writing extended narrative and informational texts (National

Writing Project) into early elementary classrooms have not produced evidence of the effectiveness of such strategies in producing independent beginning readers.

Data are, in fact, available on the characteristics of schools that have beaten the odds in changing the literacy profiles of students, particularly for schools in neighborhoods that are challenged by poverty and mobility. Characteristics that have been tied specifically to reading achievement are included in Table 5. Overall characteristics of school reform such as networks of support for teachers and communication with families are present in schools that are successful in teaching children to read. But to ensure that students' reading levels improve, these schools also need access to information about specific foci and tasks that are appropriate at particular developmental levels. The content of effective instruction over a child's first six years of school is clear from the recent report of the National Reading Panel (2000). In its description of these interventions, it should be emphasized that the Panel took for granted the presence of other fundamental dimensions of a literacy program—book reading, writing, and discussions of what has been read and written. But in terms of a number of critical aspects of reading programs, the Panel provided guidelines.

According to the National Reading Panel, interventions in kindergarten and the first part of grade 1 would emphasize phoneme and letter manipulation in small groups (National Reading Panel, 2000). By the second quarter of grade 1, interventions would emphasize consistent and common features of words as well as common words that have

irregular letter-sound patterns. These interventions provide children with occasions for reading books with words that have been the focus of lessons as well as numerous occasions for spelling and writing these and other words.

In second and third grades, interventions are focused on increasing students' fluency in reading. This can involve the support of an experienced reader who guides students in repeated reading and technological aids such as tape-recorders or computers. What is essential is that the intervention provides models of and occasions for reading with fluency and prosody in texts that are not overly difficult (i.e., approximately 5 unique words that may be challenging or rare within every 100 running words of text).

These interventions need to emphasize students' summaries of the information in the texts and their interpretations of ideas in the texts (National Reading Panel, 2000). Students need to be guided in consistent use of strategies that aid them in thinking about and remembering information in text. As students move into fourth grade, interventions provide little, if any, emphasis on fluency and increasingly emphasize processing and remembering text information. In addition to the modeling and application of consistent strategies to promote comprehension and remembering, interventions at this level should also support the development of new bodies of vocabulary. These bodies of vocabulary pertain to the elementary social studies, science, health, and humanities curricula. For example, key concepts in geology, geography, and political science become the foci of interventions in grades 4–6. In these sessions, students must grapple with the morphology

and etymology of words rather than simple definitions. The meaning of "geo," for example, would be integral to studying new words in geology and geography, as would "civis" to political science and history.

Conclusion

American reading reform efforts are massive in scope, but their efficacy is hindered by a lack of integration between state and federal efforts. The reform of textbooks by America's two largest states and funding for model programs by the federal government are not at cross-purposes. However, by not treating textbooks as model programs, federal efforts ignore a primary vehicle for reform. Textbook programs are already used extensively, a part of the school lives of millions of students and their teachers. If textbook programs were placed in the foreground rather than in the background within federal research and implementation efforts, the characteristics of these programs could be more closely scrutinized.

Just as the vehicles for reform are available, so too are guidelines for instruction. The report of the National Reading Panel (2000) described effective practices in the primary grades. The lack of integration within states and between states and the federal government, however, means that implementation of these findings on a large scale is unlikely to occur. For example, the Panel's conclusions regarding instruction that supports fluency have not filtered down to the model reform efforts of the federal government or the teachers' guides of the large states' textbook programs.

Without such integration, many new teachers who will teach in America's poorest schools over the next decades will struggle to bring their students to basic levels of literacy. Without integration of the vehicles for reform and the vision underlying this reform, literacy levels of those most in need in America's schools are unlikely to meet the demands of the digital age.

References

Adler, M. A., & Fisher, C. W. (2001). Early reading programs in high poverty schools: A case study of beating the odds. The Reading Teacher, 54, 616–619.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). Becoming a nation of readers: The report of the Commission on Reading. Champaign, IL: The Center for the Study of Reading.

Association of American Publishers. (2001). Industry statistics for the school division [On-line]. Available http://www.publishers.org/school/.

Barr, R., & Dreeben, R. (1983). How schools work. Chicago: University of Chicago Press.

Baumann, J. F., Hoffman, J. V., Duffy-Hester, A. M., & Ro, J. M. (2000). The first R yesterday and today: U.S. elementary reading instruction practices reported by teachers and administrators. Reading Research Quarterly, 35, 338–377.

Bond, G., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. Reading Research Quarterly, 2, 5–142.

Buehler, M. (2001, May 8). Personal communication with author.

Bush, G. W. (2001, January 29). No child left behind (Education Policy of the U.S. President) [On-line]. Available http://www.ed.gov/inits/nclb/index.html.

California English/Language Arts Committee. (1999). English-Language Arts content standards for California public schools (kindergarten through grade twelve). Sacramento: California Department of Education.

California English/Language Arts Committee. (1987). English-Language Arts framework for California public schools (kindergarten through grade twelve). Sacramento: California Department of Education.

Campbell, J. R., Donahue, P. L., Reese, C. M., & Phillips, G. W. (1996). NAEP 1994 reading report card for the nation and the states: Findings from the National Assessment of Educational Progress and trial state assessments. Washington, DC: National Center for Education Statistics.

Chall, J. S., & Squire, J. R. (1991). The publishing industry and textbooks. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), Handbook of reading research (Vol 2, pp. 120–146). Mahwah, NJ: Erlbaum.

Darling-Hammond, L. (1997). The quality of teaching matters most: What teachers know and can do makes the most difference in what children learn. Journal of Staff Development, 18, 38–44.

Doherty, K. M. (2000). Early implementation of the Comprehensive School Reform Demonstration (CSRD) program. Washington, DC: U.S. Department of Education.

Donahue, P. L., Finnegan, R. J., Lutkus, A. D., Allen, N. L., & Campbell, J. R. (2001). The nation's report card for reading: Fourth grade. Washington, DC: National

Center for Education Statistics.

Farr, R. C., Strickland, D. S., Beck, I. L., Abrahamson, R. F., Ada, A. F., Cullinan, B. E., McKeown, M., Roser, N., Smith, P., Wallis, J., Yokota, J., & Yopp, H. K. (2001). Collections: A Harcourt Reading/Language Arts program. Orlando, FL: Harcourt School Publishing.

Fisher, R., & Singleton, C. (2000). Symposium: The National Literacy Strategy. Journal of Research in Reading, 23, 242–244.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. Journal of Educational Psychology, 90, 37–55.

Herman, R., Aladjem, D., McMahon, P., Masem, E., Mulligan, I., O'Malley, A. S., Quinones, S., Reeve, A., & Woodruff, D. (1999). An educators' guide to schoolwide reform. Washington, DC: American Institutes for Research.

Hiebert, E.H. (2001, April). Pace and repetition: The forgotten variables in the design of beginning reading programs. Paper presented as part of the symposium Texts that support beginning reading acquisition: What we know and how we know it at the annual meeting of the American Educational Research Association, Seattle, WA.

Hiebert, E. H. (in press). Standards, assessment, and text difficulty. In A. E. Farstrup & S. J. Samuels (Eds.), What research has to say about reading instruction (3rd ed.). Newark, DE: International Reading Association.

Hiebert, E.H. (1994). Reading Recovery in the United States: What difference does it make to an age cohort? Educational Researcher, 23, 15-25.

Hiebert, E. H., & Taylor, B. M. (2000). Beginning reading instruction: Research on early interventions. In M. Kamil, P. Mosenthal, R. Barr, & P. D. Pearson (Eds.), Handbook of reading research (Vol. 3; pp. 455–482). Mahwah, NJ: Erlbaum.

Hiebert, E. H., Liu, G., Levin, L. Huxley, A., & Chung, K. (1995, November). First graders reading the new first-grade readers. Presented at the annual meeting of the National Reading Conference, New Orleans, LA.

Hoffman, J. V., McCarthey, S. J., Abbott, J., Christian, C., Corman, L., Dressman, M., Elliot, B., Matherne, D., & Stahle, D. (1994). So what's new in the "new" basals? A focus on first grade. Journal of Reading Behavior, 26, 47–73.

Hoffman, J., Roser, N., Patterson, E., Salas, R., & Pennington, J. (2000). Text leveling and little books in first-grade reading (CIERA Report #1-010). Ann Arbor, MI: Center for the Improvement of Early Reading Achievement.

International Reading Association (IRA) and National Council of Teachers of English (NCTE). (1996). Standards for the English Language Arts. Newark, DE: Authors.

Juel, C., & Roper/Schneider, D. (1985). The influence of basal readers on first grade reading. Reading Research Quarterly, 20 (2), 134–152.

Lohnes, P.R., & Gray, M.M. (1972). Intellectual development and the cooperative reading studies. Reading Research Quarterly, 8, 52-61.

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. Washington, DC: National Institute of Child Health and Human Development.

Northwest Regional Educational Laboratory. (2001). Components of comprehensive school reform programs [On-line]. Available http://www.nwrel.org.

Phillips, G. W. (2001, April 6). The release of the National Assessment of Educational Progress (NAEP) Fourth-grade Reading 2000 [On-line}. Available http://www.nces.ed.gov/commissioner/remarks2001/.

Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B., & Beatty, A. S. (1995). Listening to children read aloud: Data from NAEP's integrated reading performance record (IRPR) at Grade 4. Washington, DC: OERI /U.S. Department of Education & ETS.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). Preventing reading difficulties in young children. Washington, DC: National Academy Press.

Stotsky, S. (2000). The state of literary study in national and state English language arts standards: Why it matters and what can be done about it. In S. Stotsky (Ed.), What's at stake in the K–12 standards wars: A primer for educational policy makers (pp. 237–257). New York: Peter Lang.

Texas Education Agency. (1990). Proclamation of the State Board of Education advertising for bids on textbooks. Austin, TX: Author.

Texas Education Agency. (1997). Proclamation of the State Board of Education advertising for bids on textbooks. Austin, TX: Author.

U.S. Department of Education (2001). www.ed.gov

Table 1. Reading/Language Arts Reform Models

	Developer Data	External Reviewer Data
•Breakthrough to Literacy (K–2)*	No published evaluations are available.	No studies are claimed.
•Carbo Reading Styles Program (K–8)	Evaluations are claimed but are not available on www or in published journals.	1 dissertation, 1 description in practitioner journal, and 1 ERIC document.
CELL/ExLL (PreK–6)	Evaluations are claimed but are not available on www or in published journals.	No studies are claimed.
CORE (K–8)	No published evaluations are available.	1 evaluation is claimed but is not available on www or in published journals.
Early Intervention in Reading (K–4)	2 descriptions in practitioner journal exist (Reading Teacher).	1 evaluation is claimed but is not available on www or in published journals.
Exemplary Center for Reading Instruction (K–12)	1 evaluation is claimed but is not available on www or in published journals.	3 descriptive articles in practitioner journals exist.
•First Steps (K–10)	Evaluations are claimed but are not available on www or in published journals.	3 evaluations by Australian Council of Educational Research exist (available on www).
Junior Great Books (K–12)	1 evaluation is claimed but is not available on www or in published journals.	1 dissertation on 5th graders exists; 2 evaluations are claimed but are not available on www or in published journals.
Literacy Collaborative (K–2)	Evaluations are claimed but are not available on www or in published journals.	No studies are claimed.
•National Writing Project (K–16)	1 evaluation is claimed but is not available on www or in published journals.	No studies are claimed.

•Reading Recovery (1)	1 evaluation is claimed but is not available on www or in published journals; 1 book chapter.	1 study in an international journal and 2 reports (one available through US Dept. of Ed.) exist; external reviews in archival journals by, e.g., Hiebert, 1994 were not cited.

*Indicates grade level