Routledge
Taylor & Francis Group

THE EFFECTS OF TEXT DIFFICULTY ON SECOND GRADERS' FLUENCY DEVELOPMENT

ELFRIEDA H. HIEBERT

University of California, Berkeley, California, USA

In this study, two groups of second graders participated in Fluency-Oriented Reading Instruction (FORI) but their repeated reading experiences used different kinds of texts. One group—the Literature group—read texts from the district's literature-based, basal reading program. The second group—the Content group—read from a set of science and social studies texts that were written to have few rare, multisyllabic, single-appearing words. Control-group students read from the district's literature-based program.

Both intervention groups made greater gains in reading rate than Control group students and Content students made greater gains in reading rate than Literature students. Content and Literature groups outperformed Control students on the comprehension measure but did not perform significantly differently from one another. The gains made by the Content classrooms were made in approximately half the amount of time allocated to reading instruction as by Literature classrooms.

Example 1—*The man who kept house*: Once upon a time there was a farmer who believed that his work was too hard. Each evening when he returned from his fields, he would ask his good wife what she did all day. "I kept house," she always answered. The farmer would say, "Oh, your work is easy." (Aaron et al., 1983).
Example 2—*Annie's Gifts*: Once there lived a family that loved music. Every morning the children, Lee, Patty, and Annie, turned on some music. The floors trembled as they stomped their feet to the loud bass beat. Soon they were moving down the street to catch the school bus. (Medearis, 1997, in Afflerbach et al., 2000)

The stories from which these two examples are excerpted share more than the classic beginning of evoking a time in the past. Both stories come from the same place in the second-grade textbooks of the basal program of the same publisher. Both passages are narratives, and the plot of each story revolves around a topic

familiar to children—the dynamics of home life. There are also differences between these two programs, published 14 years apart, such as the types of words in the two texts. All of the unique or different words in the excerpt of the earlier era are among the 1,000 most-frequent words (Zeno, Ivens, Millard, & Duvvuri, 1995, or simple derivatives of those words, e.g., *farmer*). Within the excerpt from the more recent program, 80% of unique words are among the 1,000 most-frequent words and the other 20% are less frequently used words. This difference is represented in the use of common words, rather than less-frequent words, to describe particular actions in the first text, such as *hard* rather than *strenuous*. In contrast, less frequent words rather than common ones are used to describe actions, such as the use of *trembled* rather than *shaking* in the current text.

Bruce (1984) and Beck, MeKeown, Omanson, and Pople (1984) have suggested that the presence of words with greater connotative specificity can contribute to the engagingness and even comprehensibility of texts for students. At the same time, infrequent vocabulary can put demands on the reading abilities of developing and struggling readers. In accurate, automatic reading, proficient readers turn their attention to unknown words when they occur. However, if too many words in a text require such attention, comprehension suffers. When developing readers encounter infrequent words that they have likely not encountered in text before, they may need to stop and figure them out. When infrequent words are multisyllabic (e.g., *music, trembled*) or have uncommon letter-sound relations (e.g., *bass*), developing readers may need to devote considerable attention to unknown words.

The thesis underlying the current work is that the number of rare, multisyllabic words in texts is a factor in the fluency development of beginning and struggling readers. As a result of the National Reading Panel's (2000) report and the use of that report in designing the No Child Left Behind (NCLB) agenda (U.S. Congress, 2001), the current attention to fluency instruction has emphasized the instructional techniques of guided and repeated reading but has ignored the dimension of text. Since the period when techniques such as repeated reading (Samuels, 1979) and neurological impress reading (Heckleman, 1966) were proposed to support fluency, texts have moved from the features that are represented in Example 1 to those in Example 2. After the middle of

first grade, current textbook programs consist of trade book selections rather than texts that are controlled for either decodability or high-frequency words (Foorman, Francis, Davidson, Harm, & Griffin, 2004). The benefits of literature in elementary classrooms are many. For students who are not fluent with particular corpora of words, however, an almost exclusive diet of texts with high percentages of rare, multisyllabic words—especially ones that occur a single time—may not provide them with the opportunity to become fluent with those words that account for large portions of written language.

This paper addresses the nature of texts that promote fluency in two ways. First, theory and research are reviewed to demonstrate why texts with particular features are hypothesized to support fluency, while texts with other features may detract from fluency. The second aim of this paper is to report on the results of a study where second graders' repeated reading experience was with either literary texts or content texts. True to the genre, the literary texts had a high percentage of rare, multisyllabic words. Similar to the description of content texts (Duke & Kays, 1999), the rare words in the content texts were fewer and, when they appeared, were repeated several times.

The Role of Text in Fluency

Samuels (1979) and other scholars theorizing about automaticity did not address text directly, but the features of texts that influence fluency can be extrapolated from the theory. Automaticity in reading a text means that readers' attention is directed to generating meaning. Automaticity does not preclude readers from focusing on particular words to determine their meaning. However, automaticity does mean that this attention to individual words is not taxing enough to detract from meaning-making. The number of unknown words that readers can focus on within a text while continuing to focus on the meaning-making process would be expected to differ as a function of the corpus of words automatically recognized by readers and the nature of unknown words in a text. Based on Betts's (1946) hypotheses, reading educators have considered a ratio of 1:10 as the maximum number of unknown to known words that readers can encounter and continue to make meaning of text (Rasinski, 1999).

TABLE 1 Number of Unique and Root Words per One Million Word Corpus

Number of Appearances in 1 million words	Number of Unique Words (cumulative)[1]	Number of Root Words[2] (cumulative)	Proportion of Total Words: One million word corpus (cumulative)	Word Zones (Hiebert, in press)
30,000	1	1	.07	0
10,000	8	8	.21	
3,000	38	37	.37	
1,000	107	99	.48	
300	310	258	.57	1
100	930	637	.67	2
30	2606	1477	.74	3
10	5586	2710	.79	4
3	11240	5497	.82	5
1	19468	8459	.87	
.99 and fewer	154941	58584	1.0	6

[1]Based on Zeno et al. (1995).

[2]Through 10 appearances in 1 million words, the number of root words comes from a word-by-word analysis (Hiebert, in press). For words with 3 or fewer appearances per 1 million words, the number of root words has been estimated based on Nagy and Anderson's (1984) calculations.

Since the body of words in school texts contains as many as 156,000 different words (Zeno et al., 1995), the task of becoming automatic with 90% of the words in texts appears daunting, if not impossible. However, the typical distribution of these 156,000 words in school texts, as illustrated in Table 1, provides an indication of what words need to be recognized automatically and what words fall into the 10% category.

This distribution indicates that a small percentage of all unique words accounts for a substantial portion of the total words in text. Further, as Nagy and Anderson (1984) have shown, these high-frequency words represent a "family" or group of words that share simple derivates—inflected endings, comparisons e.g., (-er, -est), and certain affixes (e.g., ly, pre). A group of 930 words accounts for 67% of all of the words in texts read through college. When simple derivates of these words are included, a significant portion of the next group of 1,676 words (i.e., those with appearances of 30 per 1 million words) is accounted for as well (Hiebert, in press). The remaining words in written English do not occur with the frequency of this group of words. Presumably, readers apply

word- and text-level knowledge to recognize and understand the words that occur less frequently. However, if all words, including the 930 most frequent and their derivatives, are treated as words to be figured out, it is doubtful that a text will be comprehended adequately.

The nature of the progression that readers follow in becoming automatic with this group of 930 words and their derivatives has yet to be documented in the research literature. Presumably, particular words from within this group are the basis for students' initial automaticity and, based on this foundation, their automaticity extends to other, less frequent words. While at this point the research literature is not specific as to how quickly the majority of students become automatic with different corpora of words, several lines of research support the hypothesis that practice with texts that have high percentages of rare words will be less useful for beginning and struggling readers than texts with low percentages of rare words.

In the first group of studies, changes in fluency have been examined as a function of the degree to which practice and criterion texts share words. While the degree of rareness of words has not been reported in these studies, texts with the highest percentage of shared words have been found to produce the greatest gains in reading speed. In examining the relative contribution of either text features or the technique of repeated reading to student achievement, Rashotte and Torgesen (1985) modified texts to create one set in which the overlap of vocabulary across stories was low and a second set where overlap was high. The condition with the highest percentage of shared words yielded the greatest gains in reading speed. When the percentage was not great, repeated reading did not make a difference.

In another study, Faulkner and Levy (1994) found that texts that share content and words produce gains in the speed and accuracy of both good and poor readers. However, the speed and accuracy of poor readers improved even when texts shared words but pertained to different content. Faulkner and Levy argued that this finding of shared vocabulary across texts explained results of early studies on repeated reading. For example in Dowhower's (1987) study, there was a 77% overlap between words in the practice and final texts.

The second group of studies that supports the hypothesis of fluency with texts that have low percentages of rare words comes

from the meta-analysis on which the NRP based its conclusions about fluency. Hiebert and Fisher (2002) categorized the 13 texts in the meta-analysis and the 46 studies in the extended database into four text types: high-interest/low-vocabulary texts (HI/LV; modeled after the I Can Read series to which Minarik and Dr. Seuss contributed), skill-builders such as the Barnell-Loft series, pre-1990 basal textbooks that had texts controlled by readability formulas, and post-1990 basal textbooks that contained literature from trade books. Around 80% of the studies in both levels of the NRP review used three types of texts with controlled vocabulary: pre-1990 basal texts, skill-builder texts, and high-interest/low-vocabulary (HI/LV) books. Of the three studies in the meta-analysis that used literature to promote fluency, only one reported on fluency rates (Eldredge, Reutzel, & Hollingsworth, 1996) and it found no differences in reading rate between the two groups who read literature in either a shared book experience (with repeated reading) or round robin reading.

An analysis of samples of the three types of controlled texts at the third-grade level showed that percentages of unique words in the pre-1990 basals, skill-builders, and HI/LV that were among the 1,000 most frequent words ranged from 68–75%. In an analysis of samples from a 2000 mainstream textbook program (i.e., a post-1990 basal that contained literature from trade books), 60% of the words were of this type. While the three types of controlled texts had an average of 15% of the words that were beyond the 1,000 most frequent words and multisyllabic, the percentage in the 2000 textbook was 28%. Further, three-quarters of these rare, multisyllabic words in the 2000 textbook appeared a single time. What can be concluded from these analyses is that the findings on repeated and guided oral reading were achieved with texts that had a substantially lower percentage of rare, multisyllabic, single-appearing words than is the case with the current literature-based texts of instruction.

Hiebert (2003a) examined the effects of repeated reading of texts with different percentages of rare, multisyllabic words on students' reading rate, prosody, and comprehension. One group of students read repeatedly from literary trade selections where approximately 20% of the unique words were rare and multisyllabic. The repeated reading of the other group occurred with a set of content selections where 2% of the unique words were rare and

multisyllabic. The adjusted means for the two groups approached significance, with the students in the Content repeated reading group gaining an average of 6 words Words Correct Per Minute (WCPM) more than the Literature repeated reading group over the 10-week intervention.

The Current Study

This study was a replication and extension of the previous study (Hiebert, 2003a). As in the previous study, the interest lay in classroom applications of fluency practices. As Kuhn (this volume) has noted, a majority of the studies on which the NRP (2000) based their findings were conducted in experimental contexts where students were either in tutorial or small groups of three or four. Further, tutors or teachers in these interventions have typically been members of experimental teams and many of the studies were conducted for short durations. For example, the studies of Rashotte and Torgesen (1985) and Faulkner and Levy (1994) consisted of a handful of sessions. If instructional applications of fluency research are to be as useful, issues related to optimal implementation in classrooms need to be addressed. While the NRP (2000) was able to designate optimal lengths of instruction for phonemic awareness, the research literature on fluency was neither extensive nor robust enough to permit such recommendations. With the aim of extending the data base, this study used the same texts that were used in the previous study but doubled the length of the intervention.

Many additional questions are yet unanswered about instructional practices that support fluency. Central among these are questions related to the amount of time that it takes students, especially those who are far below benchmarks on state and national assessments, to become fluent readers. One of the responses by policy-makers to perceived achievement gaps has been to mandate increased allocations of time to reading instruction (see, e.g., California State Board of Education, 2002). While academic learning time (a variable that includes the nature of the task, level of student engagement, and amount of time spent on pertinent tasks) predicts reading achievement (Berliner & Fisher, 1985), allocation of instructional time in itself is not highly predictive of student success. However, the amount that students read does appear to

influence reading achievement. Amount of self-reported reading in the classroom and for homework has been found to have a consistently, positive relationship with reading performance on the National Assessment of Educational Progress (NAEP; Donahue, Finnegan, Lutkus, Allen, & Campbell, 2001). The average scale score on the NAEP of students who reported reading 5 pages or less daily fell in the below-basic level, while students who reported reading 6 through 10 pages fell into the basic level or above. In analyzing the data on amount of time spent reading inside and outside of school among Maryland fourth graders and their reading performances on the 1994 NAEP, Guthrie, Schafer, and Huang (2001) described these reported amounts of text as opportunity to read. The current study's focus was on an aspect of opportunity to read: opportunity to read texts repeatedly. This study included a descriptive component of the amount of time that students spent in reading and rereading of texts in their classrooms.

As in the previous study, second grade was targeted because it is the period when students' reading rates increase substantially. Regardless of their achievement level, second graders gain an average of 40 words per minute in reading speed (Hasbrouck & Tindal, 1992). By contrast, the average gain in third grade is 20 words per minute, in fourth grade 3 words, and in fifth grade, 9 words.

Both groups of students in the current study were in classrooms where the same repeated reading procedures were used: Fluency-Oriented Reading Instruction (FORI) (Stahl, Heubach, & Cramond, 1997). However, what the students read during the repeated and monitored readings differed. The Literature group's repeated reading experience occurred with the district's literature-based textbook program. The Content group read texts from the literature-based textbook program for part of their reading period. However, their repeated reading experiences occurred with a set of science and social studies texts that had few rare, multisyllabic words that appeared a single time.

Method

Participants

Students in the two intervention groups as well as the Control group came from the same school district in a mid-sized, suburban

school district in a southeastern American state. The intervention occurred over a 20-week period from mid-October through mid-March of the school year.

Because of state sanctions and policies, random assignment of teachers to intervention and Control conditions was not possible. Teachers' participation in a project that involved changes in classroom practice was seen as part of the school improvement effort in which one of the schools was engaged. A substantial change in classroom practice, such as the implementation of an intervention, required that grade-level teams work together. Consequently, the three treatments were assigned randomly to schools. The repeated reading with Literature intervention (Literature RR) was conducted in one school (School A) with 4 second-grade classrooms. In these 4 classrooms, data were gathered on 43 students. The repeated reading with Content text (Content RR) was implemented in one school (School B) with three classrooms, where data were gathered on 45 students. The Control condition involved one school with six classrooms (School C). Data were gathered on 27 students from these classrooms. This school maintained its typical literature-based reading instruction.

Instructional Procedures

In one important feature, the reading instruction in all three types of classrooms (Control, Literature RR, and Content RR) was similar: The same basal textbook program was used in a whole-class format in the district. While the textbook program for the main part of the reading program was the same for all three groups, the use of repeated reading differentiated the two intervention groups from the control group and the type of text for repeated reading differentiated the two intervention groups.

INTERVENTION PROCEDURES

Both intervention groups followed the basic procedures of FORI (Stahl et al., 1997). The FORI instructional cycle provides students with multiple experiences with a text. The cycle includes: (a) teacher initiation of a text with modeling of fluent reading, comprehension, and review of key vocabulary; (b) partner rereading of the text; (c) teacher-led choral and/or echoic reading; and (d) extension activities that focus on comprehension. Home

reading that includes reading aloud of focus texts, particularly for struggling students, and free-choice reading (both in the class-room and at home) are additional components of the procedure, although not the focus of this study.

A text from the basal anthology provided the focus for an instructional cycle over a week for the students in the Literature RR group. In the Content RR group, the text from the basal an-thology was presented in a teacher-guided format over one to two lessons. Repeated reading experiences occurred with the content texts. Each content text was read at least three times in a teacher-led, partner or choral reading, and individual context. Discussions that focused on comprehension and vocabulary were part of each round of repeated reading of a Content text.

PROFESSIONAL DEVELOPMENT

Teachers in both intervention groups participated in a two-day professional development seminar prior to the initiation of the study. The repeated reading procedures were demonstrated by presenters as well as on videotaped lessons. The first reading of a text emphasized discussion and understanding of the text. Once the text has been read and discussed, teachers lead students through the texts in echo reading. Partner reading with the text followed. The technique of choral reading for additional exposure to a text was also demonstrated.

Teachers were provided with sample lessons as well as with letters to send to parents regarding the home reading program. Other topics of discussion included ways to create dyads for part-ner reading. In summary, the procedures for the four dimensions of the FORI program—guided reading with comprehension and vocabulary instruction; repeated reading of the focus text in echo, partner, and choral contexts; classroom choice reading; and home reading—were presented and reviewed during the two-day work-shop. The teachers in the Content RR intervention were advised that they should present the texts in the basal program to their stu-dents but that repeated reading should be done with the content texts that they were provided.

OPPORTUNITY TO READ TEXT REPEATEDLY

Four measures of opportunity to read texts repeatedly were identified from scholarship on academic learning time (Berliner

TABLE 2 Anchors for Scoring "Opportunity to Read" Categories

Measure	Score of 6	Score of 3	Score of 1
Opportunity for Reading Instruction	150 minutes +	90 minutes daily	30 minutes daily
Opportunity for Repeated Reading	30 minutes daily	15 minutes daily	0
Opportunity for Feedback & Modeling	Teacher monitors individuals, small groups during repeated reading sessions	Repeated reading is done primarily with partners	Students are asked to reread texts by themselves without task monitoring
Opportunity to Read "new" words	Text read repeatedly averages 120 different words daily (i.e., approximately 600 words weekly)	Text read repeatedly averages 60 different words daily (i.e., 360 words weekly)	Very little or no text read repeatedly

& Fisher, 1985): (a) length of a typical reading period and the consistency of these reading periods over the intervention period; (b) length of time devoted to repeated reading during the course of these reading periods; (c) opportunities for feedback and modeling of proficient reading; and (d) the amount of text that was read repeatedly. For each dimension, a 6-point scale was used with 6 indicating high levels of a dimension and 1 indicating low levels. Descriptions of the four scales appear in Table 2.

No single dimension is necessarily dependent on another rating. For example, it is possible that the total amount of time devoted to reading could be limited but all of that time could be devoted to repeated reading. Further, long periods could be devoted to repeated reading with partners of equivalent reading proficiency, thus limiting opportunity to receive feedback and modeling from a proficient reader (at least for dyads of low-performing students).

These four measures were established on a class level rather than for individual students and only in the intervention classrooms. In that repeated reading is an activity that has been

advocated in professional development seminars and materials, individual teachers in School C may have used repeated reading in their programs. However, repeated reading had not been a focus of that school's professional development. Since repeated reading was the component that had been added to the programs of School A (Literature RR intervention) and School B (Content RR Intervention), observations were limited to those two schools.

Assignment of a class-wide score was established for the four measures in the following manner. Formal observations were conducted in each classroom four times over the intervention, with one observation occurring on one of the four days of the week—Monday through Thursday—in each classroom. One research assistant, a doctoral student in reading education, conducted the observations. Two additional graduate assistants visited each classroom at least once to ensure interrater reliability in the rating of activities. During the formal observation, the observers used a rating instrument to record the activities of the lesson, how much time was spent on reading connected text, and the frequency of use of strategies (e.g., echo reading, partner reading). The graduate assistant used the rating form to write a summary of the observation period. The observations on this summary were verified by weekly visits by the same graduate assistant to the classrooms. During these informal visits, the graduate assistant met briefly with each teacher, addressing questions about procedures, student progress, and scheduling.

The summary forms were used to establish the four measures of opportunity to read. Two raters read through the transcripts and assigned scores for each of the four dimensions for each of the seven intervention classrooms. The two raters agreed on 75% of the assignments of scores.

The Texts

The literature-based, basal reading program was LiteratureWorks[1] (Pearson et al., 1998). The Content texts were *QuickReads*[1] (Hiebert, 2003b). Both sets of texts were analyzed with a Hyper-Card application that summarizes the number of unique words

---

[1]The principal investigator of this paper is an author on both programs that were used in the instructional intervention—the literature-based basal reading program and the content reading program.

within a text and the high-frequency ratings and decodability patterns of unique words. High-frequency ratings are based on Hiebert's (in press) six word zones that appear in Table 1. Zones one through four correspond to the curriculum of grades one through four. Data for both programs appear in Table 3 and an example of each type of text appears in Table 4.

While the two programs had close to the same number of total words, the Literature Texts had twice as many unique words per 100 running words as the Content Texts: 14 to 7. Of these unique words, less than 50% of the Literature Texts and over 75% of the Content Texts were in zones 0–2. Further, 17% of the unique words in the Literature Texts were multisyllabic, rare, and single-appearing words, while any rare, multisyllabic word in the Content Texts was repeated.

Assessments

All students were individually assessed at the beginning and end of the intervention period on two texts. These texts were each 100 words long and dealt with the topic of defense mechanisms of a different group of animals: (a) Text 1: animals that look alike (coral and king snakes), and (b) Text 2: animals that use sharp bones or fins (lion fish).

The comparability of the text difficulty of the two passages was examined on three indices: (a) average decodability based on an 8-point scale where 1 was assigned to words with a C-V pattern and 8 assigned to multisyllabic words; (b) average frequency ratings where 1 was assigned to the 100 most-frequent words (Zeno et al., 1995) and 11 to words beyond the 1,000 most-frequent words; and (c) the number of unique or different words. According to these indices, the texts were comparable. Decodability ratings were 4.4 and 4.8, where category 4 represents words with "a silent e" and category 5, vowel digraphs (e.g., ee in keep). Frequency ratings were 3.0 and 3.3, meaning that words typically were among those ranked 300–330 which corresponds to zones 1 and 2 in the distribution in Table 1. Both texts had the same number of unique words—52.

An analysis was also conducted to see if the assessment passages favored either of the sets of texts that were used for repeated reading in the two interventions. Of the 81 unique words in the

TABLE 3 Characteristics of the Two Texts Used for Repeated Reading

	Total Words (18 passages)	Total Words per Passage	New, Unique Words per 100	Percentage of Unique Words in Word Zones					Percentage of Unique words: Multisyllabic, Singletons (< = 10)
				> = 100 (Zones 0–2)	30 (Zone 3)	10 (Zone 4)	1 (Zone 5)	< = .99 (Zone 6)	
Literature Text (Pearson et al., 1998)	11,153	620	14	46	16	13	16	9	17
Content Text (Hiebert, 2003)	9,263	515	7	77	9	0	14	0	0

TABLE 4 Examples of Texts Repeatedly Read by Two Intervention Groups

Text type	Excerpt
Literature text	Uncle McAllister lived in Scotland. Every year he sent Louis a birthday gift for his nature collection. "This is the best one yet!" cried Louis. The next day he took his entire collection to school for show-and-tell. "Class, this is a tadpole," said Mrs. Shelbert. She asked Louis to bring it back. (Kellogg, 1977, in Pearson et al., 1998)
Content text	George Washington Carver was a scientist who knew about plants. He learned that soil wears out when farmers grow the same crop every year. When soil wears out, crops are poor. George Washington Carver showed farmers how to grow one crop in one year. Then they would grow a different crop in the next year. (Hiebert, 2003b)

two assessment texts, 69 words appeared an average of 39 times in Literature texts and 59 appeared an average of 28 times in the Content texts. If both groups repeatedly read the designated texts for their treatment, the Literature group would have had approximately 40% more exposure to the words on the assessments than the Content group.

The assessment administrators were all graduate students in reading education who had been trained in administration of clinical assessments, including the Gray Oral Reading Test (GORT; Wiederholt & Bryant, 2001), but were not involved in the design or implementation of the intervention. They used the administration procedures of the GORT, counting substitutions, insertions, and omissions as errors and discontinuing the task when students made 10 consecutive oral reading errors. Assessment administrators wrote down beginning and ending times of students' oral readings.

When students finished reading passages, the assessment administrator asked four comprehension questions. The first two questions pertained to literal aspects of the text and the final two pertained to interpretations of the text. Answers to each of the four questions were scored on a three-point scale where 0 was assigned to incorrect, non-topic related responses, 1 to a basic understanding of the text but incomplete, and 2 to accurate and complete responses. Comprehension scores ranged from 0–8. Two research assistants who were not involved in the study implementation or

design rated a randomly selected sample of the assessments until an inter-rater agreement of 90% was achieved. One rater then scored half of the protocols. The second rater independently rated 10% of this group of protocols to ensure that 90% agreement was maintained. With the agreement level confirmed, the remaining protocols were scored.

Assessment administrators also rated students' prosody on the 4-point scale used by Pinnell et al. (1995). Ratings of 4 were given to oral reading renditions where students read in large, meaningful phrase groups; 3 to three to four word phrase groups; 2 to two-word phrases with some three- or four-word groupings; and 1 to word-by-word reading. Prior to their administration of the assessments, administrators achieved a 97% inter-rater agreement level on coding prosody, using audiotapes of students from another project.

Results

An analysis of the effects of the interventions on student performance is presented first. In examining the effects of the interventions, the distribution of student performances was considered as well as comparisons of group means. Additional analyses explored opportunities to read repeatedly and the relationship of these opportunities to student performances.

Student Performances as a Function of Treatment

The standards-based movement has directed policy and practice to the influence of instructional treatments on students of differing entry levels, not simply group effects (U.S. Congress, 2001). The need for this emphasis was underscored when the initial profiles of the three groups were examined. An examination of the entry distributions of the three groups showed that the two intervention groups had a substantially greater number of students in the bottom quartile according to the fluency norms of Hasbrouck and Tindal (1992) relative to the Control group. The distributions of all three groups are summarized in Table 5. In both repeated reading groups, 49% of the students fell into the bottom quartile (Hasbrouck & Tindal, 1992). By contrast, 7% of the control group had entry fluency levels that placed them in this group. Subsequent

TABLE 5 Pretest and Gain Scores and Size of Quartile Groups in Three Conditions

	Q1			Q2			Q3			Q4		
	Pre	Gain	*n*	Pre	Gain	*n*	Pre	Gain	*n*	Pre	Gain	*n*
Content RR	110.7	17.6	3	67.1	23.0	9	43.0	29.8	11	2.4	41.6	22
Literature RR	100.6	14.7	3	68.6	29.7	10	35.2	18.5	9	4.7	37.4	21
Control	121.1	21.9	7	66.8	27.6	12	37.3	19.7	6	22.0	16.8	2

analyses of group differences and of group profiles were selected to accommodate these differences between the control and two treatment groups (Glass & Hopkins, 1995; Ravid, 1994).

EXAMINATION OF GROUP DIFFERENCES

Group sizes, means and standards deviations for pretest and posttest scores, and gain scores of students in the three groups—Literature RR, Content RR, and Control—and on the three measures—fluency, comprehension, and prosody—are reported in Table 6.

TABLE 6 Means (and Standard Deviations) for Control and Treatment Groups: Pretest, Posttest, and Difference Scores

Measure	Type of Score	Content Rereading *n* = 45	Literature Rereading *n* = 43	Control *n* = 27
Fluency				
	Pretest	32.5 (35.0)	32.6 (33.1)	71.0 (35.7)
	Posttest	65.9 (32.4)	62.6 (34.4)	94.6 (39.3)
	Difference	33.4 (21.0)	30.0 (21.9)	23.6 (16.0)
Comprehension				
	Pretest	2.1 (2.1)	1.9 (1.9)	4.0 (1.8)
	Posttest	3.0 (2.0)	3.0 (1.9)	4.4 (1.2)
	Difference	.08 (2.1)	1.1 (2.0)	0.4 (1.6)
Prosody				
	Pretest	1.2 (1.3)	1.6 (1.5)	2.5 (1.1)
	Posttest	1.6 (1.1)	2.0 (1.1)	2.7 (0.6)
	Difference	0.4 (0.9)	0.4 (1.2)	0.2 (0.8)

Note: These are unadjusted means. Standard deviations are shown in parentheses.

ANCOVAs were run on the three dependent variables, using the fall entry scores as covariates. Three one-way ANVOCAs were run: The first two compared the interventions with the control ground; the third compared the Literature RR group with the Content RR group. On prosody and comprehension scores, no significant differences were found. However, on the fluency variable, where the intervention groups had similar entry-level scores ($X = 32.5$ WCPM for the Content RR group and $X = 32.6$ for the Literature RR group), the students in the Content RR group significantly outperformed the Control group, $F(1, 70) = 4.4$, $p < .04$. An examination of Table 6 indicates that the Content RR group's mean gain was approximately 10 WCPM higher than that of the Control group. The gain of the Literature RR group on the fluency measure was greater than that of the Control group; however, this difference was not statistically significant, $F(1, 68) = 1.76$, $p < .2$. The difference between the Content RR and Literature RR groups on the fluency measure was not statistically significant.

STUDENT PERFORMANCES AS A FUNCTION OF ENTRY LEVEL

Pretest assessment scores (fall) were used to place students in quartile groups based on Hasbrouck and Tindal's (1992) norms. Students who scored up to and including 22 words correct per minute (WCPM) were assigned to the fourth quartile, and students scoring from 23 through 52, 53 through 81, and 82 or more WCPM were assigned to the third, second and first quartiles respectively. The numbers of students in these quartile groups varied considerably across the three treatment conditions (see Table 5). Because of unequal cell sizes, data on performances of quartile groups are presented descriptively.

The scores of the Content RR students were considerably higher in the third and fourth quartiles than those of students in either the Literature RR or the Control groups (see Table 5). Growth in the first quartile was comparable for all students regardless of condition. The second quartile is of particular interest in that it was the only quartile where the cell sizes for the three conditions were similar and where performances were comparable at the beginning of the study. In the second quartile, the Content RR group had lower gain scores than both the Literature RR and Control conditions.

Opportunity for Repeated Reading

The final analyses pertained to opportunity for repeated reading. Data are presented in Table 7 for the four indices that were used to measure opportunity for reading: (a) overall opportunity for reading; (b) specific opportunity for repeated reading; (c) opportunity to have feedback from a proficient model; and (d) amount of text covered. Table 7 also includes fluency means for the seven intervention classes. No data on opportunity for repeated reading were available for the Control group.

On the first measure–overall opportunity for reading—the data in Table 7 indicate that different amounts of time were allocated to reading instruction in the two intervention schools. The treatments were confounded by school policies regarding allocation of time to reading/language arts. As part of a federal reading improvement grant that the school with the Literature

TABLE 7 Opportunity to Read and Performances For Classrooms

	Fall: WCPM	Spring: WCPM	Class Gain	Opportunities to read measures			
				1	2	3	4
Literature RR intervention classes							
Literature RR 1	29	62	33	6	5	4	3
(*n* = 13)							
Literature RR 2	46	70	26	6	5	5	2
(*n* = 11)							
Literature RR 3	37	68	31	6	5	4	3
(*n* = 8)							
Literature RR 4	21	52	31	6	5	4	3
(*n* = 11)							
Content RR intervention classes							
Content RR 1	20	56	36	3	3	4	4
(*n* = 14)							
Content RR 2	35	73	38	3	3	4	4
(*n* = 13)							
Content RR 3	40	69	29	2	2	2	2
(*n* = 18)							

1 = overall opportunity to read.
2 = opportunity to read repeatedly.
3 = opportunity for feedback and modeling.
4 = amount of text read repeatedly.

RR had received, second-grade teachers were required to devote three hours to reading/language arts daily. All of the second-grade teachers complied with this mandate. The reading block was substantially less in the school with the Content RR treatment. The school with the Content RR treatment operated under the district guidelines that called for approximately 75 minutes daily for reading instruction.

The criterion for the highest level of opportunity for repeated reading was 30 minutes daily. This amount of time would have been possible in the shorter, 75-minute daily reading period in the Content RR classrooms of School B. However, the Content RR teachers provided approximately 15-minute sessions for their students to read repeatedly. Further, unlike the Literature RR treatment where teachers applied the treatment quite consistently, one of the teachers in the Content RR chose to provide little time for repeated reading. In general, students in the Literature RR classrooms had more opportunity for repeated reading than students in Content RR classrooms. However, the gains in WCPM for Literature RR students were not commensurately higher than that of students in the Content RR classes. Despite almost 60% more time spent on reading instruction on a daily basis than students in the Content RR classrooms, students in the Literature RR classrooms had lower gains in WCPM than those of students in the classrooms with the Content RR treatment.

Discussion

As part of a movement to increase scientifically based classroom practice, findings from experimental studies are the basis for mandates and reform efforts in thousands of American classrooms (U.S. Department of Education, 2003). While the research literature in domains such as phonemic awareness or phonics has considered many of the issues that arise in classroom applications (National Reading Panel, 2000; Snow, Burns, & Griffith, 1998), the research literature on fluency has not been as extensive, applied, or robust. While the NRP (2000) addressed the pedagogical techniques that support fluency, the equally critical instructional components of text features and time allocation were not addressed. This study focused on the first of these components and, secondarily, on the second.

The features of texts made a difference over and above the application of the repeated reading techniques that were advocated by the NRP (2000). The greatest difference lay between the Content RR and the Control group—a difference of almost 10 WCPM. At least for second graders, opportunities to read repeatedly in texts with considerably few rare, multisyllabic words resulted in greater gains in WCPM.

Similar to the finding of Faulkner and Levy (1994) who found that initial reading proficiency interacted with text features, the text level of the content texts was most effective for students in the bottom two quartiles. This pattern is promising because of the consistently sizable group that fails to attain even a basic standard on the NAEP (Donahue et al., 2001). Gains of 10 WCPM more for the Content RR group than for the Control group are not trivial when classes have as many students who enter second grade unable to read a fundamental group of high-frequency words as in the experimental schools in this study. While students in the bottom quartile made up only 7% of the Control group, they made up 49% of both intervention groups. As would be expected, the same texts did not benefit all students to the same degree. Among students who began at the 55th percentile or higher according to fluency norms (Good, Wallin, Simmons, Kame'enui, & Kaminski, 2002), reading literature produced similar gains as repeatedly reading the content texts.

There has been little research to date on the effects that mandates on increased allocations of time to reading instruction (and less time on science, social studies and other subjects) have had on students' reading achievement. At least when it comes to fluent reading, simply increasing the length of reading periods does not mean that students will read more fluently. However, ensuring that students spend their time in particular activities does appear to increase reading fluency. The students in the two Content RR classrooms where repeated reading was consistently implemented made higher gains in reading rate than their peers in Literature RR classrooms where the mandated length of reading periods was substantially longer.

No matter how accessible a text may be, developing readers need opportunities to repeatedly read it. When opportunities to read the content texts repeatedly were not offered to students, their gains in reading rate were less than those of their counterparts

in classrooms in the same school where the opportunity to read texts repeatedly was consistent—approximately .4 WCPM for each of 20 weeks. This difference is near to the .5 that Fuchs, Fuchs, Hamlett, Walz, and Germann (1993) have identified as the level of ambitious growth needed per week if initially struggling second graders are to close the gap.

Another finding related to time allocation was that doubling the length of an intervention did not have the effect of doubling students' gains in fluency. In a previous study with similar procedures but lasting 10 weeks (Hiebert, 2003a), the gain was 25 words for the Literature RR students and 31 for the Content RR. The gains for a 20-week period in the present study were not substantially greater: 27 for the Literature RR students and 33 for the Content RR group. Several explanations can be offered for this pattern. For one, a renewed professional development effort was not included as part of the extended intervention in the present study. It may also be that two 10-week periods at different points in the year are preferable to one sustained effort. For example, one period might occur at the beginning of a school year when students are making up for the loss that they suffered over the summer. Norms of oral reading fluency suggest that beginning readers in the bottom quartile lose around 25% of their fluency levels over the summer (Good et al., 2002; Hasbrouck & Tindal, 1992). Another period of focused fluency practice might occur at the end of the year, giving students the opportunity to consolidate what they have learned over the year.

This study has uncovered numerous questions that require attention, if the many students who are not highly fluent readers are to develop the skills they need to be fully literate. The concluding section of this paper addresses three issues that arose in the course of this study. The first has to do with what it takes to bring a school where a majority of students are below fluency norms to the standards set by NCLB. Bringing a cohort of second graders to approximately 66 WCPM on a grade-level passage by spring was a substantial accomplishment on the part of Content RR teachers, in that half of their students could not recognize a handful of words in the fall. However, these students have considerable ground to cover if they are to attain the 130 WCPM that Pinnell et al. (1995) identified as necessary for proficient reading on assessments like the NAEP at grade four. On average, students would need to make

similar gains in reading rate—32 to 33 WCPM—in grades 3 and 4 as they did during the second-grade intervention. In schools where many children fall substantially below state and national standards, interventions need to extend over several years, rather than be limited to a single grade.

As part of future research, data are needed on the amount of reading that leads to changes in fluency levels. While research provides descriptions of the amount of text that students at different reading levels read after-school (Anderson, Wilson, & Fielding, 1988), no comparable data exists on the amounts that students of different proficiencies read in classrooms. An analysis of the two basal programs that are the centerpiece of California's massive state reform effort indicates that these programs provide approximately 130 words of text per day over grades one through three (Hiebert, 2003a). If students read all of the available texts in the basal program's core components (i.e., anthologies and, for grades one and two, decodable texts), they would read approximately 70,500 words from grades one through three. This number of words is similar to the amount that fourth and fifth graders at the 25th percentile in the Anderson et al. (1988) study claimed to read at home in a year. Is 70,500 words over three grades of reading/language arts instruction a sufficient amount of text to develop sufficient fluency? Is this amount of text sufficient when rereading is part of the instructional routine? Questions such as these have yet to be addressed. If the goals of NCLB are to be attained for the students who are in highly challenged schools, systematic attention to optimal opportunities for reading text in classrooms is needed.

Another area that requires work has to do with the genres of texts used in reading programs, particularly for establishing fluency. The choice of informational text for fluency development in the content intervention in this study was theoretically derived. Research of the 1980s (e.g., Beck et al., 1984; Bruce, 1984) questioned the control of vocabulary in texts. What was not considered in these studies is the role of repetition in different genres of texts. The texts that provided the focus of the studies of the 1980s were narrative. In narrative texts, authors use rare or infrequent words to communicate a salient feature of a character, setting, or event. Informational texts—even those sold by trade rather than text divisions of publishers—use fewer rare words and the rare words that

are used are repeated more frequently (Hiebert, 2003a). Thus, the repetition of those words that were rare and/or multisyllabic in the content texts used in this study complies with the style of authentic informational text. However, informational texts may make demands on students' background knowledge that direct students' attention away from reading rate and fluency. Work on the role of genre needs to go beyond the argument for a better balance of informational and narrative text in the primary grades (Duke, 2000). The role of different types of texts needs to be considered for different functions of reading, including the development of background knowledge, interest, and engagingness for different groups of students and at different points in students' development.

Finally, much more work is needed on the decision-making of teachers. The aim of this study was not to establish why some teachers are compliant in implementing an intervention and others are not. Teachers may have agreed to participate in the current study because of the promise of new materials or because of pressure from administrators and colleagues to participate. It is also possible that teachers make different choices depending on the profiles of their classrooms. For example, the teacher in the Content RR treatment with the students who scored the highest on the pretest was the one who did not implement the intervention systematically. It may also be that fluency was not well understood by some of the teachers. Mandates can be issued by state and federal agencies regarding fluency levels but, without teacher understanding, opportunities for repeated reading may not be available to the children who need them. Teachers' knowledge about fluency, their choices about fluency instruction and the effects of these choices on student outcomes, and discussions with teachers about these choices need to be a focus of future research on fluency.

In conclusion, the findings of this study add to the evidence on the efficacy of repeated reading and on the features of texts that influence students' fluency. The findings also suggest that simply increasing the length of reading instruction or of repeated reading occasions is not necessarily the answer. Longer instructional periods may not compensate for the difficulty of texts. Many questions remain in establishing the manner in which accessible texts and the amount that students read influence fluency over the elementary grades.

References

Aaron, I. E., Jackson, D., Riggs, C., Smith, R. G., Tierney, R. J., & Jennings, R. E. (1983). *Scott, Foresman Reading*. Glenview, IL: Scott Foresman.

Afflerbach, P., Beers, J., Blachowicz, C., Boyd, C. D., Diffily, D., Gaunty-Porter, D., Harris, V., Leu, D., McClanahan, S., Monson, D., Perez, B., Sebesta, S., & Wixson, K. K. (2000). *Scott Foresman Reading*. Glenview, IL: Scott Foresman.

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

Beck, I. L., McKeown, M., Omanson, R., & Pople, M. (1984). Improving the comprehensibility of stories: The effects of revisions that improve coherence. *Reading Research Quarterly, 19*, 263–277.

Berliner, D. C., & Fisher, C. W. (1985). One more time. In C. W. Fisher & D. C. Berliner (Eds.), *Perspectives on instructional time* (pp. 333–347). New York: Longman.

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

Bruce, B. C. (1984). A new point of view on children's stories. In R. C. Anderson, J. Osborn, & R. J. Tierney (Eds.), *Learning to read in American schools: Basal readers and content texts* (pp. 153–174). Hillsdale, NJ: Erlbaum.

California State Board of Education (2002). *Reading/language arts/English language development publishers list* (2002–2008 State Board adopted). Sacramento, CA: California 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.

Dowhower, S. L. (1987). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. *Reading Research Quarterly, 22*, 389–406.

Duke, N. K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly, 35*, 202–224.

Duke, N. K., & Kays, J. (1998). "Can I say 'once upon a time'?": Kindergarten children developing knowledge of information book language. *Early Childhood Research Quarterly, 13*, 295–318.

Eldredge, J. L., Reutzel, D. R., & Hollingsworth, P. M. (1996). Comparing the effectiveness of two oral reading practices: Round-robin reading and the shared book experience. *Journal of Literacy Research, 28*, 201–225.

Faulkner, H. J., & Levy, B. A. (1994). Fluent and nonfluent forms of transfer in reading: Words and their message. *Psychonomic Bulletin and Review, 6*, 111–116.

Foorman, B. R., Francis, D. J., Davidson, K. C., Harm, M. W., & Griffin, J. (2004). Variability in text features in six grade 1 basal reading programs. *Scientific Studies of Reading, 8*, 167–197.

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.

Glass, G. V., & Hopkins, K. D. (1995). *Statistical methods in education and psychology* (3rd ed.). Boston: Allyn & Bacon.

Good, R. H., Wallin, J. U., Simmons, D. C., Kame-enui, E. J., & Kaminski, R. A. (2002). *System-wide percentile ranks for DIBELS Benchmark Assessment* (Technical Report 9). Eugene, OR: University of Oregon.

Guthrie, J. T., Schafer, W. D., & Huang, C. (2001). Benefits of opportunity to read and balanced instruction on the NAEP. *Journal of Educational Research, 94,* 145–162.

Hasbrouck, J. E., & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in grades 2 through 5. *Teaching Exceptional Children, 24*(3), 41–44.

Heckleman, R. G. (1966). Using the neurological impress remedial reading technique. *Academic Therapy Quarterly, 1,* 235–239.

Hiebert, E. H. (2003a, April). *The role of text in developing fluency: A comparison of two interventions.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Hiebert, E. H. (2003b). *QuickReads.* Parsippany, NJ: Pearson Learning Group.

Hiebert, E. H., & Fisher, C. W. (2002, April). *Text matters in developing reading fluency.* Paper presented at the annual meeting of the International Reading Association, San Francisco, CA.

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

Kellogg, S. (1977). *The mysterious tadpole.* New York: Dial Books for Young Readers.

Kuhn, M. (in press). A comparative study of small group fluency instruction. *Reading Psychology.*

Medearis, A. S. (1997). *Annie's gifts.* Glenview, IL: Scott Foresman.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19,* 304–330.

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.

Pearson, P. D. et al. (1998). *Literature works.* Chicago: Pearson Education.

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: Office of Educational Research and Improvement, U.S. Department of Education; Princeton, NJ: Educational Testing Service.

Rashotte, C., & Torgesen, J. K. (1985). Repeated reading and reading fluency in learning disabled children. *Reading Research Quarterly, 20,* 180–188.

Rasinski, T. V. (1999). Exploring a method for estimating independent, instructional, and frustration reading rates. *Reading Psychology, 20,* 61–69.

Ravid, R. (1994). *Practical statistics for educators.* New York: University Press of America.

Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher, 32,* 403–408.

Stahl, S. A., Heubach, K., & Cramond, B. (1997). *Fluency-oriented reading instruction* (Reading Research Report No. 79). Athens, GA, and College Park, MD: Universities of Georgia & Maryland, National Reading Research Center.

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

U. S. Congress. (2001). *No Child Left Behind Act: Reauthorization of the Elementary and Secondary Education Act (PL 107-110).* Retrieved April 10, 2003, from www.ed.gov

U. S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide.* Washington, DC: Author.

Wiederholt, J. L., & Bryant, B. R. (2001). *Gray Oral Reading Test* (4th ed). Austin, TX: Pro-Ed.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide.* New York: Touchstone Applied Science Associates, Inc.