Instruction, Student Engagement, and Reading Skill Growth in Reading First Classrooms

Carol McDonald Connor
Lara J. Jakobsons
Elizabeth C. Crowe
Jane Granger Meadows

Florida Center for Reading Research and Florida State University

The Elementary School Journal
Volume 109, Number 3

0013-5984/2009/10903-0001$10.00

Abstract

This study explored first-, second-, and third-grade reading instruction and students' (n = 1,586) reading skill growth in Florida Reading First (RF) classrooms (n = 95). The goal of RF is to improve students' literacy outcomes through the use of research-based instruction, assessment, teacher training, and program evaluation. Whereas survey data have shown that schools are generally implementing the major components of RF, we found no studies that related observed classroom instruction and student reading outcomes, which was the purpose of this study. Students attended schools that were historically high poverty with generally weak student achievement. Standardized assessments of vocabulary and reading comprehension, as well as progress-monitoring oral reading fluency tests, were administered. Classrooms were observed in the spring of the school year for 1 hour during reading instruction. In general, first graders demonstrated expected reading comprehension skills by the end of the school year; however, second and third graders did not. Hierarchical linear models revealed that the effect of instruction on students' reading comprehension was highly complex, varying substantially across grades. Further, the effect of instruction depended on the grade observed as well as students' vocabulary and fall oral reading fluency skills.

Recent reports of the National Assessment of Educational Progress (NAEP, 2005) have revealed that almost 40% of U.S. fourth graders fail to reach basic levels of reading. This means that over 15 million children do not have even "partial mastery of the knowledge and skills that are fundamental for proficient work at a given grade" (NAEP, 2003, p. 2). Somewhat encouraging is that, across the United States, students' fourth-grade reading scores have improved

from previous years—one point since 2003 and two points since 1992, the year in which the assessment was initiated. Moreover, for seven states, including Florida, the percentage of students scoring below a basic level decreased (e.g., for Florida 2003 = 37% below basic; 2005 = 35% below basic). Although the reasons for this somewhat encouraging trend are unclear and undoubtedly related to simultaneous national, state, local, community, classroom, and home sources of influence (Connor, Son, Hindman, & Morrison, 2005; Morrison, Bachman, & Connor, 2005), one possible contributing factor in Florida is the state's Reading First initiative. Reading First (RF) is a national program, created as part of the No Child Left Behind Act (NCLB, 2002), which provides grants to states with the goal of improving students' reading achievement through encouraging the use of "scientifically proven methods of instruction" (Reading First, 2007). The recent interim report on RF nationwide (U.S. Department of Education, 2006) revealed that "Reading First schools appear to be implementing the major elements of the program as intended by the legislation, such as providing scientifically based reading instruction in grades K–3, increased amounts of time for reading instruction, interventions for struggling readers, wider use of classroom-based reading assessments, and more professional development activities" (p. 3). This finding was based on results from surveys completed by teachers, principals, and reading coaches at over 1,000 schools receiving RF funds and over 500 Title 1 non-RF schools. There was, however, neither observation of the instruction provided to students nor consideration of student reading outcomes. Therefore, the purpose of this study was to (1) describe the nature and variability of reading instruction strategies used during classroom observations conducted in a randomly selected sample of Florida Reading First schools; and (2) to examine the relation of RF-supported instruction, across multiple dimensions, to students' reading skill growth.

Reading First

Reading First represents a major policy shift in the federal government with regard to education. Attention has changed from a focus on resources to a focus on student achievement (Cohen, Raudenbush, & Ball, 2003). This shift has been coupled with a major financial investment, almost $6 billion, with funds targeted to students enrolled in schools with high percentages or numbers of students reading below grade level from kindergarten through third grade and schools with large numbers of students from low-income families. RF provides funds for professional development for teachers and encourages the use of reading curricula and resources that are based on scientific reading research, as well as screening, diagnostic and progress-monitoring assessments (FCRR, 2007). Statewide accountability structures are funded as well, including site visits.

With its mission to support all children's literacy achievement, RF in Florida focuses primarily on improving teachers' use of effective reading instruction practices through the adoption of a scientifically based, core reading curriculum and by providing professional development. Additionally, teachers are encouraged to use student assessment results to guide their instructional practices.

In Florida, trained assessment teams assess students four times per year using the Dynamic Indicators of Basic Early Literacy Skills (DIBELS, Good & Kaminski, 2002). These progress-monitoring results are kept in a statewide, networked electronic database. Teachers are encouraged to access these results and use them to plan differentiated instruction. End-of-year assessments include measures of vocabulary and reading in kindergarten through third grade in addition to the state-mandated assessment, the Florida Comprehensive Achievement

Test (FCAT), administered to all third-grade students statewide. These results are used to assess the overall effectiveness of the Reading First effort.

Generally, students' reading scores improved steadily when their schools participated in RF. For example, the mean fourth-grade NAEP score for children living in Florida in 1998 (before RF) was 206, compared to a national mean of 215. By 2005 (after 2–3 years of RF support) the Florida mean NAEP score rose 13 points to 219 and exceeded the national average of 217 (Torgesen, 2006). Although overall the trends in Florida on the state-mandated third-grade test of reading showed a decrease in the number of students performing in the lowest quartile, the decrease was greatest for schools participating in RF, although those schools tended to serve students and communities with greater challenges than the non-RF schools.

In an effort to understand how (or whether) teachers in RF schools were using research-based classroom practices in response to the professional development and resources provided by RF funds, in the study reported in this article, we conducted site visits in a randomly selected subset comprised of about 10% of participating schools. These visits included interviews with staff and unannounced observations of randomly selected classrooms during the mandated 90-minute block of reading instruction (FCRR, 2007). For this study, we investigated the nature and variability of content and time spent in various classroom literacy activities, the extent to which teachers differentiated instruction, and levels of student engagement using site visit data from the first through third grades.

Amount and Type of Instruction

Multiple factors affect children's learning (Bronfenbrenner, 1986; Morrison et al., 2005), with instruction and classroom environment emerging as important influences within a larger system (Connor et al., 2005), especially when conceptualized across a number of dimensions. RF in Florida focuses on the components of reading described in the National Reading Panel report (NRP, 2000). These components include phonological awareness, phonics, fluency, vocabulary, and comprehension. Consequently, the site visit observations in this study focused on these areas of instruction, as well as on how well teachers differentiated instruction and the extent to which students were engaged in literacy activities.

Viewing instruction at the curriculum level (e.g., Dahl & Freppon, 1995; Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Wharton-McDonald, Pressley, & Hampston, 1998) tends to underestimate the complexity of instruction that children experience. Research has also revealed that teachers may use a combination of strategies to enhance students' reading comprehension skills (Chall, 1967; National Institute of Child Health and Human Development–Early Child Care Research Network [NICHD-ECCRN], 2002; NRP, 2000). Moreover, important differences have been observed in children's learning when instruction is largely teacher managed (e.g., teachers reading a book to students) compared to the same activity in which children are expected to work independently or with peers (e.g., sustained independent silent reading; Connor, Morrison, & Katch, 2004). Thus, we used a multidimensional view of literacy instruction.

In this study we investigated two dimensions of instruction, code versus meaning focused (i.e., the content of instruction) and teacher versus child managed (who is managing instruction), across three grades. These dimensions operate simultaneously (see Table 1). Thus, any language arts activity falls in one of four quadrants. Additionally, observers globally rated each classroom regarding the extent to which the teacher differentiated instruction according to students' reading skills and students' general on-task behavior. Each dimension is discussed below.

TABLE 1. Instructional Activities According to Dimension of Instruction

Variable	Teacher Managed	Child Managed
Code focused	Phonemic awareness: Rhyming Blending/segmenting sentences and syllables Onset/rime Blending/segmenting phonemes Alphabetic knowledge letter identification/recognition Phonics: Letter/sound relationships Application of letter/sound knowledge to reading/writing/spelling Irregular words Word reading Integration of word study Spelling Letter or sound naming fluency Word fluency	
Meaning focused	Concepts of print Supported oral reading Choral reading Comprehension: Vocabulary Prior knowledge/predicting Monitoring Listening comprehension monitoring Comprehension strategy instruction/use Oral language development: In expansion of student-initiated language In structured opportunities to talk with teacher/peers Repeated reading of text Writing composition: Independent composition Independent writing/publishing Grammar and punctuation Handwriting instruction and copying	Independent silent reading Independent oral reading Independent writing/publishing

Code- versus Meaning-Focused Instruction

The dimension, code versus meaning focused, denotes the overall goal or content of the instructional activity. Activities designed to help children learn to "crack the code" are considered code focused (Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001). These activities include phonological awareness, phonics, word-level fluency, and alphabetic knowledge. In comparison, meaning-focused activities are designed to teach children how to extract and construct meaning from text (Snow, 2001). Such activities include reading aloud, vocabulary, predicting, inferring, and discussion (see Table 1).

Teacher- versus Child-Managed Instruction

The dimension, teacher versus child managed, denotes who is managing or focusing students' attention on the learning opportunity at hand—the teacher or the child. This dimension should not be confused with child centered and teacher directed instruction (Bredekemp & Copple, 1997). Child-centered activities speak to the motivator of the activity but may still be teacher managed. For example, if a student is interested in fire engines, the teacher may allow her to select a book about fire engines and then read it to her. Using this conceptualization, the child-centered activity would be considered teacher managed, meaning fo-

cused (TM, meaning focused). Any time the teacher is actively involved in an instructional activity, even one that is highly interactive or initiated by a child, the activity is considered to be teacher managed. Activities where children work independently or with peers are considered to be child managed because the students, not the teacher, are managing the learning opportunity. In this study, we considered peer- and child-managed instruction to be child managed. There is some evidence that peer-managed activities differ substantially from teacher-managed activities (Christian & Bloome, 2004; Palincsar, Collins, Marano, & Magnusson, 2000). Examples of teacher- and child-managed activities are provided in Table 1. Importantly, the dimensions operate simultaneously so that, for example, instruction can be teacher managed, code focused or teacher managed, meaning focused.

Differentiating Instruction and Child by Instruction Interactions

In the literature on effective schools and schools in high-poverty neighborhoods whose students consistently demonstrate high achievement, effective teachers use small, homogeneous groups based on reading skills and assessment to guide instruction (Pressley et al., 2001; Pressley, Wharton-McDonald, Mistretta-Hapston, & Echevarria, 1998; Taylor, Pearson, Clark, & Walpole, 2000; Wharton-McDonald et al., 1998). This kind of instruction has been called differentiated or individualized. A recent randomized field trial revealed that students whose teachers individualized instruction on the basis of assessment results demonstrated significantly greater reading comprehension skill growth compared to students in control classrooms (Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007). Moreover, the effect of an instructional strategy appears to depend on students' language and literacy skills. These have been called child characteristics by instruction interactions (Connor, Morrison, & Katch, 2004). In the present study, observers rated the extent to which each teacher differentiated instruction and provided extra instruction to the lowest-performing students in the classroom. Additionally, we tested for specific child characteristics by instruction interactions using hierarchical linear models.

Reading First Instruction Viewed Multidimensionally

Several important reports, such as *Preventing Reading Difficulties in Young Children* (Snow, Burns, & Griffin, 1998), the National Reading Panel Report (NRP, 2000), and the Rand report on reading comprehension (Snow, 2001), have reviewed the literature and have been used as guides in the NCLB Act and RF guidelines to define scientifically based reading research. Following the recommendations of these documents, what kinds of changes would one expect to see in instruction in RF schools compared to instruction prior to NCLB? A general consensus exists that, for children at risk for reading difficulties, explicit code-focused instruction is more effective than embedded or less systematic instruction (Foorman et al., 1998; NRP, 2000). Thus, teachers who implement research-informed instruction should provide substantial code-focused instruction across all grades (NRP, 2000), particularly in first grade. Additionally, most researchers agree that teaching children reading comprehension strategies contributes to stronger comprehension and general reading overall (NRP, 2000; Palincsar & Brown, 1984; Snow, 2001). Thus, in RF schools, one would expect to observe meaning-focused instruction, especially in the second and third grades where the research evidence is most robust.

Additionally, emerging research has suggested that teacher-managed instruction is generally more effective with chil-

dren at risk for reading difficulties or those struggling with reading than are child-managed activities (Connor, Morrison, & Katch, 2004; Connor, Morrison, & Petrella, 2004; Foorman et al., 1998; Torgesen et al., 2001). Therefore, in RF schools, which serve many children at risk for significant reading difficulties, one should generally see more time in teacher- rather than child-managed activities.

Child Characteristics

In this study we examined students' end-of-year reading comprehension skills. Arguably one of the most important skills in society (NRP, 2000), reading comprehension has been defined as the active extraction and construction of meaning from text (Snow, 2001). Comprehension calls on students' background and cultural knowledge as well as their sophistication with language and its use; it is influenced by the child's motivation (Guthrie et al., 2004), text difficulty, and genre (e.g., narrative, expository) of the text (Snow, 2001). Researchers have frequently overlooked it as an outcome of interest until the middle- and upper-elementary grades. The understanding of what factors contribute to students' reading comprehension and ways to increase their understanding of what they have read has become increasingly important. Research has demonstrated the importance of fluent phonological decoding skills in the early grades (Rayner et al., 2001). Moreover, there is strong evidence that reading comprehension can be improved through consistent instruction and the use of reading comprehension strategies (Williams et al., 2005). However, the contribution of instruction to students' reading comprehension, especially in first and second grade, where the focus is on establishing basic decoding skills, is less well understood and was one of the principal foci of this study.

Just as we viewed instruction multidimensionally, so too did we examine multiple child characteristics that influence reading outcomes. We chose oral reading fluency, decoding, phonological awareness, and vocabulary as variables of interest because of the theoretical importance of these skills to the development of proficient reading and reading comprehension (Dickinson & Tabors, 2001; Hart & Risley, 1995; Morrison & Cooney, 2002; NRP, 2000; Rayner et al., 2001; Snow, 2001; Storch & Whitehurst, 2002; Therrien, 2004).

Beyond the language and cognitive skills associated with students' academic success, research has revealed that, in contrast to disengaged readers, engaged readers have higher reading achievement and comprehension scores (Campbell, Voelkl, & Donahue, 1997; Cunningham & Stanovich, 1997). Likewise, numerous studies have shown that students' inattentive behavior in the classroom is related to lower academic performance (e.g., Finn, Pannozzo, & Voelkl, 1995; Haskins, Walden, & Ramey, 1983). For example, Greenwood, Terry, Marquis, and Walker (1994) demonstrated that student engagement in academic responding mediated instruction and achievement. Another investigation showed that students' time on task predicted growth in reading comprehension (Taylor, Pearson, & Rodriquez, 2003). In sum, these studies underscore the importance of better understanding the relation between student engagement and reading achievement.

Research Questions

The following research questions guided this inquiry: (1) What was the achievement of children in these RF classrooms and how did their performance compare to state and national norms? We hoped to see greater achievement than might be predicted for children in historically underperforming schools that face many challenges in teaching students to read proficiently.

(2) What is the nature and variability of classroom literacy instruction in first-,

TABLE 2. Demographics and Sample Characteristics, in Percentages

Demographic	First Grade (n = 493)	Second Grade (n = 550)	Third Grade (n = 543)
Gender:			
Male	49	54	52
Race/ethnicity:			
Caucasian/white	33	35	33
African American	39	40	40
Hispanic	22	20	23
Other	6	5	4
Free/reduced-price lunch:			
Did not qualify or apply	22	20	24
Free lunch	66	62	60
Reduced-price lunch	9	14	12
Enrolled in USDA-approved Provision Z school	3	4	4

second-, and third-grade RF classrooms? We anticipated substantial variability in amount and type of instruction observed both among classrooms and among grades. Additionally, we hypothesized that amount of teacher-managed, code-focused instruction would be fairly high across all three grades, particularly in first and second grade. Further, we anticipated a fairly high amount of teacher-managed, meaning-focused instruction in third grade RF classrooms. These patterns of instruction would be anticipated in RF schools that were implementing instruction based on extant reading research.

(3) What is the effect of amount and type of instruction, across multiple dimensions and grades, on students' reading comprehension, and are there child characteristic by instruction type interactions? Because our outcome was reading comprehension rather than word-reading skill, we hypothesized that meaning-focused instruction would have a greater effect on student outcomes than would code-focused instruction. We anticipated child by instruction interactions, however, such that the effect of instructional strategies would depend on students' initial vocabulary and reading skills. Additionally, we hypothesized that higher student engagement and more differentiated instruction would be positively related to higher student reading skill across grades.

Method

Participants

In all, 1,586 children taught by 95 teachers in 33 schools participated in this cross-sectional study. We selected these schools randomly from among the 328 schools that were participating in Florida RF during the 2003–2004 school year. Students were in grades 1, 2, and 3 (n = 493, 550, and 543, respectively) and attended primarily low- and middle-socioeconomic status (SES) schools in 17 districts. All schools met the RF criteria for low student achievement or generally low student SES. Characteristics of the students by grade are presented in Table 2. Not surprisingly, this sample included a high percentage of children from underrepresented minorities and low-SES households.

In this sample, there were 31 first-grade, 32 second-grade, and 32 third-grade teachers. All teachers met state certification requirements; each was licensed to teach elementary grades and held a bachelor's degree at a minimum. No one core curriculum was used in all of the classrooms, although all curricula had to meet the RF guidelines and were based on current reading research. For each classroom, complete assessment data were available for about 17 students.

Child Assessments

As part of the RF assessment protocol, children were assessed four times during

TABLE 3. Descriptive Statistics for Student-Level Measures, by Grade

Measure	First Grade	Second Grade	Third Grade
Spring:			
Oral vocabulary (PPVT-III) standard score:			
Mean	89.27	90.99	92.33
SD	13.78	13.75	14.29
n	504	554	553
Reading comprehension (SAT-10) scaled score:			
Mean	553.53	589.78	616.54
SD	48.56	36.66	35.61
n	493	550	543
Fall:			
Oral reading fluency:			
Mean	13.54	48.47	72.45
SD	17.50	25.93	28.37
n	500	552	554
Nonsense word fluency:			
Mean	27.99	47.93	...
SD	19.46	26.86	...
n	500	552	...
Phoneme segmentation fluency:			
Mean	28.42	...	...
SD	16.29	...	...
n	500	...	...
Letter naming fluency:			
Mean	42.87	...	...
SD	15.76	...	...
n	500	...	...

the 2003–2004 school year, using a battery of progress-monitoring reading measures that varied depending on grade level (FCRR, 2007a). End-of-year spring assessments included vocabulary and general reading. School staff who had been trained to administer the assessments did so. The assessments are described below. Means and standard deviations are provided in Table 3. As expected, children's scores were moderately positively correlated, and correlations ranged from a low of .38 (first-grade vocabulary and fall oral reading fluency, $p < .01$) to a high of .67 (third-grade fall oral reading fluency and spring reading comprehension, $p < .01$).

Reading comprehension. The outcome in our study, reading comprehension, was assessed using the Stanford Achievement Test, 10th edition (SAT-10, Harcourt Educational Measurement, 2003). The reading comprehension subtest of the SAT-10 taps students' initial understanding, critical analysis, interpretation, and awareness and usage of reading strategies in grades 1, 2, and 3. Students read functional, recreational, and textual passages and answer multiple-choice questions about the passages. We used the scaled IRT scores for grades 1, 2, and 3 in these analyses; we obtained grade-equivalent scores using published norms. Reported reliability was .89 for first-grade, .91 for second-grade, and .93 for third-grade reading comprehension.

Vocabulary. Children's receptive vocabulary was assessed using the Peabody Picture Vocabulary Test–third edition (PPVT-III, Dunn & Dunn, 1997). In this task, the examiner said increasingly unfamiliar words and, for each word, the child selected one of four picture alternatives that he or she felt best depicted the definition of the word. Dunn and Dunn (1997) reported an internal consistency reliability of .95 and a split-half reliability of .94. Children were assessed in April, so we used standard scores in the analyses. We relied on the well-documented stability of vocabulary

skills to justify this variable's inclusion as a predictor variable in our models.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS). Grade-appropriate assessments (DIBELS, Good & Kaminski, 2002) were administered to all students four times per year in September, December, February, and April (FCRR, 2007c). For all of the assessments, we used the raw scores in the analyses; grade-equivalent scores were obtained using Florida published norms, which were accessed from the Web (FCRR, 2007a). In general, the DIBELS assessments have adequate reliability (.52–.91) (Shaw & Shaw, 2002) and validity when used as progress-monitoring assessments (see also University of Oregon, 2007) but do not have the strong psychometric properties of the SAT-10 and the PPVT-III. For this reason, we used the SAT-10 as the outcome of interest in this study, with fall status captured by the grade-appropriate DIBELS assessments administered in the fall. We used the oral reading fluency subtest (described below) to examine students' progress during the school year because it was the only subtest administered at all three grades.

Oral reading fluency: Children's fluency and accuracy with text were assessed using the DIBELS oral reading fluency (ORF). In this task, the child reads three grade-level passages orally for 1 minute each. The oral reading fluency rate is the number of words read per minute in the passage minus errors (i.e., total correct). Errors are words substituted or omitted, or pauses greater than 3 seconds. We used the median score in the analyses.

Letter naming fluency: Administered only in fall of first grade, the letter naming fluency (LNF) subtest of DIBELS was used to measure how quickly a child could name letters of the alphabet. In this assessment, the child is provided with letters arranged in random order on a sheet of paper. The examiner asks the child to name as many letters as he or she can within 1 minute. The reported raw scores are the number of letters correctly named in 1 minute.

Phoneme segmentation fluency: Children's ability to fluently segment words into their separate phonemes was assessed only in first grade using the phoneme segmentation fluency (PSF) subtest of DIBELS. This measure taps children's phonemic awareness overall. In this task, the examiner states three- and four-phoneme words and asks the child to orally identify the individual phonemes in each word. The raw score is the number of correct phonemes named in 1 minute.

Nonsense word fluency: Administered in first and second grade, the nonsense word fluency (NWF) subtest of DIBELS was designed to measure a child's ability to blend letters, using letter-sound correspondence, into words in which the letters signify their most frequent sounds (Kaminski & Good, 1996). The child's score is computed from the total number of letter sounds read aloud (blended or not) in nonsense words correctly in 1 minute.

Classroom Observations and Coding Systems

Site visits. We observed classroom instruction during the spring of the 2003–2004 school year in 33 RF schools selected randomly by region. Student progress-monitoring and outcome data were available for 31 first-grade classrooms and 32 second- and third-grade classrooms. Missing-data analyses revealed no differences in school or classroom characteristics for the classrooms for which we were unable to match student data and teachers.

We notified schools several weeks in advance of the visit; however, randomly selected classrooms (one per grade K–3) were not notified in advance that they would be observed. All K–3 teachers were told that any classroom could be observed during the reading block. Observations lasted, on average, about 47 minutes in first grade, but this ranged from 40 to 55 min-

utes; 45 minutes in second grade, ranging from 30 to 60 minutes; and 46 minutes in third grade, ranging from 35 to 50 with one outlier at 102 minutes. For this reason, the effect of observation length was controlled in each model.

Trained reading experts conducted the observations using the revised edition of the Instructional Content Emphasis observation tool (ICE-R, Vaughn & Briggs, 2003). This instrument was designed to systematically code and categorize the content of reading and language arts instruction. Content validity was established through an extensive review of literature describing beginning reading instruction, including national and state reading standards and best practices in literacy instruction. Observers collected data in 10 major instructional categories: (1) concepts of print, (2) phonological awareness, (3) alphabetic knowledge, (4) word study/phonics, (5) spelling, (6) oral language development/discussion, (7) fluency, (8) text reading, (9) comprehension, and (10) writing or language arts (see Table 4 and App. Fig. A1). Descriptive indicators for the 10 categories were developed to increase reliability of the instrument. During the development of the instrument, an interrater reliability rate of 91% was achieved.

Classroom observers were selected for Florida's RF site-visit project based on their expertise in reading instruction and on successful completion of a 2-day training session during which they practiced using the ICE-R instrument to code video segments of reading instruction. During the culminating phase of the training, participants were asked to code a final video segment of reading instruction and to submit their coding sheets to the training facilitator, who checked the coding for accuracy. An interrater reliability rate of 81% was achieved.

The trained observers recorded timed narratives that documented each instructional activity type by category and duration (see App. Fig. B1). The observer followed the teacher; if the class was divided

TABLE 4. Descriptive Statistics for the ICE-R Main Categories across Grades in Minutes Observed

Category/Grade	*M*	*SD*	Maximum[a]
Concepts of print:			
First grade	.0968	.39622	2.00
Second grade	.1406	.34159	1.00
Third grade	.0625	.35355	2.00
Total	.1000	.36173	2.00
Phonological awareness:			
First grade	1.3226	2.74939	10.00
Second grade	.1250	.55358	3.00
Third grade	.0000	.00000	.00
Total	.4737	1.69381	10.00
Alphabetic knowledge:			
First grade	.2258	.95602	5.00
Second grade	.0000	.00000	.00
Third grade	.0000	.00000	.00
Total	.0737	.55048	5.00
Word study and phonics:			
First grade	11.3548	8.82250	26.00
Second grade	5.0000	6.82027	25.00
Third grade	4.7969	6.99292	23.00
Total	7.0053	8.09797	26.00
Spelling:			
First grade	.7742	2.44554	10.00
Second grade	1.2500	3.19273	13.00
Third grade	.3750	1.79156	10.00
Total	.8000	2.54157	13.00
Oral language development:			
First grade	.6129	1.49839	5.00
Second grade	.7813	1.86192	7.00
Third grade	.4063	1.36451	6.00
Total	.6000	1.58047	7.00
Fluency:			
First grade	.7742	2.70444	13.00
Second grade	2.1563	4.82674	21.00
Third grade	1.2813	2.71477	11.00
Total	1.4105	3.57442	21.00
Text reading:			
First grade	12.4839	9.68115	37.00
Second grade	8.2500	8.07225	29.00
Third grade	9.8375	9.33428	34.00
Total	10.1663	9.12106	37.00
Comprehension:			
First grade	10.3548	8.42041	31.00
Second grade	18.3281	10.38269	44.00
Third grade	22.0828	14.54415	73.00
Total	16.9911	12.32471	73.00
Writing and language arts:			
First grade	4.2903	8.22676	38.00
Second grade	1.5938	4.87164	26.00
Third grade	2.1250	4.40491	17.00
Total	2.6526	6.09808	38.00

[a]The minimum score was .00 for all categories. $N = 31$ for first grade, 32 for second grade, and 32 for third grade.

into small groups and the teacher was working with one of those groups, then that was the only activity recorded. For this reason, the amount of child-managed instruction is underreported.

The duration of each activity was reported in minutes of instruction. Means and standard deviations are provided for each grade in Table 4. Note that only half of the 90-minute language arts block was observed for any one classroom, so these data are invalid to judge a single classroom or school. Averaged across classrooms, however, the data provide an adequate snapshot of instruction.

Dimensions of instruction, instructional activities, and student engagement. Teacher versus child managed: The ICE-R activities that comprise teacher- versus child-managed instruction are provided in Table 1 and Appendix Fig. A1. All activities were teacher managed except for independent silent reading, independent oral reading, and independent writing/publishing.

Code versus meaning focused: Code-focused activities included concepts of print, phonological awareness, alphabetic knowledge, word study/phonics, spelling, and letter- and word-level fluency. Meaning-focused activities included oral language development/discussion, sentence- and passage-level fluency, text reading, comprehension, and writing.

Each recorded activity was assigned to a quadrant (see Table 1). Teacher-managed (TM) code- and meaning-focused activities and CM meaning-focused activities were identified. No CM code-focused activities were recorded but, again, this was probably an artifact of the ICE-R observation system. Based on observer report, CM code-focused activities were typically observed while the teacher worked with a small group of children.

Differentiated instruction: This rating was a global assessment of instruction based on the extent to which teachers provided differentiated instruction systematically and explicitly with concern for students who were struggling with reading. A teacher in a classroom who received a rating of 4 would have been observed consistently demonstrating behaviors such as using direct and explicit language, modeling examples for students, providing opportunities for practice, responding with immediate corrective feedback, and using appropriate pacing based on student needs. A teacher who demonstrated these behaviors inconsistently would receive a high-average (3) or low-average (2) rating. The rating of weak (1) was assigned when instruction was dominated by indirect, implicit language, lack of models or demonstrations, absence of practice opportunities, ineffective or no feedback, or poor pacing without regard to individual student needs. Observers were trained to base ratings only on observable behaviors related to classroom instruction and to assign overall ratings according to the mode of scores given during the classroom observation.

Engagement: Observers used the ICE-R to assess classrooms with regard to students' engagement. Student engagement was rated on a three-point scale. The highest rating of 3 was assigned if almost all students were actively participating in the learning activity. If most students were actively participating a rating of 2 was assigned. The lowest score of 1 was given if more than half of the students were off task and not participating in the instructional activity. In the ICE-R system, students who were following along but not necessarily vocally participating were considered participating. Observers rated student engagement for each instructional activity. Scores reported are the average rating across literacy activities weighted for the amount of time spent in each activity. In this way, engagement scores for activities lasting 15 minutes contributed three times as much to the mean score compared to activities lasting 5 minutes. Engagement ratings recorded for nonliteracy activities were not included in this score.

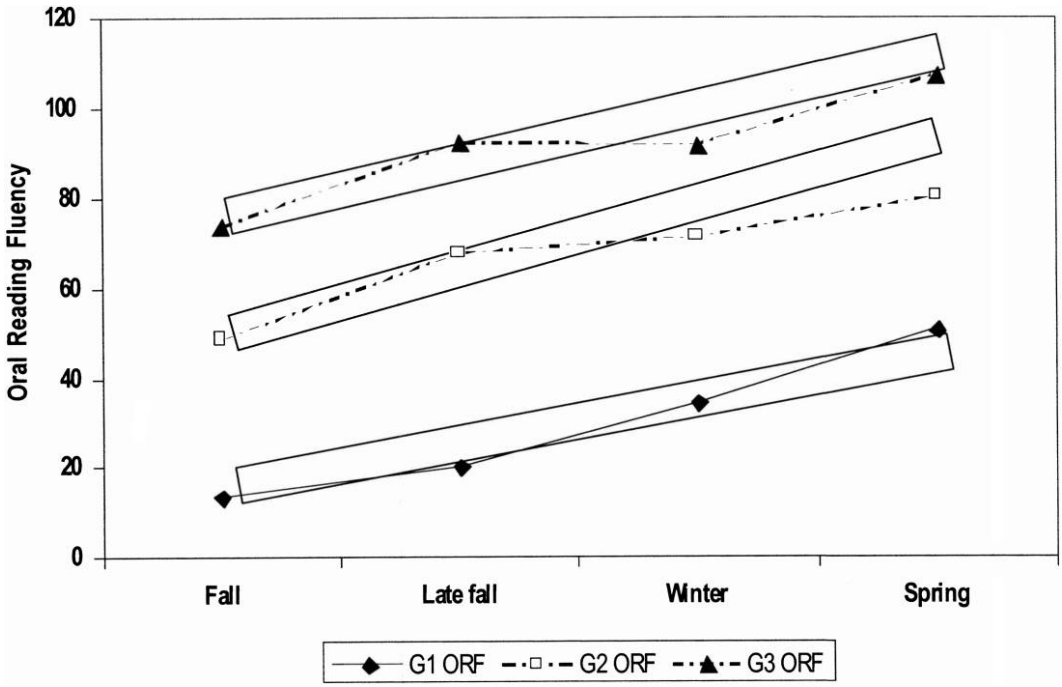

FIG. 1.—Growth in mean ORF scores from fall to spring for the first, second, and third grade. Outlined rectangles represent the 2005–2006 Florida benchmarks for grade-level performance expectations for each grade.

Results

Overall, students made progress during first, second, and third grade in their fluent text reading scores as evidenced by improvements in their ORF scores from fall to spring (see Fig. 1). First graders demonstrated an improvement of 38 words per minute, on average, between fall and spring. In second grade, students improved by 32 words per minute, and in third grade, students gained 34 words per minute, on average, between fall and spring. Based on Florida standards, this represents acceptable levels of growth in first grade (33-point gain expected) and third grade (33-point gain expected) but not in second grade (46-point gain expected). In terms of absolute achievement, children were performing at grade level at the beginning through the end of first grade, on average (ORF of 40–64 by spring is considered at first-grade level). However, second graders were below grade level (ORF of 90–108 by spring is considered second-grade level). Third graders were achieving below grade expectations, on average, as well (ORF of 110–128 by spring is considered third-grade level).

Similar findings for percentile scores on the SAT-10 reading comprehension measures revealed that in first grade, on average, children scored at the fifty-first percentile ($SD = 27$), whereas in second and third grade, children were falling behind national standards with SAT-10 scores at the forty-sixth percentile for both grades ($SD = 26$ and 27, respectively). Note the substantial variability at all grades as reflected by the standard deviations. Overall, children in these RF classrooms were achieving grade expectations in oral reading fluency and reading comprehension only for first grade.

Instruction

As can be seen in Table 4, amounts and types of instructional strategies, using the

TABLE 5. Descriptive Statistics for Classroom-Level Instructional Variables in Minutes Observed

Instruction Type	First Grade (n = 31 teachers)	Second Grade (n = 32 teachers)	Third Grade (n = 32 teachers)
Teacher managed, code focused:			
Mean	13.68	7.47	5.55
SD	9.36	9.39	7.72
Minimum	.00	.00	.00
Maximum	33.00	34.00	23.00
Teacher managed, meaning focused:			
Mean	25.55	28.28	34.68
SD	11.98	10.61	13.25
Minimum	3.00	.00	14.0
Maximum	48.00	44.00	83.00
Child managed, meaning focused:			
Mean	3.19	2.13	1.21
SD	7.25	3.08	2.26
Minimum	.00	.00	.00
Maximum	38.00	10.00	7.00
Engagement during literacy activities:			
Mean	2.70	2.64	2.72
SD	.37	.47	.59
Minimum	1.79	1.00	1.00
Maximum	3.00	3.00	3.00
Extent to which teachers differentiated instruction:			
Mean	2.94	3.03	3.19
SD	.727	.782	.592
Minimum	2	2	2
Maximum	4	4	4

ICE-R major categories, varied systematically by grade, with proportionally more time spent in word study and phonics and text reading in first grade, with amounts decreasing, on average, in second and third grade, and time spent in comprehension activities increasing from first to second to third grade. Phonological awareness instruction amounts were significantly different from zero in first grade but not in second or third. Analysis of variance (ANOVA) revealed that only the amount of phonological awareness [$F(2, 92) = 6.505$, $p = .002$], word study and phonics [$F(2, 92) = 7.572$, $p = .001$], and comprehension [$F(2, 92) = 8.614$, $p < .001$] varied significantly across grades. For all other types of instruction, there were no significant differences in amounts by grade.

We then summed the times for each type of instruction for the quadrants (e.g., TM code focused, CM meaning focused, etc.; see Table 1) and conducted descriptive analyses (see Tables 1 and 5). Using hierarchical linear modeling (HLM, Raudenbush & Bryk, 2002) to accommodate the nested structure of our data (children nested in classrooms), we investigated the effect of amount and type of instruction on students' spring reading comprehension skills, controlling for observation length at level 2 and vocabulary standard score and fall DIBELS scores at level 1. Building the models systematically, we first examined the child variables at level 1 and child by child interactions (e.g., ORF $\times$ vocabulary) that contributed to reading comprehension scores for each grade and removed from the model any that were not significant unless they were theoretically important. We then added TM code-focused, TM meaning-focused, and CM meaning-focused amounts and relevant classroom by classroom interactions at level 2 and tested interactions with child-level variables. All continuous variables were centered at their sample grand mean. We next added differentiation and engagement scores and again tested child and instruction interactions. With the exception of the level 2 in-

TABLE 6. Grade 1 HLM Final Model for Spring Reading Comprehension

Fixed Effect	Coefficient	Standard Error	*p*-value(*df*)
Intercept	554.62	2.54	<.001 (24)
Child-level variables:			
Fall letter naming fluency raw score	.49	.14	<.001 (472)
Fall oral reading fluency raw score	3.32	.56	<.001 (472)
Vocabulary standard score	.91	.15	<.001 (29)
Vocabulary by oral reading fluency	−.02	.00	<.001 (472)
Classroom-level variables:			
Observation length (minutes)	−.50	1.01	.627 (24)
Teacher managed, code focused	.12	.54	.822 (24)
Teacher managed, meaning focused	.09	.50	.861 (24)
Child managed, meaning focused	1.15	.47	.022 (24)
Engagement (1–3)	16.61	8.10	.051 (24)
Differentiation (1–4)	5.62	4.08	.181 (24)
Cross-level interactions:			
Letter naming fluency by teacher managed, meaning focused	.02	.14	.001 (472)
Oral reading fluency by teacher managed, meaning focused	−.10	.05	.031 (472)
Vocabulary by teacher managed, meaning focused	.002	.01	.821 (29)

Random Effects	Variance Component	χ^2(*df*)	*p*-value
Classroom level (u_0)	114.46	74.32 (24)	<.001
Child level (*r*)	1095.64		

NOTE.—Deviance = 4818.80. In models with only child-level variables, there was significant variability in vocabulary scores between classrooms, so this random effect was kept in the final model.

struction variables that were theoretically important, variables that did not contribute to the model were removed from the model to preserve parsimony. An exemplar model is provided in Appendix C.

First grade. Nature and variability of instruction: Teachers typically spent most of their language arts instruction time in TM meaning-focused activities, including supported choral reading and comprehension monitoring, on average about 25 minutes per day (see Table 5), but this varied widely. Teachers who spent more time in TM meaning-focused activities were less likely to spend time in TM code-focused activities ($r = -.80$, $p < .001$).

In first grade, the mean differentiated instruction rating was 3.06 ($SD = .727$), with a majority of teachers receiving ratings of 3 (see Table 5). ANOVA revealed that mean ratings did not vary significantly by grade [$F(3, 123) = .661$, $p = .578$]. The mean student engagement score was 2.7 (highest possible was a score of 3, see Table 5), which suggests that in most classes almost all students were participating in literacy activities. Teachers who were judged to have highly engaged classrooms were also more frequently observed differentiating instruction for children at risk for reading difficulties ($r = 503$, $p < .001$).

Relation of child variables and instruction to comprehension: HLM revealed that the relations among child variables as well as between child variables and instruction were highly complex. Results of the final model are provided in Table 6 with details provided in Appendix C. Overall, this model explained 49% of the variability in children's first-grade reading comprehension scores. It explained 74% of the classroom-level variance and 43% of the child-level variance (unconditional model variance; child level (r) = 1912.96, classroom level (u) = 453.07). Variance explained is computed by subtracting the full model child and/or classroom variance from the corresponding variance in the unconditional model and dividing by the unconditional model variance. The unconditional model revealed an intraclass correlation (ICC) of .19. An ICC of .19 indicates that 19% of the variation in children's comprehension

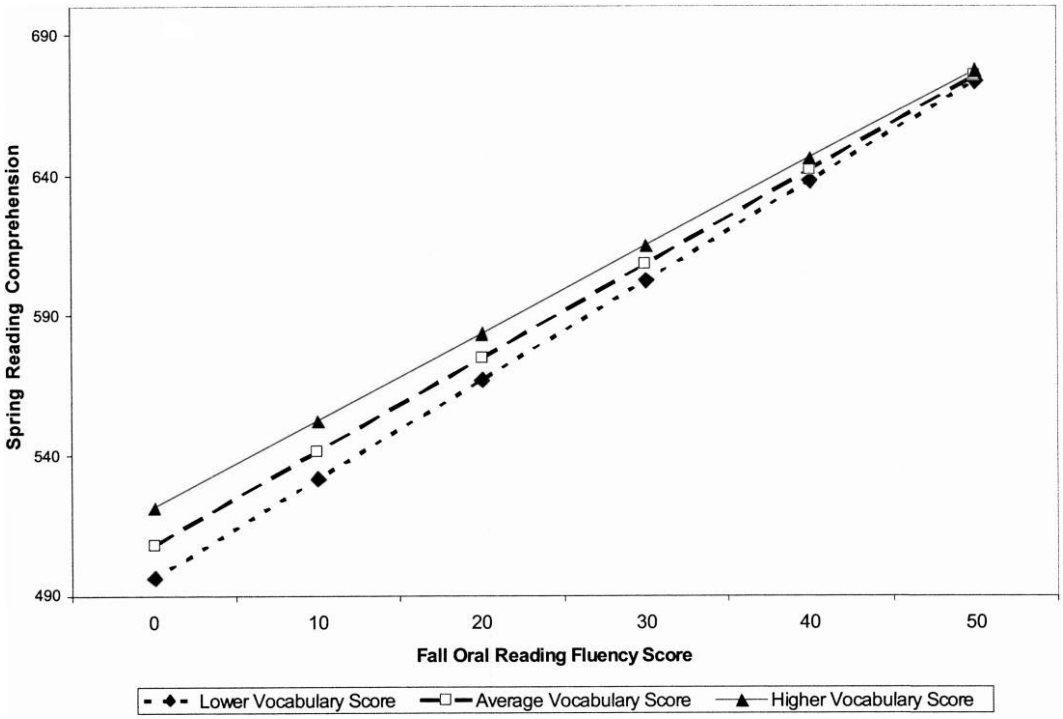

FIG. 2.—First-grade oral reading fluency by vocabulary interaction with vocabulary modeled at standard scores 75 (representing the approximate mean lower quartile of the sample), 100, and 115 (representing the approximate mean upper quartile of the sample).

is due to classroom membership and was computed from the unconditional model by dividing the classroom-level variance by the total variance.

As anticipated, children with stronger vocabulary, fall LNF, and fall ORF scores had higher spring reading comprehension scores than did children with lower scores, on average. There was also a vocabulary × fall ORF interaction (see Fig. 2). Generally, higher ORF scores mitigated, to some extent, the negative effect of lower vocabulary scores on spring reading comprehension. In other words, children with lower vocabulary scores could achieve reading comprehension scores that were similar to children with higher vocabulary scores if they had higher ORF scores. However, if they had lower ORF scores in the fall, they achieved substantially lower spring reading comprehension scores compared to children with similar ORF but higher vocabulary scores.

The effect of instruction on children's reading comprehension was also highly complex (see Table 6 and Fig. 3). Generally, children who were in classrooms that provided high amounts of CM meaning-focused instruction achieved stronger reading comprehension scores compared to children in classrooms with less CM meaning-focused instruction. More classroom time in TM meaning-focused instruction was related to higher reading comprehension scores for children with lower fall ORF scores but was related to slightly lower reading comprehension scores for children with higher fall ORF (see Fig. 3).

Differentiated instruction ratings did not significantly predict reading comprehension, but we kept the predictor in the model because it was theoretically impor-

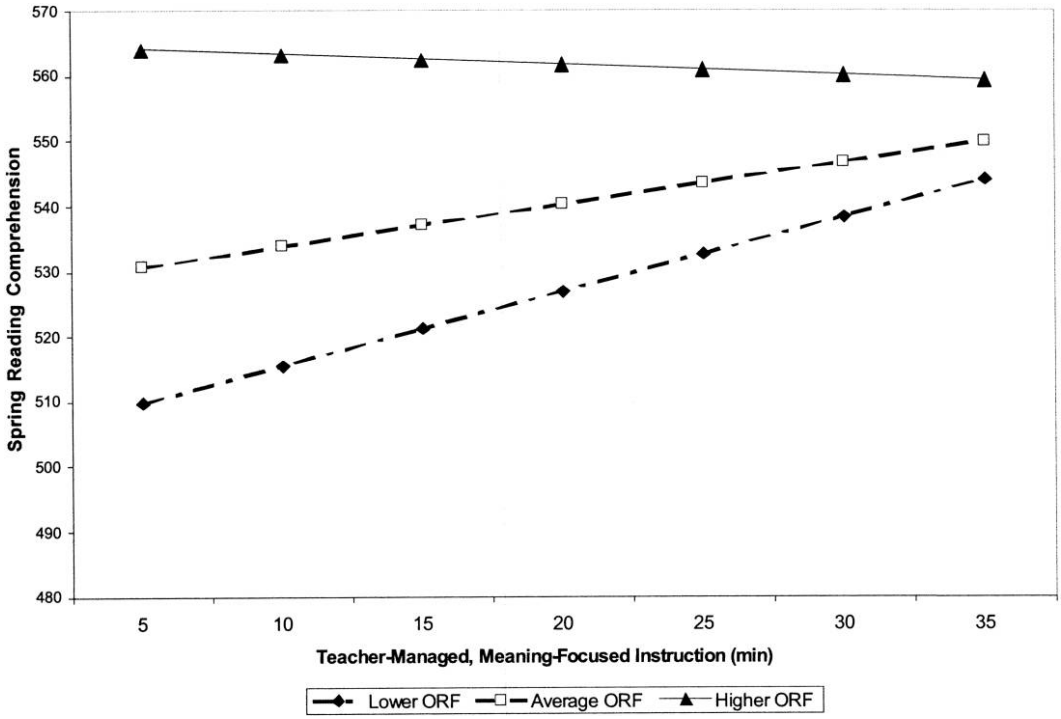

FIG. 3.—First-grade teacher managed, meaning-focused interaction with ORF modeled at the twenty-fifth, fiftieth, and seventy-fifth percentiles of the sample. There are no available norms for fall first-grade ORF.

tant. Engagement scores were positively related to students' reading comprehension scores. In considering the effect of these scores, it is important to keep in mind that engagement scores were entered into the model centered at the sample grand mean of 2.7. This means that a score of 3, the highest possible, predicted a reading comprehension score increase of 4.9, which is a small effect ($d = .10$). An engagement score of 1.7 (1 point below the mean of 2.7) predicted a reading score that was 16 points below the mean. This represents an effect size (d) of $-.33$, which is small to moderate.

Second grade. Nature and variability of instruction: In second grade, although varying widely, teachers spent the largest proportion of their language arts instruction time in TM meaning-focused activities, providing an average of 28 minutes per day (see Table 5). On average, teachers spent about 7.5 minutes per day in TM code-focused activities, such as application of word knowledge and spelling, which again varied widely. Similar to first grade, teachers spent the least time in CM meaning-focused activities (2.1 minutes) on average, including independent writing and independent silent reading. Teachers who provided more TM meaning-focused instruction tended to provide less TM code-focused instruction ($r = -735$, $p < .001$).

The mean differentiated instruction rating was 3.03 ($SD = .782$), with ratings ranging from 1 to 4. As with first grade, a majority of teachers received ratings of 3. The mean engagement score across all literacy activities for each teacher was 2.6, indicating that almost all students were engaged during literacy activities for most classes. Again, teachers who had high engagement scores also tended to be observed more frequently differentiating instruction for children at risk for reading difficulties ($r = .595$, $p < .001$).

Relation of child variables and instruc-

TABLE 7. Grade 2 HLM Final Model for Spring Reading Comprehension

Fixed Effect	Coefficient	Standard Error	*p*-value(*df*)
Intercept	590.95	.98	<.001 (23)
Child-level variables:			
Fall oral reading fluency raw score	.84	.06	<.001 (29)
Vocabulary standard score	.64	.07	<.001 (521)
Classroom-level variables:			
Observation length (minutes)	−.04	.19	.850 (23)
Teacher managed, code focused	2.20	.85	.017 (23)
Teacher managed, meaning focused	.017	.17	.334 (23)
Child managed, meaning focused	−.01	.31	.968 (23)
Engagement (1–3)	−2.90	1.94	.150 (23)
Engagement by teacher managed, code focused	−.93	.40	.027 (23)
Differentiation (1–4)	−.95	1.32	.479 (23)
Cross-level interactions:			
Oral reading fluency by engagement	−.01	.10	.898 (29)
Vocabulary by engagement	.64	.17	<.001 (521)

Random Effects	Variance Component	χ^2(*df*)	*p*-value
Classroom level (u_0)	.653	18.54	>.50
Oral reading fluency (u_1)	.06	55.75	.002
Child level (r)	630.27		

NOTE.—Deviance = 4969.50. In models with only child-level variables, there was significant variability in oral reading fluency scores between classrooms, so this random effect was kept in the final model.

tion to comprehension: Similar to first grade, HLM revealed that relations among child variables and instruction were highly complex. The intraclass correlation was .07, indicating that 7% of the variance fell between classrooms. There was some variability observed at the classroom level and substantial variability at the child level ($u_0 = 88.39$, $r = 1253.67$). Results of the final model are provided in Table 7. On the whole, the final model explained 53% of the variability in children's second-grade reading comprehension scores. This model explained 99% of the classroom-level and 50% of the child-level variance. Overall, children with stronger vocabulary standard scores and fall ORF scores had higher spring reading comprehension scores on average than did children with lower ORF and vocabulary standard scores. Moreover, there were two significant interactions, both of which involved engagement. First, there was a TM code-focused instruction by engagement interaction (see Table 8 and Fig. 4). On average, children whose teachers provided more TM code-focused instruction achieved stronger reading comprehension scores compared to children whose teachers provided less TM code-focused instruction. Further, this effect was greater in classrooms that had lower engagement ratings.

The second significant interaction was engagement by vocabulary (see Fig. 5). Generally, children with weaker vocabulary scores in higher-engagement classrooms earned lower reading comprehension scores than such children did in lower-engagement classrooms. At the same time, children with stronger vocabulary scores in high-engagement classrooms earned higher reading comprehension scores than they did in lower-engagement classrooms.

Third grade. Nature and variability of instruction: Virtually the entire language arts block in third-grade classrooms was devoted to TM meaning-focused activities, on average 35 minutes per day (see Table 5), and this varied widely. Overall, most of this time was spent in comprehension monitoring, vocabulary, comprehension strategies, and supported oral reading.

TABLE 8. Grade 3 HLM Final Model for Spring Reading Comprehension

Fixed Effect	Coefficient	Standard Error	*p*-value(*df*)
Intercept	616.49	1.10	<.001 (25)
Child-level variables:			
Fall oral reading fluency raw score	.66	.05	<.001 (25)
Vocabulary standard score	.83	.09	.038 (527)
Classroom-level variables:			
Observation length (minutes)	−.71	.25	.010 (25)
Teacher managed, code focused	.37	.24	.147 (25)
Teacher managed, meaning focused	.42	1.74	.092 (25)
Child managed, meaning focused	.19	.37	.623 (25)
Engagement (1–3)	−8.10	1.96	<.001 (25)
Differentiation (1–4)	6.34	2.06	.006 (25)
Cross-level interactions:			
Oral reading fluency by engagement	−.03	.11	.734 (30)
Vocabulary by engagement	.37	.18	.038 (527)
Random Effects	**Variance Component**	$\chi^2(df)$	***p*-value**
Classroom level	9.58	39.90	.030
Oral reading fluency	.01	44.04	.047
Child level	539.60		

NOTE.—Deviance = 4962.27. In models with only child-level variables, there was significant variability in oral reading fluency scores between classrooms, so this random effect was kept in the final model.

Substantially less time (5½ minutes per day, on average) was spent in TM code-focused activities. With the exception of activities involving teaching letter-sound relationships (3 minutes), no code-focused activity was taught consistently by all teachers. Again, teachers providing more TM code-focused instruction tended to provide less TM meaning-focused instruction ($r = -.403$). Only about 1 minute of CM meaning-focused activity per day was observed. No particular activity, such as sustained silent reading, appeared to characterize this time.

In third grade, on average, teachers earned differentiated instruction ratings of 3.19 out of 4. The overall mean engagement score was 2.8 out of 3, suggesting that virtually all of the classrooms were judged to deserve a high engagement score during literacy activities. Again, teachers who were rated high on engagement also tended to be rated as differentiating instruction more frequently ($r = .595$, $p < .001$).

Relation of child variables and instruction to student reading comprehension: In third grade, the intraclass correlation was .13. There was substantial variability at the classroom and child levels in the unconditional model ($U_0 = 161.25$, $r = 1108.45$, respectively). Results of the final model are provided in Table 8. Overall, this model explained 57% of the variance in students' reading comprehension scores.

As we found in first and second grade, children with stronger fall ORF and vocabulary scores achieved, on average, stronger spring reading comprehension scores compared to children with lower fall ORF and vocabulary scores. In general, in classrooms where teachers were observed to differentiate instruction more, students achieved higher comprehension scores than did students in classrooms with less differentiated instruction. Moreover, student engagement negatively predicted spring reading comprehension and interacted with vocabulary (see Fig. 6). Generally, the more students in the classroom were judged to be on task, the weaker was their spring reading comprehension, and this effect was greater for students with

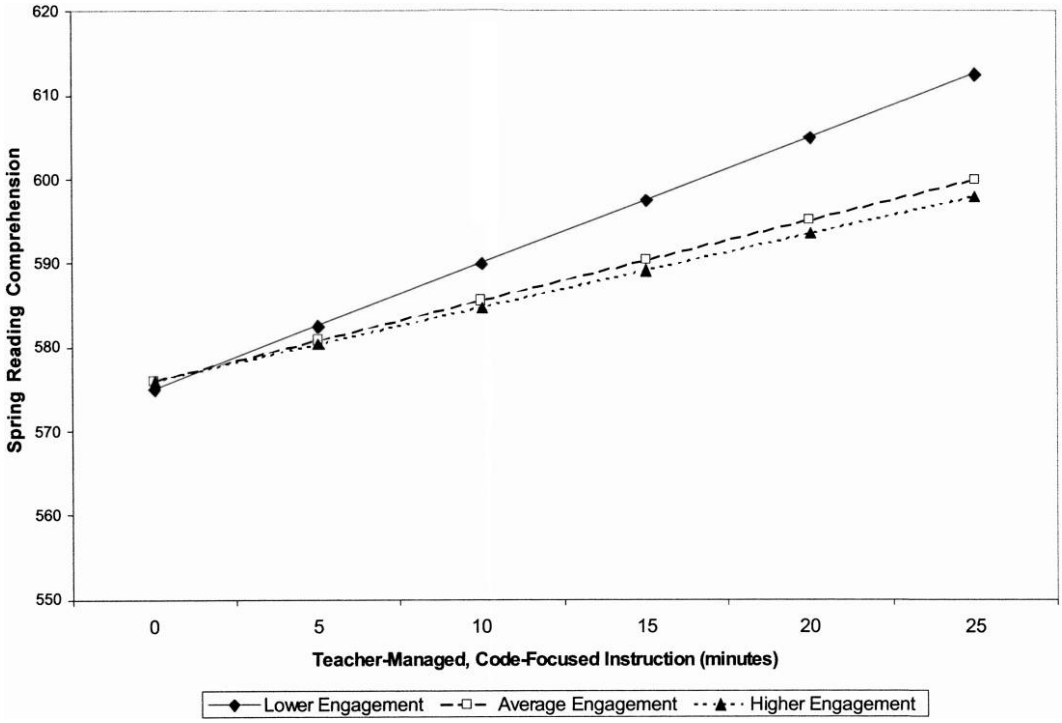

FIG. 4.—Second-grade engagement by teacher managed, code-focused interaction with engagement modeled at the twenty-fifth, fiftieth, and seventy-fifth percentiles of the sample.

higher vocabulary scores than for those with lower vocabulary scores.

Discussion

Child Outcomes

Only in first grade did students meet national norms after the initial year of RF and increase their oral reading fluency (ORF) scores from late fall to winter. From winter to spring, first graders maintained this rate of growth and, on average, exceeded grade-level performance expectations. Generally, students failed to reach state benchmarks and national achievement norms in second grade and national achievement norms in third grade. What is particularly discouraging is that in second and third grade, children typically began the year with grade-level ORF scores and demonstrated strong growth through the late fall (see Fig. 1). From late fall to winter, however, growth stagnated in both grades. Rates improved again from winter to spring for third graders but not for second graders.

Second and third graders achieved SAT-10 reading comprehension scores at the forty-sixth percentile and below national norms for grade-level expectations. In contrast, first graders were performing at the fiftieth percentile on the SAT-10, a notable achievement for children attending historically underperforming schools.

Reading First and Classroom Instruction

Our hypothesis that RF would influence teachers' practices with regard to amount and type of literacy instruction was supported. Classroom instruction guided by rigorous reading research would tend to provide more time in explicit and systematic teacher-managed instruction in phonological awareness, phonics, and reading

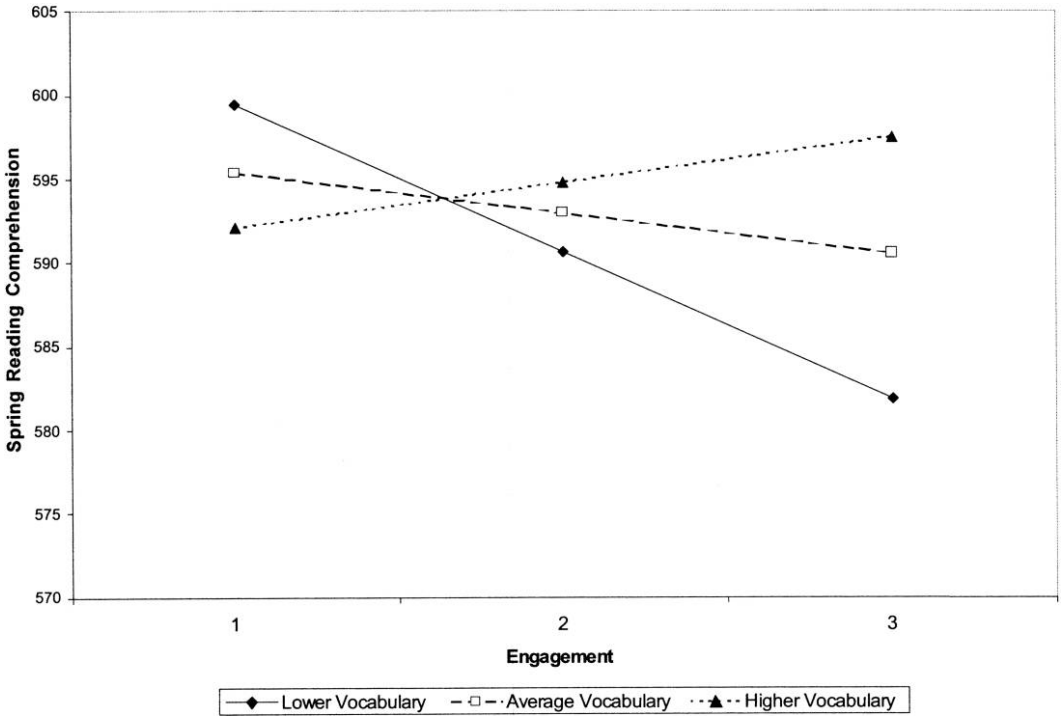

FIG. 5.—Second-grade vocabulary by engagement interaction with vocabulary modeled at the twenty-fifth, fiftieth, and seventy-fifth percentiles of the sample.

comprehension strategies (NRP, 2000), and this was observed. Moreover, when compared to classroom observations conducted prior to NCLB (e.g., Connor, Morrison, & Katch, 2004; NICHD-ECCRN, 2002), generally RF teachers spent more time in TM activities—both meaning and code focused—than did teachers in pre-NCLB classrooms.

Teacher-Managed, Code-Focused Instruction

More TM code-focused instruction was associated with stronger spring reading outcomes only in second grade. Specifically, in second grade, the more TM code-focused instruction children received, the greater were their spring reading comprehension skills, which replicates the second-grade findings of Connor, Morrison, and Underwood (2007). Our findings suggest that in second grade, children might have continued to make progress had they received more time in TM code-focused instruction. The finding for first grade is surprising inasmuch as TM code-focused instruction has been associated with stronger word reading and other decoding outcomes (Foorman et al., 1998; Rayner et al., 2001). Although our outcome was reading comprehension, especially in the early grades, students with stronger decoding skills also tend to have stronger comprehension skills. Possibly, children received so much TM code-focused instruction that a threshold was reached so that virtually all children received enough or more than enough of this type of instruction. Models based on child by instruction research, which compute recommended amounts of specific types of instruction that predict stronger student outcomes, based on children's language and literacy skills, are nonlinear (Connor & Morrison, 2006b; Connor,

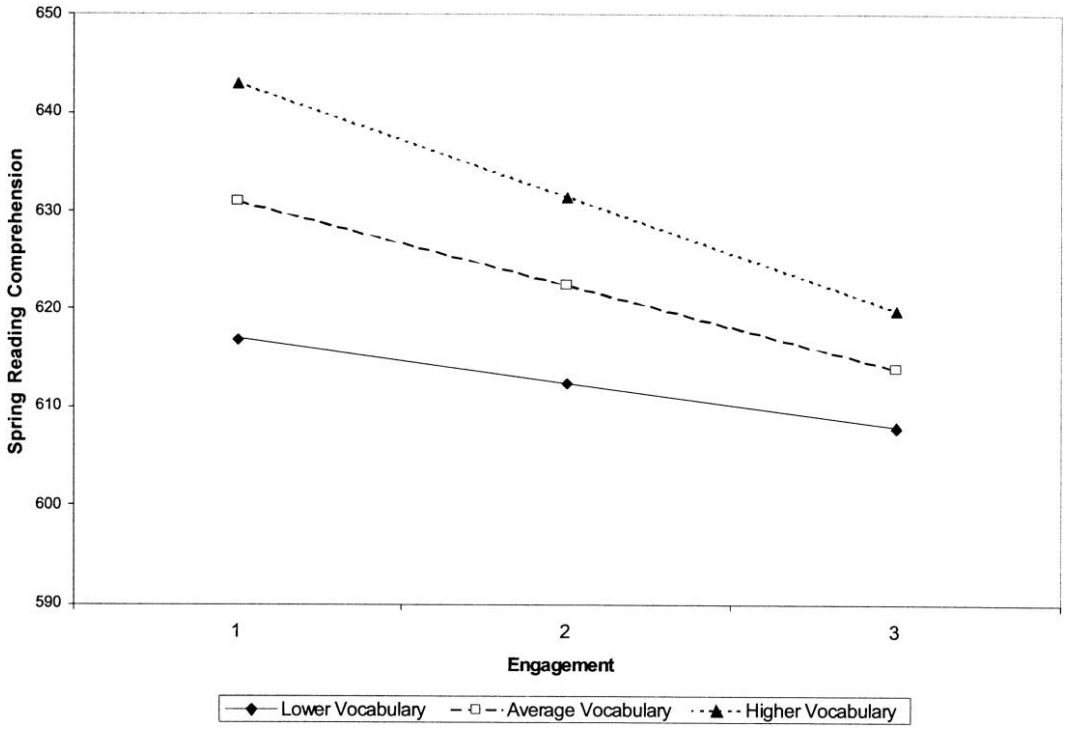

FIG. 6.—Third-grade vocabulary by engagement interaction with engagement modeled at the twenty-fifth, fiftieth, and seventy-fifth percentiles of the sample.

Morrison, Fishman, Schatschneider, & Underwood, 2007). This suggests that there may be diminishing returns for children reading close to grade level, beyond which more time in specific instructional activities does not contribute to students' skill development.

Teacher-Managed, Meaning-Focused Instruction

Our hypothesis that more time in TM meaning-focused instruction would be associated with stronger reading comprehension skills (Connor, Morrison, & Katch, 2004; NRP, 2000) was supported, but not completely. In third grade, there was a trend ($p < .10$) that children in classrooms with more TM meaning-focused instruction achieved stronger spring reading comprehension outcomes. In a more complex relation, first graders in classrooms with more TM meaning-focused instruction demonstrated stronger reading comprehension skills by spring, but only if they showed typical to below-average ORF skills in the fall. For first graders whose ORF skills were well above average, TM meaning-focused instruction had no substantial effect on their outcomes. Other studies have produced similar or stronger findings (Connor, Morrison, & Katch, 2004; NRP, 2000; Pressley & Wharton-McDonald, 1997; Snow, 2001). It is interesting that, before we added the variable differentiated instruction to the model, TM meaning-focused instruction did significantly and positively predict students' reading comprehension outcomes (coefficient = 0.69, $t(26) = 2.50$, $p = 0.019$). We conjecture that total amount of instruction, although important, appears to be less important than tailoring instruction to meet individual students' needs.

In the second grade, our hypothesis re-

garding TM meaning-focused instruction was not supported. Children may have been receiving enough of this type of instruction and passed a threshold. Also, the strategies used during TM meaning-focused instruction may not have been explicitly focused on instruction and practice in comprehension strategies. For instance, most of the time in this type of instruction in the third grade was spent in teacher read-aloud (8 minutes), which has not been strongly associated with gains in students' comprehension skills (Connor, Morrison, & Petrella, 2004).

Child-Managed, Meaning-Focused Instruction

Although little was observed, CM-meaning-focused instruction had an important effect on students' reading comprehension outcomes in first grade but not in second or third grade. First graders in classrooms of teachers who spent more time in CM meaning-focused activities demonstrated stronger reading outcomes. These activities include repeated reading of books as well as independent reading and writing skill activities. Research suggests that these activities and peer-assisted reading support fluency (Kuhn & Stahl, 2002; NRP, 2000; Therrien, 2004), which in turn support reading comprehension (Fuchs & Fuchs, 2005). Accumulating evidence appears to confirm the importance of these activities for early readers.

Engagement and Differentiated Instruction

The findings regarding the positive effect of differentiated instruction are as hypothesized and consistent with accumulating evidence of the effect of child by instruction interactions on elementary students' outcomes (e.g., Connor & Morrison, 2006a; Connor, Morrison, Fishman et al., 2007; Connor, Morrison, & Katch, 2004; Connor, Morrison, & Slominski, 2006; Foorman et al., 1998; Juel & Minden-Cupp, 2000) and with the effective schools literature (Pressley et al., 2001; Taylor et al., 2000; Wharton-McDonald et al., 1998). Across all grades, when teachers differentiated instruction, students tended to be more on task and, in third grade, differentiated instruction positively predicted students' reading comprehension scores. Overall, these findings suggest that the extent to which teachers differentiate instruction should have a generally positive effect on student academic outcomes and on behavior.

In the present study we found students to be generally on task during literacy activities in RF classrooms. Interestingly, we found support for our hypothesis that engagement would relate positively to student outcomes in the first grade but not in the second or third grade. Specifically, engagement demonstrated an interacting effect in grades 2 and 3. For second graders with lower vocabulary skills, classrooms rated as highly engaged were associated with weaker reading comprehension skills by spring. In third grade, generally higher engagement ratings were associated with lower spring reading comprehension scores, and this effect was greater for children with stronger vocabulary skills. Accumulating evidence strongly indicates that student engagement is a critical component of effective learning (Guthrie et al., 2004), so these negative findings are perplexing. In other studies as well as ours, students' on-task behavior has been an inconsistent predictor of achievement across grades (Karweit & Slavin, 1981). Perhaps students' on-task behavior is not a good proxy for their cognitive engagement because, in this study, it clearly failed to capture important aspects (cognitive, motivational, and emotional) of students' engagement during literacy activities.

Another explanation for our findings is that as early as second grade, students may learn how to "do school"—especially students who are struggling academically. Specifically, they learn how to appear to pay attention (i.e., behavioral engagement) in school during activities related to in-

TABLE 9. Summary of Coefficients for Instructional Variables, across Grades, as They Relate to Spring Reading Comprehension

	First Grade		Second Grade		Third Grade	
	Lower ORF	Higher ORF	Lower Vocabulary	Higher Vocabulary	Lower Vocabulary	Higher Vocabulary
Teacher managed, code focused			2.20	2.20		
Teacher managed, meaning focused	++	−				
Child managed, meaning focused	1.15	1.15				
Engagement	16.61	16.61	++	+	−	−−
Teacher managed, code focused (by engagement)			Less ENG −−	More ENG ++		
Differentiated instruction					6.33	6.33

NOTE.—Results obtained when controlling for vocabulary and fall scores (e.g., oral reading fluency, letter naming fluency). The general direction of interactions for students with higher (standard score = 115) and lower (standard score = 75) vocabulary, oral reading fluency (ORF), or engagement (ENG) scores are indicated by pluses (generally positive effect), minuses (generally negative effect), or 0 (no significant effect). All effects, $p \leq .05$.

struction but are not actively and cognitively engaged in learning (e.g., Bloome, 1983; Grek, 2000).

Child by Instruction Interactions

The findings of this study are highly complex inasmuch as we found a child characteristic by child characteristic interaction (child × child; ORF by vocabulary interaction in first grade), an instruction type by instruction type interaction (instruction × instruction; TM code-focused by engagement interaction in second grade), and child by instruction interactions in all three grades (see Table 9). It is the latter in which we are most interested because there are important theoretical and practical implications (Connor, Morrison, Fishman et al., 2007). These findings add to the accumulating evidence that the effect of instructional strategies depends on each child's language and literacy skills. Indeed, for two children in the same classroom, effective instruction for the student with strong vocabulary and reading skills differs sharply from effective instruction for the student with weaker skills. This theoretical view emphasizes the importance of progress monitoring, including curriculum-based assessment (Deno et al., 2002) and using assessment to guide instruction content and strategy and amount of instruction.

The first-grade ORF by vocabulary interaction is intriguing and raises interesting questions about the role of fluency and vocabulary in the prediction of reading comprehension. Whereas researchers generally agree that fluency and comprehension have a reciprocal relation (Therrien, 2004), the role of vocabulary and oral language in this complex relation is relatively unstudied. At all grades, ORF and vocabulary uniquely predicted reading comprehension and interacted with specific instructional strategies in first and second grade.

Limitations of the Study

There are limitations to our study, and thus these results should be interpreted carefully. The Progress Monitoring Reporting Network and the site-visit observations were not designed and implemented specifically for this study. As with any secondary analysis of existing data, there are limitations that cannot be rectified. For example, there was only one observation to capture an entire school year, and only half of the 90-minute reading block was observed. Watching a classroom for 45 minutes in the spring clearly cannot fully capture the complexity of

classroom instruction and the important interactions that occur between teachers and students. However, with this protocol, we observed 95 classrooms across Florida between March 30 and May 20, 2004.

Limitations are evident in the ICE-R coding protocol used in 2003–2004. Specifically, child-managed activities were certainly underreported in this study as an artifact of the observation system. Nevertheless, the ICE-R results were reliable and sensitive enough to reveal systematic relations among instructional strategies and students' reading comprehension outcomes.

We have some concerns regarding multicollinearity. In the first and second grade, TM meaning-focused and TM code-focused instruction were strongly negatively correlated with a more moderate negative correlation in third grade. Teachers who provided more TM code-focused instruction were more likely to use less TM meaning-focused instruction and vise versa. Multicollinearity tends to lower the size of the coefficients of correlated variables because they cancel each other, so to speak. By centering the variables in our HLM analyses, we were able to mitigate the effect of multicollinearity to some extent but not completely. We ran all models deleting one of the highly correlated instruction variables, and in all instances, there was some change in the size of the remaining coefficient, but the major results remained unchanged.

The strong negative correlation between TM meaning-focused and TM code-focused instruction may be a phenomenon represented by different core curriculum priorities, teaching preferences, or theoretical perspectives. Alternatively, this correlation may have occurred because we observed only half of the reading blocks, or this could have been an artifact of the observation method. Unfortunately, it is not possible to separate these different explanations with the current data. Nonetheless, many of our findings are consistent with or extend existing research findings and, taken in the spirit of identifying and clarifying future research questions, are interesting and informative.

Changing Classroom Practices

Taking the most optimistic view, Reading First in Florida appears to be changing classroom instructional practices and, at least for first graders, contributing to improved student reading comprehension outcomes. Although these changes cannot be attributed only to the policies and requirements of RF and NCLB, still, the literacy instruction observed in these classrooms reflects the types of practices that current literacy research suggests should be most effective in promoting students' reading comprehension outcomes. Our results for second and third graders, however, demonstrate that there is much to learn about how to support children's literacy development.

Our findings indicate that patterns of instruction that are effective appear to depend not only on students' language and literacy skills but also on the grade of interest (see Table 9). An instructional focus that is important, or assumptions about what constitutes effective instruction in one grade, may change in another grade. Overall, the literature generally provides fewer insights for second grade compared to first grade or third grade. Consequently, the lack of rigorous research on second-grade instruction might lead to limitations in breadth, depth, and focus in the core curricula used in this grade. Furthermore, children in this grade may never make up a year of ineffective instruction (Connor, Morrison, & Underwood, 2007; Sanders & Horn, 1998).

A potential weakness of RF is an overreliance on the core curriculum to support research-guided instruction. Even the best of these curricula have not met the rigorous requirements of random field trials at all grades and with all levels of readers (e.g., Foorman et al., 1998, which used a strong quasi-experimental design but reported re-

sults only for Title 1 students). Moreover, because it takes so long to publish any core curriculum and to make it available to teachers, no curriculum can fully integrate the most current research. Much has occurred since the National Reading Panel report in 2000 (Pressley, 2006). Recently, researchers have produced more information about the importance of morphosyntactic awareness (Carlisle, 2000), motivation (Guthrie et al., 2004), parenting and preschool (Morrison et al., 2005), and child characteristic by instruction interactions. The state of educational research is changing rapidly. Only by becoming active consumers of rigorous research will teachers and the administrations that support them ensure that they are using the most effective instructional strategies for their students.

Additionally, changing and expanding teacher practice is challenging, and we examined only the first year of implementation of RF. Research suggests that, even with intensive classroom-based professional development (e.g., literacy coaches), it takes 2 to 5 years for most teachers to fully adopt new instructional approaches (Sadoski & Willson, 2006), especially if they run counter to teachers' preservice training (Klingner, Vaughn, Arguelles, Hughes, & Leftwich, 2004). Unfortunately, as the National Council on Teacher Quality (Walsh, Glaser, & Wilcox, 2006) reveals, only 11% of randomly selected teacher education programs appeared to incorporate the latest reading research in their teacher-preparation courses.

All in all, the results of this study paint an encouraging picture of instruction based on research. These findings also highlight how complex implementing effective instruction is when children in the same classroom respond to different types and amounts of instruction. Even with the extra resources, training, progress monitoring, and oversight that Reading First provides, meeting the charge of No Child Left Behind, that all children read at grade level by the end of third grade, is a challenging albeit worthwhile goal.

Appendix A

Content Categories	1: Concepts of Print	2: Phonological awareness	3: Alphabetic knowledge	4: Word study/Phonics	5: Spelling
Descriptors	1. Concepts of print	1. Rhyming 2. Blending or segmenting sentences/ syllables. 3. Onset/rime 4. Blending or segmenting phonemes 5. Isolation tasks 6. Other	1. Letter identification and/or recognition. 2. Other	1. Letter/sound relationships 2. Provides opportunities for application of letter/sound knowledge to reading /writing/spelling. 3. Irregular words 4. Word reading 5. Integration of word study 6. Other	1. Spelling
Content Categories	6: Oral language development	7: Fluency	8: Text reading	9: Comprehension	10: Writing or language arts
Descriptors	1. Teacher initiated structured opportunities to talk with teachers/peers. 2. Expansion of student initiated language (incidental language strategies) 3. Other	1. Letter or sound naming fluency 2. Word fluency 3. Repeated reading of text 4. Other	1. Supported oral reading 2. Choral reading 3. Independent silent reading 4. Independent oral reading 5. Teacher reads aloud 6. Teacher reads aloud while students read along 7. Other	1. Vocabulary 2. Prior knowledge/predicting 3. Reading comprehension monitoring 4. Listening comprehension monitoring 5. Comprehension strategy instruction/use 6. Other	1. Shared writing 2. Writing composition 3. Independent writing/publishing 4. Grammar and punctuation 5. Handwriting instruction 6. Copying 7. Other

FIG. A1.—Instructional Content Emphasis, Revised (ICE-R) content categories. From Edmonds and Briggs (2003), adapted by Florida Center for Reading Research. Table reprinted with permission of author.

Appendix B

Time	Observational Record	A 1-10 Main instruction content category	B varies Sub-instruction category	Student Engagement 1-3	Differentiation 1-4
8:30-8:36	Small group with 5 students. T: Let's make some words. Write the first sound in "map" (white boards & markers). Write the next sound in "map". Tap it out. Say it fast. Now spell it. T: "Lip". Write the sounds. Break it down. Say it fast. Spell it. (Teacher repeated with other words: mat, nap.) Note: 10 centers with very actively engaged students at each. Paraprofessional helped with centers and listening word work.	4	2	3	4
8:36-8:39	Teacher uses magnetic board with small group. T: Come up and make this word: "mud". Point to the first sound. Point to the next sound. Point to the last sound. Now say each sound, put them together. Teacher repeated with other words, individuals take turns; unison response.	4	2	3	4
8:39-8:42	T: Let's read some words. (shows flashcards) Get Ready. What word? "a, me, here, like, have, my, do, go, the, one, little, I , had, on, some, see, are, a, we" Teacher called different groups efficiently and without loss of control.	7	2	3	4

FIG. B1.—Example of observation coding record. The main instruction content category (1-10) refers to the content areas in Appendix A, which provides the 10 ICE-R coding categories. Subinstruction categories are the types of instruction within the main instruction content category. For example, the first activity has been coded 4 for main content category, which is "word study/phonics." The subcategory, 2, indicates that the specific activity was "provides opportunities for application of letter/sound/knowledge to reading/writing/spelling."

Appendix C

The First-Grade Hierarchical Linear Model as an Exemplar

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{fall letter naming fluency}_{ij}) + \beta_{2j}(\text{fall oral reading fluency}_{ij}) + \beta_{3j}(\text{vocabulary}_{ij}) + \beta_{4j}(\text{vocabulary by oral reading fluency}_{ij}) + r_{ij}.$$

Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{observation length}_j) + \gamma_{02}(\text{TM-code-focused}_j) + \gamma_{03}(\text{TM-meaning-focused}_j) + \gamma_{04}(\text{CM-meaning-focused}_j) + \gamma_{05}(\text{engagement}_j) + \gamma_{06}(\text{level of differentiation}) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{TM-meaning-focused}_j)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(\text{TM-meaning-focused}_j)$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}(\text{TM-meaning-focused}_j) + u_3$$

$$\beta_{4j} = \gamma_{40} + \gamma_{41}(\text{TM-meaning-focused}_j).$$

Y_{ij}, which is the spring reading comprehension score for child i in class j, is a function of the respective coefficients (β_j) at level one as they pertain to child i's fall letter naming fluency, oral reading fluency, vocabulary, and vocabulary by oral reading fluency, as well as a residual (r_{ij}). β_{0j} is a function of the fitted mean spring reading comprehension score for the sample of students (γ_{00}) plus the effect of the classroom instruction variables for classroom j, plus error (u_j). Coefficients are interpreted in the same way as regression coefficients, including within- and cross-level interactions. Residuals (u, r) were assumed to be normally distributed with a mean of zero. Models deleting either TM code-focused or TM meaning-focused were tested for the first- and second-grade models because these variables were correlated. In these models, the overall results, including direction and significance of coefficients, remained the same, although the remaining TM instruction coefficients' values changed: In first grade, TM meaning focused = −.0077 after TM code focused was deleted, and in second grade, TM meaning focused = 1.757 after TM meaning focused was removed from the model.

Note

We thank Stephanie Al Otaiba and Joseph Torgesen for their helpful comments on this work and Courtney Rankin for her assistance. We are also grateful to the Reading First site-visit team and Steve Nettles at FCRR. We also wish to thank the children, parents, teachers, and school administrators without whom this research would not have been possible. This study was funded in part by the U.S. Department of Education, Institute for Education Sciences grant no. R305H04013; the State of Florida Project no. 371-2096A-6CF03 Specific Funding Authority–No Child Left Behind Act, 2001 Title 1, Part B, Subpart 1, Reading First CFDA Program no. 84.357A; and by the National Institute for Child Health and Human Development grant no. R01HD48539. The opinions expressed are ours and do not represent views of the funding agencies. Contact: Carol McDonald Connor, cconnor@fsu.edu.

References

Bloome, D. (1983). Classroom reading instruction: A socio-communicative analysis of time on task. *National Reading Conference Yearbook,* **32**, 275–281.

Bredekemp, S., & Copple, C. (Eds.). (1997). *Developmentally appropriate practice in early childhood programs.* Washington, DC: National Association for the Education of Young Children.

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology,* **22**, 723–742.

Campbell, J. R., Voelkl, K. E., & Donahue, P. L. (1997). *NAEP 1996 trends in academic progress.* Washington, DC: National Center for Education Statistics.

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading & Writing,* **12**(3–4), 169–190.

Chall, J. S. (1967). *Learning to read: The great debate.* New York: McGraw-Hill.

Christian, B., & Bloome, D. (2004). Learning to read is who you are. *Reading and Writing Quarterly,* **20**, 365–384.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research.

Educational Evaluation and Policy Analysis, **25**(2), 119–142.

Connor, C. M., & Morrison, F. J. (2006a, May). *Individualizing student reading instruction: Planning and Implementation.* Paper presented at the annual meeting of the International Reading Association, Chicago.

Connor, C. M., & Morrison, F. J. (2006b, April). *Supporting teachers' efforts to use assessment to guide instruction.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). THE EARLY YEARS: Algorithm-guided individualized reading instruction. *Science,* **315**(5811), 464–465.

Connor, C. M., Morrison, F. J., & Katch, E. L. (2004). Beyond the reading wars: The effect of classroom instruction by child interactions on early reading. *Scientific Studies of Reading,* **8**(4), 305–336.

Connor, C. M., Morrison, F. J., & Petrella, J. N. (2004). Effective reading comprehension instruction: Examining child by instruction interactions. *Journal of Educational Psychology,* **96**(4), 682–698.

Connor, C. M., Morrison, F. J., & Slominski, L. (2006). Preschool instruction and children's literacy skill growth. *Journal of Educational Psychology,* **98**(4), 665–689.

Connor, C. M., Morrison, F. J., & Underwood, P. (2007). A second chance in second grade? The independent and cumulative impact of first and second grade reading instruction and students' letter-word reading skill growth. *Scientific Studies of Reading,* **11**(3), 199–233.

Connor, C. M., Son, S., Hindman, A., & Morrison, F. J. (2005). Teacher qualifications, classroom practices, family characteristics and preschool experience: Complex effects on first graders' vocabulary and early reading outcomes. *Journal of School Psychology,* **43**, 343–375.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology,* **33**, 934–945.

Dahl, K. L., & Freppon, P. A. (1995). A comparison of innercity children's interpretations of reading and writing instruction in the early grades in skills-based and whole language classrooms. *Reading Research Quarterly,* **30**(1), 50–74.

Deno, S. L., Espin, C. A., Fuchs, L. S., Shinn, M. R., Walker, H. M., & Stoner, G. (2002). Evaluation strategies for preventing and remediating basic skill deficits. In Anonymous (Ed.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 213–241). Washington, DC: National Association of School Psychologists.

Dickinson, D. K., & Tabors, P. O. (2001). *Beginning literacy with language.* Baltimore: Paul H. Brookes.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–III.* Circle Pines, MN: American Guidance Service.

Edmonds, M. S., & Briggs, K. L. (2003). Instructional Content Emphasis instrument. In S. R. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for observing teaching and learning.* Baltimore: Paul H. Brookes.

FCRR. (2007a). *DIBELS risk levels.* 2007 [On-line]. Available: http://www.fcrr.org/pmrn/pdffiles/k3_risklevels_sy0506.pdf

FCRR. (2007b). *FCRR technical reports* [On-line]. Available: http://www.fcrr.org/science/scienceTechnicalReports.htm

FCRR. (2007c). *Florida Center for Reading Research: Florida Assessment Plan* [On-line]. Available: http://www.fcrr.org/assessment/PDFfiles/Fl_Assess_Plan_Final.pdf

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and inattentive-withdrawn behavior and achievement among fourth graders. *Elementary School Journal,* **95**, 421–434.

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.

Fuchs, D., & Fuchs, L. S. (2005). Peer-assisted learning strategies: Promoting word recognition, fluency, and reading comprehension in young children. *Journal of Special Education,* **39**(1), 34–44.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for Development of Educational Achievement.

reenwood, C. R., Terry, B., Marquis, J., & Walker, D. C. (1994). Confirming a performance-based instructional model. *School Psychology Review,* **23**, 625–668.

Grek, M. L. (2000). *First grade non-natural readers: A descriptive study of the teaching and learning of the reading process.* Florida State University, Tallahassee. *Dissertation Abstracts International,* **61**(05). (UMI No. 9971751)

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., et al. (2004). Increasing reading comprehension and engagement through concept-oriented

reading instruction. *Journal of Educational Psychology,* **94**(3), 403–423.

Harcourt Educational Measurement. (2003). *Stanford achievement test 10 (SAT10).* Orlando, FL: Harcourt.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: Paul H. Brookes.

Haskins, R., Walden, T., & Ramey, C. T. (1983). Teacher and student behavior in high- and low-ability groups. *Journal of Educational Psychology,* **75**, 865–876.

Juel, C., & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly,* **35**(4), 458–492.

Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review,* **25**, 215–227.

Karweit, N., & Slavin, R. E. (1981). Measurement and modeling choices in studies of time and learning. *American Educational Research Journal,* **18**, 157–171.

Klingner, J. K., Vaughn, S., Arguelles, M. E., Hughes, M. T., & Leftwich, S. A. (2004). Collaborative strategic reading: "Real-world" lessons from classroom teachers. *Remedial and Special Education,* **25**(5), 291–302.

Kuhn, M. R., & Stahl, S. A. (2002, December). *Developing fluency in classrooms.* Paper presented at the annual meeting of the National Reading Conference, Miami.

Morrison, F. J., Bachman, H. J., & Connor, C. M. (2005). *Improving literacy in America: Guidelines from research.* New Haven, CT: Yale University Press.

Morrison, F. J., & Cooney, R. (2002). Parenting and academic achievement: Multiple paths to early literacy. In J. Borkowski, S. Ramey Landesman, & M. Bristol-Power (Eds.), *Parenting and the child's world: Influences on academic, intellectual, and social-emotional development* (pp. 141–160). Mahwah, NJ: Erlbaum.

NAEP. (2003). *The nation's report card: Reading highlights* (no. NCES 2004-452). Washington, DC: National Center for Education Statistics.

NAEP. (2005). *The nation's report card.* Washington DC: National Center for Education Statistics.

NICHD-ECCRN. (2002). The relation of global first-grade classroom environment to structural classroom features and teacher and student behaviors. *Elementary School Journal,* **102**(5), 367–387.

National Reading Panel. (2000). *National Reading Panel report: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (no. NIH pub. 00-4769). Washington, DC: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Child Health and Human Development.

No Child Left Behind Act of 2001. (2002). *Public Law #107-110, 115 Stat. 1425* [On-line]. Available: http://www.ed.gov

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction,* **1**, 117–175.

Palincsar, A. S., Collins, K. M., Marano, N. L., & Magnusson, S. J. (2000). Investigating the engagement and learning of students with learning disabilities in guided inquiry science teaching. *Speech, Language, and Hearing Services in Schools,* **31**(3), 240–251.

Pressley, M. (2006, April). *What the future of reading research could be.* Paper presented at the annual meeting of the International Reading Association, Chicago.

Pressley, M., & Wharton-McDonald, R. (1997). Skilled comprehension and its development through instruction. *School Psychology Review,* **26**(3), 448–466.

Pressley, M., Wharton-McDonald, R., Allington, R., Block, C. C., Morrow, L., Tracey, D., et al. (2001). A study of effective first-grade literacy instruction. *Scientific Studies of Reading,* **5**(1), 35–58.

Pressley, M., Wharton-McDonald, R., Mistretta-Hapston, J., & Echevarria, M. (1998). Literacy instruction in 10 fourth and fifth grade classrooms in upstate New York. *Scientific Studies of Reading,* **2**(2), 159–194.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2d ed.). Thousand Oaks, CA: Sage.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest,* **2**(2), 31–74.

Reading First. (2007). [On-line]. Available: http://www.ed.gov/programs/readingfirst/index.html

Sadoski, M., & Willson, V. L. (2006). Effects of a theoretically based large-scale reading intervention in a multicultural urban school district. *American Educational Research Journal,* **43**(1), 137–154.

Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and re-

search. *Journal of Personnel Evaluation in Education,* **12**(3), 247–256.

Shaw, R., & Shaw, D. (2002). *DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado State assessment program (CSAP).* Eugene: University of Oregon.

Snow, C. E. (2001). *Reading for understanding.* Santa Monica, CA: RAND Education and the Science and Technology Policy Institute.

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

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology,* **38**(6), 934–947.

Taylor, B. M., Pearson, D. P., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: Lessons about primary-grade reading instruction in low-income schools. *Elementary School Journal,* **101**(2), 121–165.

Taylor, B. M., Pearson, P. D., & Rodriquez, M. C. (2003). Reading growth in high-poverty classrooms: The influences of teacher practices that encourage cognitive engagement in literacy learning. *Elementary School Journal,* **104**, 3–28.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education,* **25**(4), 252–261.

Torgesen, J. K. (2006). Learning to Read in Florida: Issues at the school and classroom level. Presented as part of the seminar series for the National Center for Teacher Transformation, Tarpon Springs, FL. [On-line]. Available: http://www.fcrr.org/science/pdf/torgesen/Tarpon_Springs_April.pdf

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voelier, K. K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities,* **34**(1), 33–58.

U.S. Department of Education. (2006). *Reading First implementation evaluation: Interim report.* Washington DC: Office of Planning, Evaluation, and Policy Development, Policy and Program Studies Service.

University of Oregon. (2007). DIBELS [On-line]. Available: http://dibels.uoregon.edu

Vaughn, S., & Briggs, K. L. (Eds.). (2003). *Reading in the classroom: Systems for the observation of teaching and learning.* Baltimore, MD: Paul H. Brookes.

Walsh, K., Glaser, D., & Wilcox, D. D. (2006). *What education schools aren't teaching about reading and what elementary teachers aren't learning.* Washington, DC: National Council on Teacher Quality.

Wharton-McDonald, R., Pressley, M., & Hampston, J. M. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *Elementary School Journal,* **99**(2), 101–128.

Williams, J. P., Hall, K. M., Lauer, K. D., Stafford, B., DeSisto, L. A., & deCani, J. (2005). Expository text comprehension in the primary-grade classroom. *Journal of Educational Psychology,* **97**(4), 538–550.