Reading Growth in High-Poverty Classrooms: The Influence of Teacher Practices That Encourage Cognitive Engagement in Literacy Learning

Barbara M. Taylor
University of Minnesota

P. David Pearson
University of California, Berkeley

Debra S. Peterson
Michael C. Rodriguez
University of Minnesota

The Elementary School Journal
Volume 104, Number 1
© 2003 by The University of Chicago. All rights reserved.
0013-5984/2004/10401-0001$05.00

Abstract

In this study we investigated the effects of teaching on students' reading achievement. More specifically, based on a framework of reading instruction maximizing students' cognitive engagement in literacy learning, curricular and teaching variables, including aspects of word-recognition and comprehension instruction, approaches to teaching such as telling versus coaching, and the enabling of students' active versus passive responding to literacy activities, were investigated to explain growth on reading comprehension, fluency, and writing measures over a school year in grade 1–5 classrooms. Participants included 88 teachers and 9 randomly selected students per classroom in 9 high-poverty schools across the United States that were engaged in a literacy instruction reform project. Teachers were observed 3 times across the school year during a reading lesson. Hierarchical linear modeling showed that a number of teaching variables explained substantial variation in student growth on several measures of reading achievement. Looking across all of the data, the most consistent finding was that teachers who emphasized higher-order thinking, either through the questions they asked or the tasks they assigned, promoted greater reading growth among the 9 target students in their classrooms. We examine the results of our work in relation to a framework of teacher effectiveness maximizing students' cognitive engagement in literacy learning.

There is an unprecedented emphasis in the United States on improving the teaching of reading in elementary classrooms, and the press for improvement has been increased by the Reading First provisions of Title I, the No Child Left Behind Act of 2001 (2002). Fortunately, there is a great deal of research-based knowledge about effective reading instruction to guide teachers' efforts to become more accomplished reading teachers.

The National Reading Panel Report (NRP, 2000), central to the conceptualization of research and effective reading instruction in the No Child Left Behind legislation, is attracting most of the attention nationally and within states in terms of defining effective reading instruction. The National Reading Panel, charged with reviewing reading research to determine the effectiveness of various approaches to teaching children to read, focused on curricular components of an effective reading program. The NRP found five areas that merited immediate implementation: (*a*) phonemic awareness instruction, (*b*) explicit, systematic phonics instruction, (*c*) repeated oral reading practice with feedback and guidance, (*d*) direct and indirect vocabulary instruction, and (*e*) comprehension strategies instruction. It is important to note that the authors of the NRP report recognized that their report did not speak to many aspects of reading instruction because the research in those areas was neither definitive nor extensive. Moreover, even for those issues for which there was broad consensus (e.g., phonics and comprehension), the panel noted that many caveats were required to add the necessary nuance, qualification, and context that would be required to support valid policy recommendations. Other broad reviews, such as the Report of the National Academy of Education on Preventing Reading Difficulties in Young Children (Snow, Burns, & Griffin, 1998) and particular chapters of the third volume of the *Handbook of Reading Research* (Kamil, Mosenthal, Pearson, & Barr, 2000), have corroborated many of the NRP findings (albeit with substantial elaboration and nuance).

Knowledge about effective reading instruction has also come from the close examination of effective teachers of reading (e.g., Duffy et al., 1987; Knapp et al., 1995; Pressley et al., 2001; Taylor, Pearson, Clark, & Walpole, 2000). In an extensive study of teachers and students in 140 high-poverty classrooms, Knapp et al. (1995) found that effective teachers of reading stressed higher-level thinking skills in addition to lower-level skills. Knapp et al. posited that this extra emphasis on higher-level thinking increased students' understanding of what they were doing and encouraged them to be meaning makers. Even so, relatively few stimulating discussions took place in the classrooms Knapp et al. (1995) studied. Based on extensive examination of classrooms populated by low-income children, Knapp et al. urged teachers to promote students' understanding and to build the quest for meaning into their learning experiences.

Duffy, Roehler, and colleagues (Duffy et al., 1987; Roehler & Duffy, 1984) identified the cognitive processes supported by excellent teachers. More effective teachers engaged in modeling and explanation to teach students strategies they could use to decode words and understand texts. Implicit in their empirically based model of teaching effectiveness was the concept of scaffolding (what others have called coaching)—supportive actions by the teacher to move either an individual or a group of students to the next level of independence in completing a task, strategy, or activity. Pressley et al. (2001) found that effective primary-grade reading teachers provided a balanced literacy program. They taught skills, actively engaged students in a great deal of reading and writing, and fostered self-regulation through a combination of modeling, scaffolding, and providing informative feedback to students as they tried to apply strategies.

Based in part on work suggesting that "engaged" readers have better text comprehension and reading achievement than disengaged readers (Campbell, Voelkl, & Donahue, 1997; Cunningham & Stanovich, 1997), Guthrie and his colleagues (Baker, Dreher, & Guthrie, 2000; Guthrie, 1996; Guthrie et al., 2000) have stressed that teachers need to know how to promote reading engagement in order to teach reading well. Engaged readers are motivated, read to learn, use cognitive strategies, and

interact in a classroom community (Guthrie, 1996). In a study of effective primary-grade reading teachers, Pressley et al. (2002) found that these teachers were skilled at providing motivating instruction, instruction that was as concerned about student involvement as it was about achievement. Chin, Anderson, and Waggoner (2001) looked at fourth-grade classrooms and found that in comparison to recitation, literacy discussions that stressed collaborative reasoning fostered greater engagement and higher-level thinking. These researchers explained this increase by suggesting that students had more control over and participation in the discussions, which in turn led to deeper and more productive cognitive processing. These findings point to instructional practices in which teachers "engage" students' cognitive functioning rather than simply cover key curricular components. In other words, they suggest that the "how" of instruction may be as important as the "what."

Our own work has supported the findings on effective teachers reported above. In an earlier study (Taylor et al., 2000) examining effective schools and teachers, we learned that primary-grade teachers in effective schools differed from their counterparts in less effective schools in a number of ways, including (*a*) providing more high-level questioning, (*b*) coaching students in strategies for applying their word-recognition skills to everyday reading, and (*c*) allowing students more active reading practice (more time for independent reading). When we examined the most effective teachers irrespective of the schools in which they taught, we also found that they, like the teachers in the effective schools, provided more higher-level questioning and emphasized applying word-recognition strategies. Additionally, they provided much more coaching and promoted a higher level of on-task behavior among their students than did the less effective teachers (Taylor et al., 2000).

Toward a Framework for Effective Reading Instruction Maximizing Cognitive Engagement in Literacy Learning

The overall purpose of the CIERA (Center for the Improvement of Early Reading Achievement) School Change study, which is the broader line of work in which this current study of classroom instruction is embedded, was to investigate the efficacy of a school-based reading improvement model. Within the wider scope of that study, we were struck by the consistency, even in the early stages of our work on effective schools and teachers (Taylor et al., 2000), with which particular instructional practices were associated with relatively high student growth. This finding prompted us to wonder whether this set of practices could be conceptualized as a more explicit framework of effective instruction. By moving back and forth between our data and the literature on effective reading instruction, we developed the current framework of teaching for cognitive engagement to partially account for what is known about effective reading instruction. A fundamental tenet of this framework is that what teachers do to maximize students' cognitive engagement in literacy activities will matter as much as what they cover in their instructional interactions with students.

The framework of instruction maximizing cognitive engagement in literacy learning that we are putting forth combines key ideas from the work of Knapp et al. (1995), which stresses teaching for meaning, with the engagement construct of Guthrie et al. (2000). Although the framework we are investigating does not account for all that is known about effective reading instruction (due to the constraints of our observation system, described below), it contains four teaching dimensions: (1) supporting higher-level thinking (in both talk and writing about text), (2) encouraging independent use of word-recognition and comprehension strategies during reading activities (both instruction and text reading), (3) us-

ing a student support stance (in contrast to a teacher-directed stance) during literacy lessons, and (4) promoting active, as opposed to passive, involvement in literacy activities.

In the present study, using an observation system designed to investigate multiple aspects of literacy lessons (Taylor & Pearson, 2000), we evaluated the relative contributions of an array of curricular and teaching variables to children's reading and writing growth. Because our system accounts for a wide range of curricular and teaching variables, not all were relevant to evaluation of the cognitive engagement framework. By the same token, our investigation into elements of effective reading instruction was limited by the aspects of reading instruction included in our observation coding scheme. Fortunately, however, two types of variables were relevant to our investigation—those included in the cognitive engagement framework and those that ran counter to it. Variables in the observation scheme that we were able to consider included explicit phonics skill instruction, instruction in applying word-recognition strategies to text, comprehension skill instruction, comprehension strategies instruction, lower- and higher-level thinking related to text, teachers' use of various interaction strategies (e.g., coaching, modeling, telling, and recitation), student time on-task, and active (as opposed to passive) pupil response to a reading lesson. (The observation system is described in the Method section.)

A key step in instantiating this framework for effective reading instruction was to identify its features in our coding scheme. In most instances, this process proved transparent. For example, the research on higher-level talk about text (Knapp, 1995; Taylor, Peterson, Pearson, & Rodriguez, 2002; Taylor et al., 2000) and use of reading strategies (Guthrie et al., 2000; Pressley et al., 2001) had been a part of our coding scheme from the outset. In the scheme (and in the research literature), when students are engaged in higher-level thinking about text, they are making connections to their prior knowledge, considering thematic elements of the text, interpreting characters' motives and actions, and so on. Similarly, during comprehension and word-recognition strategy instruction, students are engaged in metacognitive thinking and monitoring as they try to solve reading problems.

In contrast, one would expect the practices in the following list to stimulate more mechanistic or routine than active and strategic thinking by students and, hence, lead to relatively less reading growth when they occurred frequently: lower-level questioning, mechanical practice of comprehension skills, and explicit phonics skill instruction past grade 1. When engaged in lower-level thinking about text, students are typically responding to questions about detail or questions that require relatively little thought and can be answered with a word or two. With mechanical practice of comprehension skills, students may be completing a worksheet on a skill such as main idea or fact/opinion, or they may be responding to a teacher prompt to predict or retell in a turn-taking manner. In either case, these comprehension skill activities are likely to require less cognitive effort than when students are trying to apply comprehension strategies to actual reading. In kindergarten and grade 1, when the content is novel and students are attempting to grasp the alphabetic principle, students are likely to be cognitively engaged in phonics lessons. However, past grade 1, phonics lessons typically cover content that either most readers know or some have not yet mastered. In either case, phonics lessons past grade 1 are likely to involve most students in mechanical practice rather than in active cognitive engagement.

A close reading of the research leads one to expect more reading and writing growth in classrooms in which students are more actively engaged in their literacy learning than in other classrooms (Pressley et al.,

2001; Taylor et al., 2002). The more students are performing literacy activities themselves, as opposed to listening to or watching others performing literacy activities, the greater their active involvement in learning and hence the greater their opportunity for growth. Within our observation scheme, we included many indicators of active responding (reading, writing, manipulating, chorally responding, and sharing ideas with a partner) as well as passive responding (reading turn-taking, oral turn-taking, and listening to the teacher).

Teacher stance, which we defined as the mode of interaction between a teacher and his or her students, is also implicated as an important instructional variable in the research on effective teaching (e.g., Duffy et al., 1987; Pressley et al., 2001; Taylor et al., 2000). A teacher's stance toward instruction is related to the nature of the literacy activity (e.g., level of questioning, strategy vs. skill instruction) and to the type of pupil responding being encouraged (e.g., active vs. passive responding). One would expect more reading and writing growth in classrooms in which teachers frequently coach, model, and provide feedback, all of which are important when students face challenging literacy activities (Duffy et al., 1987; Taylor et al., 2000). For example, in these challenging settings a teacher may need to coach students to answer a higher-level question, model for and coach students in the use of specific comprehension or word-recognition strategies, and offer feedback as students attempt to use those strategies during a guided reading activity. In contrast, the practice of telling students information about a comprehension skill or leading them in recitation with low-level questions about a story is likely to lead to less cognitive engagement by students than are more supportive actions.

In analyzing the data for the current study, we were able to fulfill two purposes: (*a*) to determine which elements of classroom instructional practice accounted for the greatest growth in student achievement across a school year to better understand effective reading instruction and (*b*) to evaluate the efficacy of a framework of teaching for cognitive engagement. If such a framework were supported, we reasoned, it would provide teachers with a coherent body of information, beyond that provided by the NRP report, about what they must do to improve teaching, learning, and reading achievement.

Method

Participants

Nine schools were part of the CIERA School Change Project in 2000–2001. All schools were high poverty, with 70%–95% of the students qualifying for subsidized lunch. Across schools, 2%–68% of the students were non-native speakers of English, and 67%–91% were members of minority groups. The nine project schools were from eight districts in a rural area in the Southeast, an eastern city, two small towns in the Midwest, a large city in the Midwest, and a large city in the Southwest. In order to become a project school, at least 75% of the teachers in a building had to agree to participate in the project. In all schools, two teachers per grade were randomly selected and invited to participate in the classroom observations and interviews. This article focuses on students' reading and writing growth in grades 1–5. A total of 88 teachers and 792 students across these grades participated in the data collection.

Teachers from each school varied in teaching experience, ranging from 0 to 35 years of service on average. Almost all teachers held a bachelor's degree in education or a related field, and about 40% at each school had a master's degree.

Because schools had different assessment procedures in place, teachers were asked to use their judgment to divide their classes into thirds (high, average, and low) in terms of reading ability so we could obtain a stratified, random sample. It was from these thirds that we randomly selected nine children in the fall as target students,

three each from the high, middle, and low thirds of the classroom continua of reading ability.

Student Assessments

The children were assessed in the fall and spring on a number of literacy measures that varied depending on grade level. Although the study focused on reading instruction and, for the most part, reading achievement, we included a measure of writing ability to assess any serendipitous effect that the reading instructional practices we observed may have had on the highly related process of composition. Assessments included a standardized reading comprehension test (grades 1–5), a comprehension test from a basal reader program (grades 2–5), fluency (words correct per minute; Deno, 1985) (grades 1–5), and writing (responding to a common prompt) (grades 1–5), as well as tests considering letter-name knowledge (grade 1), phonemic awareness (grade 1), and word dictation (grade 1) (see Table A1).

In the fall in grade 1, children were individually assessed on letter-name knowledge and phonemic segmentation and blending (Taylor, 1991). Children were also assessed in small groups on word dictation. In the spring all students were individually assessed on reading fluency (in which students read aloud for 1 minute to obtain a score for the number of words read correctly; Deno, 1985) based on a grade-level passage from the Basic Reading Inventory (BRI; Johns, 1997). In a group setting, students took the reading comprehension subtest of the Gates-MacGinitie Reading Test (MacGinitie, MacGinitie, Maria, & Dreyer, 2000) and responded to a writing prompt (the same one as used in grades 2–5).

In the fall, students in grades 2–5 were assessed individually on fluency (words correct per minute, wcpm) based on their reading of a BRI passage one grade level below their grade placement. In a group they took the comprehension subtest of the Gates-MacGinitie Reading Test (MacGinitie et al., 2000), a comprehension test from a basal reader program (Houghton Mifflin, 1999) in which they read a two-page passage and answered five multiple-choice and five short-answer questions and responded to a writing prompt (Michigan Literacy Progress Profile, 1998). In the spring, all students were assessed on fluency on a passage at grade level (Johns, 1997), reading comprehension (Gates), basal reader comprehension, and writing to a prompt (using the same prompt as in the fall).

For the writing prompt, each paper was scored by one person according to a four-point rubric. Twenty-five percent of the writing samples at each grade level were scored by a second scorer with 83% agreement between the two scorers.

Use of the School Change Framework

Teachers in the project agreed to meet for a minimum of 1 hour a week, on average, in focused study groups. Study group activities included within-grade and/or across-grade groups that concentrated on particular aspects of classroom reading instruction and student work (e.g., comprehension instruction and phonemic awareness instruction). Groups were encouraged to review research and video clips on the CIERA School Change web site and to read and discuss articles on research-based practices related to their group's focus area. Members of study groups also raised issues, solved problems, and developed action plans related to their focus area in order to make changes in their classroom reading instruction.

Use of data emanating from the project was an important aspect of the process. At the beginning of the 2000–2001 school year, principals received a summary of the Beating the Odds research (Taylor et al., 2000) on characteristics of effective schools and teachers to share with teachers in their schools. The schools also received a report on the 1999–2000 project that presented means across schools on the school and classroom variables under investigation.

Also, the report included correlations identifying school and classroom factors related to growth in students' reading and writing ability. Two purposes of this report were to help schools consider possible strengths and weaknesses in their classroom reading instruction and to help them determine topics to focus on in study groups.

Documenting Program Characteristics and Classroom Practices

Teachers participating in the data collection were interviewed in the fall, winter, and spring. Each interview lasted about 30 minutes. The interview data were used primarily to document school reading program features as well as participants' beliefs about school reform, leadership, collaboration, professional development, and parent partnerships. In this article, however, we used interview data in a very limited manner to shed light on the beliefs about reading instruction in the classrooms of 25 teachers who tended to ask higher- or lower-order questions. No quotes from the interviews are provided, and the interview questions are not given.

On three occasions (fall, winter, spring) each teacher was also observed for an hour during reading instruction to document her classroom practices in the teaching of reading. All observations were scheduled during the portion of a teacher's literacy block devoted primarily to reading instruction. The observers consisted of graduate students in literacy education and retired elementary teachers. They were trained to use the CIERA Classroom Observation Scheme (Taylor & Pearson, 2000) through an on-site visit by one member of the research team and through use of an observation training kit that contained a manual, a practice video, and a CD with video clips of teachers illustrating the various codes in their teaching. The observers then watched an inter-rater "test" video, and they had to demonstrate at least 80% agreement at each of the seven levels of the coding scheme with a "standard" coding (Taylor & Pearson, 2000) prior to conducting classroom observations.

The observation system (influenced by the work of Greenwood, Carta, Kamps, & Delquadri, 1995; Scanlon & Gelzheiser, 1992; and Ysseldyke & Christenson, 1993) combined note-taking with a quantitative coding process. The observer took field notes for a 5-minute period, recording a narrative account of what was happening in the classroom, including, where possible and appropriate, what the teacher and children were saying. At the end of this 5-minute period, the observer first recorded the proportion of all students in the classroom who appeared to be on-task (i.e., doing what they were supposed to be doing). Then the observer coded who was providing the instruction (level 1), the grouping pattern in use for that event (level 2), and the major literacy activity (level 3). Next, the observer coded the two or three most salient literacy events (level 4 codes) occurring during that 5-minute episode. For every level 4 event, the observer also coded the materials being used (level 5), the teacher interaction styles observed (level 6), and the expected responses of the students (level 7). An example of a 5-minute observational segment is provided in Appendix B. (See Table A2 for a list of the codes for the seven levels.)

The level 4 codes dealt primarily with curricular and teaching aspects of reading instruction. Through a curricular lens, we classified the content of instruction. For example, was it phonemic awareness, phonics, vocabulary, or comprehension instruction? We also attempted to capture the nature of the instruction. For example, when the teacher was teaching comprehension skills or strategies, was she treating them as elements to be rehearsed and practiced or as processes to be understood metacognitively and incorporated into students' independent reading? During talk or writing about text, was the teacher asking lower- or higher-level questions after reading? In the case of phonics, we distin-

guished between an explicit phonics lesson and a teacher's attempts to coach students to apply a range of word-recognition strategies as they read text. Level 6 codes dealt with teaching components of reading instruction (e.g., Was telling or coaching observed?) as did level 7 codes (e.g., Was the teacher having students engage in reading or reading turn-taking?).

The observations were used as a source of feedback to individual teachers. Teachers received a copy of each observation, a description of the codes, a brief summary of research related to the major categories of codes being analyzed for the project (e.g., incidence of whole-class instruction, incidence of higher-level questioning), and a table summarizing the data from the observation codes for a national sample of teachers at their grade level. External facilitators received training in how to interpret observations so they, in turn, could help teachers understand the information contained in their observations without interpreting it for them. Teachers were encouraged to go to the facilitators with questions.

Establishing Trustworthiness of the Observation Data

Because the core analyses in this project hinged on the trustworthiness of the data from the observations, we took several steps to ensure the reliability and validity of the codes recorded at each site. First, as mentioned earlier, each observer was required to meet a standard (i.e., achieving at least 80% agreement with a standard set of codes for an observation) prior to conducting the observations used in the data analysis. Second, while visiting each site one research team member conducted a fidelity check with each site observer to obtain data on the interrater reliability of the coding scheme. The two coded the same classroom for 30 minutes. The observer's codes for each 5-minute observation segment were compared with those of the research team member to determine consistency for each of the seven levels of the coding system. Across 12 observations, mean agreement with the research team member ranged from 82% to 95% across the levels of the coding system: 95% agreement at level 2 (grouping), 95% agreement at level 3 (major literacy focus), 82% agreement at level 4 (specific literacy activity), 87% agreement at level 5 (material), 85% agreement at level 6 (teacher response), and 82% agreement at level 7 (student response). The classroom teacher was always coded at level 1 because this was the person we were observing.

Although these indices of congruence meet conventional standards for interrater reliability, we remained concerned about the consistency and the validity of the codes, particularly for levels 4 and 7. Thus we undertook a post hoc review to enhance their validity. An expert observer, who had used the observation scheme for 2 years and had helped to revise and refine it, read the notes taken during each observation by each observer to assess the degree to which observers were using the codes consistently across sites. She did not code the observations "blind." Instead she recorded a different code if her reading of the notes prompted her to disagree with an observer's original code. For a random sample of 10% of the observations, the agreements between the observers and the expert observer at each of the levels of coding for the codes used in the data analysis (described below) were as follows: 97% agreement at level 2 (grouping), 96% at level 3 (major literacy focus), 84% at level 4 (literacy activity), 100% at level 5 (material), 81% at level 6 (teacher response), and 89% agreement at level 7 (student response). In all instances in which a discrepancy was noted, not only for the 10% sample but for the entire corpus, the code assigned by the expert was used in subsequent statistical analyses.

The variability between the observers and the expert, especially at levels 4 and 6, prompted us to undertake yet another reliability check. To ensure that we had not introduced any bias in this first review pro-

cess, a second reviewer read the notes from the same 10% random sample of observations and assigned her own codes. She agreed with the first reviewer over 90% of the time, specifically: 99% agreement at levels 2 and 3, 91% at level 4, 99% at level 5, 93% at level 6, and 92% agreement at level 7. We were thus confident in the reliability and validity of the coding scheme; further, any slippage that may have occurred was detected and corrected during the post-hoc review process.

Observational Categories Used in the Analyses

Certain categories from classroom observations (i.e., those found to be important in previous research) were most relevant to the current study, that is, to determine which elements of classroom instructional practice accounted for the greatest growth in student achievement across a school year and to investigate the efficacy of our framework of teaching for cognitive engagement. The classroom practices (see App. C for descriptions of these categories) that we entered into our statistical analyses were as follows:

1. Whole class/large group
2. Small group
3. Phonemic awareness skill instruction
4. Phonics skill instruction
5. Coaching in word-recognition strategies
6. Active reading practice
7. Vocabulary instruction
8. Comprehension skill instruction
9. Comprehension strategy instruction
10. Lower-level questioning or writing about text
11. Higher-level questioning or writing about text
12. Telling
13. Recitation
14. Modeling
15. Coaching
16. Watching/listening/giving feedback
17. Students actively responding
18. Students passively responding
19. Time on-task

Statistical Analysis

We used hierarchical linear modeling (HLM; Bryk & Raudenbush, 1992) to investigate the effect of classroom-level characteristics on students' reading growth. Descriptive analyses were also conducted to elaborate on the quantitative findings.

HLM is a method of computing regression at multiple levels. The analyses in this study employed a two-level HLM model in which students were nested within classrooms. HLM estimates a regression within each classroom and combines them to see if there is a common regression across classrooms. When regressions (either the intercepts or slopes) vary across classrooms, researchers can examine classroom-level characteristics that may explain such variation. This is a common method for evaluating classroom factors and their effects on student outcomes. It would be inappropriate to use simple regression in these situations because one would be violating the independence assumption of regression.

In addition, HLM partitions variance components across levels, providing an estimate of variance in student performance that exists within classrooms or schools and between classrooms or schools. An unconditional HLM is one without an explanatory variable that allowed us to answer the question, How much variance in student outcome can be attributed to factors on which classrooms differ systematically from one another? This analysis is equivalent to a random-effects analysis of variance. Estimation using HLM rests on assumptions similar to multivariate multiple regression.

The formula for a correlation is $r = \mathrm{cov}_{XY}/s_X S_Y$. The formula for a regression coefficient is $b = \mathrm{cov}_{XY}/s^2_X$. To evaluate the relative importance of regression coefficients based on variables in different scales, one computes a standardized regression coefficient as $\beta = bs_X/s_Y = \mathrm{cov}_{XY}/s^2_X\, s_X/s_Y = \mathrm{cov}_{XY}/s_X s_Y$, which results in something like a partial correlation controlling for other variables in the equation (when there are more than one).

If the resulting effect, standardized beta, is 0.35, for example, we would interpret this as meaning that a change of one standard deviation in the predictor (observation scheme) variable is associated with a 0.35 standard deviation change in the outcome (achievement) variable.

Because of the improved estimation employed by HLM, including the use of maximum likelihood and empirical Bayes estimates, interpretation of statistical results can be broadened to include a larger *p* value associated with statistical tests. Also, statistical results with associated *p* values at or near 0.10 should be included in interpretation and explored in future studies with smaller numbers of cases (fewer teachers, in our case) because such results indicate that there are relations worth exploring further. For a more complete description of estimation in HLM, see Bryk and Raudenbush (1992, pp. 32–56). HLM (Raudenbush, Bryk, & Congdon, 2000) is recognized as a standard program for estimating multilevel models (Bryk & Raudenbush, 1992; Kreft & DeLeeuw, 1998).

Results

Reading Growth and Teacher Practices across Schools

Through the HLM analyses we investigated the relation between teacher practices during literacy instruction and students' reading growth. Because six of nine schools were in their first year of the reform in year 2 of the project, we were unable to look at changes in instruction during the school year with only three observations per teacher. The analyses were conducted on the relationships between teacher practices and each of the major outcome variables: reading fluency, comprehension as measured by a standardized reading test, comprehension as measured by a basal reader test, and writing. Typical results of multilevel models yield 10%–33% of the variance between schools. In studies employing classrooms, some have found as much as 25%–50% of the variance between classrooms (Frank, 1998). Grade 1 data (see Table 1) were analyzed separately from grades 2–5 because different fall scores (e.g., word dictation in grade 1 vs. words correct per minute, comprehension from a standardized reading test in grades 2–5) were used as explanatory variables in the analyses.

Fluency. The HLM analysis (see Table 2) for grade 1 revealed that 35% of the variance in spring fluency scores was between teachers, after accounting for fall scores. Thirty-five percent of the between-teacher variance was accounted for by the variable of higher-level questioning. The average classroom mean spring fluency score was 53.0 (SD = 17.4). For every standard deviation increase in the coding of higher-level questioning (mean percentage of segments observed = 6%, SD = 13%), students' fluency score in a class increased by an average of 11.4 words correct per minute (β = .66).

For grades 2–5, fall scores and grade were used to adjust for differences in growth in fluency across grades. The HLM analysis (see Table 3) revealed that 46% of the variance in spring fluency scores was between teachers after accounting for fall scores. Of this between-teacher variance, coaching, active reading (as opposed to turn-taking reading), and phonics instruction (negatively related) accounted for 13% of the variance. The average classroom mean spring fluency score was 107.9 wcpm (SD = 17.9). For every standard deviation increase in the coding of coaching (mean percentage of segments observed = 14%, SD = 14%), students' fluency score in a class increased by 4.2 wcpm on average (β = .23). For every standard deviation increase in the coding of active reading practice (M = 28%, SD = 12%), students' fluency score in a class increased by 3.0 wcpm on average (β = .17). For every standard deviation increase in the coding of phonics instruction (M = 6%, SD = 10%), students' fluency score in a class decreased by 5.2 wcpm on average (β = .29).

Reading comprehension as measured by a standardized test. The HLM analysis (see

TABLE 1. Means and Standard Deviations for Student Scores, Grades 1–5

		Fall		Spring	
Assessment Tool/Grade	*N*	*M*	SD	*M*	SD
Letter names (grade 1)	124	49.44	5.03		
Phonemic awareness (grade 1)	124	6.96	4.30		
Word-level dictation (grade 1)	125	14.70	9.39		
Gates comprehension (NCE):					
Grade 1	107			48.99	18.38
Grade 2	147	43.23	19.05	45.61	18.96
Grade 3	166	36.27	14.34	37.65	16.40
Grade 4	152	35.90	17.68	35.69	17.40
Grade 5	135	35.05	18.46	38.98	17.93
Basal comprehension:					
Grade 2	155	12.22	4.54	15.03	2.93
Grade 3	151	12.62	4.87	14.27	5.39
Grade 4	146	11.71	4.86	13.66	4.82
Grade 5	134	13.54	4.92	15.09	4.46
Fluency:					
Grade 1	135			52.41	31.63
Grade 2	146	54.52	29.85	80.78	31.07
Grade 3	128	79.16	31.94	86.87	30.42
Grade 4	154	105.12	37.12	122.03	38.76
Grade 5	135	121.33	40.31	139.96	42.06
Writing:					
Grade 1	123			2.04	.79
Grade 2	147	1.67	.60	1.99	.75
Grade 3	141	1.89	1.81	1.97	.79
Grade 4	123	1.71	.66	2.01	.76
Grade 5	135	2.02	.72	2.22	.79

TABLE 2. HLM Analysis of the Relation of Grade 1 Reading Fluency to Higher-Level Questioning

	Variance Component	% Variance Between		
Initial random effects:				
Classroom means	301.41	35		
Student residual	544.51			
Total	845.92			
		% Variance Accounted for by Model		
Final random effects:				
Classroom means	197.17	35		
Student residual	543.89			
	Coefficient	*t* Ratio	*df*	*p*
Final fixed effects:				
Intercept (grand mean)	52.98	14.26	15	
High-level questioning	88.14	2.70	15	.017
Fall score	2.14	6.31	132	

Table 4) for grade 1 revealed that 35% of the variance in spring comprehension scores was between teachers, after accounting for fall scores. Twenty-seven percent of the between-teacher variance was accounted for by the variable of higher-level questioning. The average classroom mean spring comprehension NCE score was 49.7 (SD = 9.8). For every standard deviation increase in the coding of higher-level questioning (*M*

TABLE 3. HLM Analysis of the Relation of Grade 2–5 Reading Fluency (after Accounting for Fall Scores and Grade) with Phonics, Active Reading, and Coaching

	Variance Component	% Variance Between
Initial random effects:		
Classroom means	319.84	46
Fall score slope	.03	
Student residual	368.69	
Total	668.56	
		% Variance Accounted for by Model
Final random effects:		
Classroom means	278.89	13
Fall score slope	.03	
Student residual	368.86	

	Coefficient	*t* Ratio	*df*	*p*
Final fixed effects:				
Intercept (grand mean)	107.85	47.44	58	.000
Grade	21.50	9.36	58	.000
Phonics	−51.98	−2.29	58	.025
Active reading	25.31	1.99	58	.051
Coaching	29.84	1.86	58	.068
Fall score	.83	22.88	62	.000

TABLE 4. HLM Analysis of Grade 1 Reading Comprehension (Gates-MacGinitie, NCE) with Higher-Level Questioning

	Variance Component	% Variance Between
Initial random effects:		
Classroom means	93.69	35
Student residual	174.51	
Total	268.20	
		%Variance Accounted for by Model
Final random effects:		
Classroom means	68.30	27
Student residual	174.37	

	Coefficient	*t* Ratio	*df*	*p*
Final fixed effects:				
Intercept (grand mean)	49.65	20.72	14	.000
Higher-level questioning	44.06	2.30	14	.037
Fall score	1.30	7.23	122	.000

= 6%, SD = 13%), students' comprehension NCE score in a class increased by 5.7 on average (β = .59).

For grades 2–5, the HLM analysis (see Table 5) revealed that 48% of the variance in spring comprehension NCE scores was between teachers after accounting for fall scores. Of this between-teacher variance, higher-level questioning (positively related), classroom time on-task (positively related), comprehension skills (negatively related), and passive responding (negatively related) accounted for 20% of the variance. In grades 2–5 the average classroom

TABLE 5. HLM Analysis of Grade 2–5 Reading Comprehension (Gates-MacGinitie) with Higher-Level Questioning, Time On-Task, Comprehension Skills (Negatively Related), and Passive Responding (Negatively Related)

	Variance Component		% Variance Between	
Initial random effects:				
Classroom means	63.32		31	
Fall score slope	.03			
Student residual	142.68			
Total	206.03			
			% Variance Accounted for by Model	
Final random effects:				
Classroom means	53.70		16	
Fall score slope	.03			
Student residual	142.97			
	Coefficient	*t* Ratio	*df*	*p*
Final fixed effects:				
Intercept (grand mean)	40.27	38.70	61	.000
Higher-level questioning	13.99	2.57	61	.010
Time on-task	3.60	1.96	61	.049
Comprehension skills	−17.01	−2.37	61	.018
Passive responding	−14.25	−1.78	61	.075
Fall score	.67	14.42	65	.000

mean spring comprehension NCE score was 40.3 (SD = 8.0). For every standard deviation increase in the coding of higher-level questioning (M = 19%, SD = 19%), students' NCE score in a class increased by an average of 2.5 (β = .33). For every standard deviation increase in the coding of time on-task (M = 91%, SD = 5.5%), students' NCE score in a class increased by 2.0 on average (β = .27). For every standard deviation increase in the coding of comprehension skill instruction (M = 13%, SD = 14%), students' comprehension NCE score in a class decreased by 2.4 on average (β = .30). For every standard deviation increase in the coding of passive responding (M = 72%, SD = 13%), students' NCE score in a class decreased by 1.8 on average (β = .23).

Reading comprehension as measured by a basal reader test. For grades 2–5, the HLM analysis (see Table 6) revealed that 40% of the variance in spring comprehension scores was between teachers, after accounting for fall scores. Asking higher-level questions after reading accounted for 4% of the variance between teachers. The average classroom mean spring score was 14.5 (SD = 2.5). For every standard deviation increase in the coding of higher-level questioning, students' comprehension score in a class increased on average by .65 points (β = .26).

Writing. For grade 1, 39% of the variance in spring writing scores was between teachers, after accounting for fall dictation scores (see Table 7). Teaching comprehension strategies and telling (negatively related) accounted for 73% of the variance between teachers. The average classroom mean writing score was 2.0 (SD = 0.5). For every standard deviation increase in the teaching of comprehension strategies (M = 6%, SD = 13%), students' writing score in a class increased by 0.4 points on average (β = .80). For every standard deviation increase in the coding of telling (M = 51%, SD = 19%), students' writing score in a class decreased on average by 0.13 points (β = .39).

For grades 2–5, 43% of the variance in spring writing scores was between teachers,

TABLE 6. HLM Analysis of Grade 2–5 Reading Comprehension (Basal Reader) with Higher-Level Questioning

	Variance Component	% Variance Between		
Initial random effects:				
Classroom means	6.31	46		
Fall score slope	.07			
Student residual	9.23			
Total	15.61			
		% Variance Accounted for by Model		
Final random effects:				
Classroom means	6.06	4		
Fall score slope	.07			
Student residual	9.24			
	Coefficient	*t* Ratio	*df*	*p*
Final fixed effects:				
Intercept (grand mean)	14.48	43.53	63	.000
Higher-level questioning	3.49	2.00	63	.045
Fall score	.46	9.97	64	.000

TABLE 7. HLM Analysis of Grade 1 Writing with Comprehension Strategies and Telling (Negatively Related)

	Variance Component	% Variance Between		
Initial random effects:				
Classroom means	.249	39		
Student residual	.387			
Total	.636			
		% Variance Accounted for by Model		
Final random effects:				
Classroom means	.067	73		
Student residual	.388			
	Coefficient	*t* Ratio	*df*	*p*
Final fixed effects:				
Intercept (grand mean)	2.03	23.46	13	.000
Comprehension strategies	3.07	3.86	13	.002
Telling	−1.02	−2.08	13	.058
Fall score	.023	2.75	119	.006

after accounting for fall scores (see Table 8). The average classroom mean writing score was 2.0 (SD = 0.5). Asking higher-level questions after reading, and modeling accounted for 22% of the variance between teachers. For every standard deviation increase in the asking of higher-level questions (M = 19%, SD = 19%), students' writing score in a class increased by 0.2 (β = .38). For every standard deviation increase in the coding of modeling, students' writing score in a class increased on average by 0.13 points (β = .24).

Summary across measures. In three of four HLM analyses for grades 2–5, higher-level questioning contributed to students' growth in reading and writing. In grade 1, higher-level questioning contributed to students' improvement in reading comprehension and fluency. Relatively frequent phonics instruction, based on classroom observations, was negatively related to stu-

TABLE 8. HLM Analysis of Grade 2–5 Writing with Higher-Level Questioning and Modeling

	Variance Component	% Variance Between		
Initial random effects:				
Classroom means	.255	43		
Student residual	.342			
Total	.597			
		% Variance Accounted for by Model		
Final random effects:				
Classroom means	.199	22		
Student residual	.342			
	Coefficient	*t* Ratio	*df*	*p*
Final fixed effects:				
Intercept (grand mean)	2.02	32.26	59	.000
Higher-level questioning	1.00	2.95	59	.005
Modeling	1.80	1.93	59	.058
Fall score	.36	4.82	49	.000

dents' fluency growth in grades 2–5. Comprehension skill instruction was negatively related to students' growth on the standardized comprehension measure in grades 2–5, whereas comprehension strategy instruction was associated with students' writing growth in grade 1.

In terms of teacher stance and student engagement or mode of responding, time on-task was positively related to students' gains in standardized reading comprehension in grades 2–5, whereas passive responding was negatively related. Coaching and involving students in active reading (as opposed to the more passive practice of turn-taking during reading) were positively related to gains in fluency in grades 2–5. Modeling was related to improvement in writing in grades 2–5. Telling was negatively related to students' writing growth in grade 1.

Descriptive Data from the Classroom Observations

It is important to qualify the significant findings in these statistical analyses with a descriptive snapshot of the frequency of occurrence of these various instructional behaviors. These data (see Table 9) are important because they show that even those instructional practices that explained substantial growth in student achievement occurred, in real terms, so infrequently (or in the case of negative predictors, so frequently) as to suggest that much remains to be done in efforts to change instruction and improve professional development.

Grouping practices. Across all grades, whole- and small-class/group instruction were coded with equal frequency except that there was more whole-group than small-group instruction coded in grade 5. In contrast, a greater occurrence of small- than whole-group instruction characterized the most effective schools in our earlier study of primary-grade reading instruction in schools that were beating the odds (Taylor et al., 2000). That being said, a framework of cognitive engagement would suggest that what is important is that students be actively involved in their learning, and this could occur in well-run small- or large-group instruction. Poor instruction, irrespective of the size or nature of the group experiencing it, may lead to a high number of students being off-task or may involve passive responding.

Reading instruction. Not surprisingly, word-level activities during reading were observed more in grade 1 than in grades

TABLE 9. Incidence (Percentages) of Classroom Factors, by Grade

	Grade 1 (N = 17)		Grade 2 (N = 18)		Grade 3 (N = 17)		Grade 4 (N = 18)		Grade 5 (N = 17)	
Factor	M	SD	M	SD	M	SD	M	SD	M	SD
Percentage of time (5-minute segments) coded:										
Whole group	.47	.29	.48	.37	.46	.28	.46	.38	.63	.30
Small group	.49	.32	.45	.34	.49	.20	.57	.35	.37	.30
Narrative text	.53	.24	.72	.22	.55	.22	.61	.28	.44	.30
Informational text	.06	.10	.06	.10	.18	.17	.21	.21	.21	.20
Telling	.51	.19	.54	.18	.52	.16	.61	.18	.55	.20
Recitation	.59	.23	.62	.19	.73	.20	.76	.16	.66	.17
Modeling	.05	.05	.04	.06	.03	.04	.05	.09	.05	.08
Coaching	.20	.16	.16	.11	.14	.11	.15	.19	.11	.15
Percentage of reading segments coded:										
Phonemic awareness	.09	.07								
Phonics instruction	.22	.17	.10	.13	.06	.09	.02	.03	.04	.11
Word-recognition strategies	.16	.14	.18	.19	.08	.07	.08	.10	.06	.07
Vocabulary	.22	.21	.23	.16	.32	.27	.29	.15	.23	.15
Comprehension skills	.12	.17	.08	.11	.18	.17	.15	.13	.13	.16
Comprehension strategies	.06	.13	.02	.05	.09	.10	.07	.14	.16	.15
Meaning of text:										
Lower level	.28	.20	.41	.24	.52	.21	.50	.22	.46	.26
Higher level	.06	.13	.11	.11	.20	.20	.21	.21	.26	.19
Active reading practice	.37	.26	.47	.20	.31	.21	.31	.15	.26	.19
Percentage of all codes for student responding:										
Active responding	.36	.14	.38	.09	.25	.13	.20	.07	.26	.12
Passive responding	.64	.14	.62	.09	.75	.13	.80	.07	.74	.12
Time on-task	.92	.05	.92	.03	.91	.38	.89	.05	.90	.08

2–3 or 4–5. Phonics instruction was coded for 22% of the reading segments in grade 1 but coded much less often in grades 2–5. Phonemic awareness instruction was seldom coded beyond grade 1. Coaching in word-recognition strategies during reading was coded with some regularity in grades 1 and 2 but with less frequency in grades 3–5. These findings are similar to those in our study of primary-grade reading instruction in schools beating the odds (Taylor et al., 2000), in which we found that word-level activities were infrequently observed in grade 3. The findings related to word skill activities also suggested that teachers are focused on phonics instruction in first grade. This finding is compatible with the National Reading Panel Report (2000) recommendation that "Phonics instruction taught early proved much more effective than phonics instruction introduced after first grade" (p. 2-85).

Little comprehension instruction was observed. Comprehension skill instruction was coded for 8%–18% of the reading segments across grades 1–5; comprehension strategy instruction was coded for 2%–9% of the segments in grades 1–4 and 16% of the segments in grade 5.

Across all grades, little higher-level questioning or writing related to texts was observed. This was coded for 6% of the reading segments in grade 1, 11% of the segments in grade 2, and 20%–26% of the segments in grades 3–5. In contrast, lower-level questioning was coded for 28% of the reading segments in grade 1, 41% of the segments in grade 2, and 46%–52% of the segments in grades 3–5. Similarly, a low incidence of higher-level questioning was found in our earlier study (Taylor & Pearson, 2000). However, in this earlier study, we did find that more accomplished teachers (based on experts' ratings) more fre-

quently asked higher-level questions than did less accomplished teachers, just as in the current study more effective teachers (based on students' reading growth) asked higher-level questions more frequently than less effective teachers.

Materials. Across all grades, informational text was seldom a part of the lessons we observed; it was coded for only 6% of the segments in grades 1 and 2 and 18%–21% of the segments in grades 3–5. In contrast, narrative text was coded for one-half to three-quarters of the segments in grades 1–4 and 44% of the segments in grade 5.

Teacher and student actions. Telling and recitation were common interaction styles of teachers in all grades, with telling coded for 51%–61% of the segments in grades 1–5. Recitation was coded for 59%–76% of the segments in grades 1–5. In contrast, modeling was coded for 3%–5% of the segments in grades 1–5. Coaching was only observed for 20% of the segments in grade 1, and from 11%–16% of the segments in grades 2–5. In our earlier study (Taylor et al., 2000), we also found that telling was a common interaction style, with the least accomplished teachers having a preferred style of telling children information, whereas the most accomplished teachers had a preferred interaction style of coaching.

Across all grades, students in the present study were coded as more often engaged in passive responding than in active responding. Passive responding, which included reading turn-taking (e.g., round robin reading), oral turn-taking, or listening to the teacher, was coded for 62%–64% of the student responses in grades 1 and 2 and 74%–80% of the student responses in grades 3–5. In contrast, active responding (reading, writing, and manipulating) was coded for 36%–38% of the student responses in grades 1–2 and 20%–26% of the student responses in grades 3–5.

Perhaps what is most important to remember about these descriptive data, in comparison to the HLM analyses, is that even modest levels of occurrence of these key variables, such as coaching and modeling or higher-level questioning, were associated with substantial growth in student achievement. One can only wonder, if a little goes such a long way, what would happen with wholesale changes in these practices.

Examples of Effective Reading Practices

One limitation of quantitative analyses is that it is difficult to get a picture of what the results look like in everyday practice. To offer a clearer sense of what was going on inside these classrooms, we describe teachers who aptly illustrate the practices identified as positive by the quantitative analyses. We also provide some counterexamples—classrooms in which low-level questioning was apparent and a heavily teacher-directed stance was prominent. Illustrations from the field notes are provided along with some of the conversation from teachers and students. To shed further light on one of the most consistent findings in the study, we conclude this section with a description of the teachers who frequently asked higher-level questions and contrast them with teachers who relied primarily on lower-level questions.

Grade 1. The HLM analyses showed that students improved more in comprehension and fluency when their teachers were coded as asking more higher-level questions than other teachers. Students grew more in writing when their teachers taught comprehension strategies and did not often tell students information. We searched the grade 1 teacher file and located one teacher who was higher than the mean in terms of asking higher-level questions and providing comprehension strategies instruction but lower in terms of telling students information. We also located one teacher who did a fair amount of lower-level questioning but little higher-level questioning, provided little comprehension strategies instruction, and exceeded the mean in terms of telling students information.

Ms. Hernandez (all names are pseudonyms), a first-grade teacher, engaged her students in a great deal of higher-level thinking during reading. She used small-group instruction extensively. While she worked with one group, the other students worked in centers, which included writing words/word families, math, computer, library, and reading signs and charts in the room.

On one day, as the students were reading a story, the teacher introduced a "GO" chart with columns that were labeled as follows: prediction, vocabulary, understanding, interpretation, connections, retelling. She prompted students to complete information on the chart by pointing to the "Prediction—I think the story is about . . ." column. Students gave their predictions based on the title and pictures. She asked them to check in the book to look for challenging vocabulary they thought they should add. Ms. Hernandez suggested a word that had the same meaning as house. They added cottage to the chart. The teacher asked, "Other interesting words? Another word for woods?" Students suggested forest. Ms. Hernandez referred to the "Understanding—I noticed" column of the chart. A student suggested, "The giant does interesting things." The teacher referred to the "Interpretation—I wonder" column. She encouraged students to think about what would happen next, to go beyond the story, to imagine what the characters could do together. "Think if you are one of the characters in the story how would you solve the problem. What connections can you make? What is the main thing you learned in the story?" A student explained it was about friends. Ms. Hernandez asked, "What maps can you use to help you retell the story?" Students suggested various graphic organizers such as story webs, circle maps, tree maps. The teacher asked how and why they could use each one. A student suggested they could use a bubble map to describe the character.

Another day Ms. Hernandez was working with a small group on informational texts. She referred to the "GO" chart. She told the students they would quickly review the steps without looking at the chart. Students responded, "I think this story is about (prediction)," "I noticed important words in the story (vocabulary)," "I noticed (understanding)," "I wonder (interpreting)," "This reminds me (connections)," and described maps/story/illustration (retelling).

Ms. Metcalf, another first-grade teacher, primarily taught through small-group instruction, as did Ms. Hernandez. Unlike Ms. Hernandez, however, she typically did much of the work for the students, using a highly teacher-directed stance toward instruction. She minimized opportunities for students to engage in higher-level thinking by doing considerable talking about text herself. One day Ms. Metcalf was reading a story to the children. As they discussed the story, she interjected her own ideas and summarized for the group. She asked, "Why do you think they wrapped the [dinosaur] bones?" This could have served as a higher-level question, but the teacher simply acknowledged a brief answer from one student ("So they wouldn't break") and then provided a lengthy answer herself. She missed the opportunity to have the children express their ideas and elaborate on their thinking.

Grades 2–5. In grades 2–5 the HLM analyses revealed that students had greater growth in reading comprehension when teachers asked higher-level questions, maintained high levels of on-task behavior, and infrequently used comprehension skill instruction (comprehension taught as a routine as compared to a strategic approach to development of comprehension processes) or passive responding (e.g., turn-taking). Students showed more improvement in fluency when their teachers were more often observed coaching, engaging their students in active reading practice, and less often observed teaching phonics lessons. Greater writing growth was positively related to the

incidence of higher-level questions and modeling. We searched the grade 2–5 teacher file and located two teachers: one who personified these findings (i.e., illustrated relatively high levels of the factors positively related to gains and low levels of the negatively related factors) and one contrasting teacher (i.e., who illustrated relatively low levels of factors positively related to gains and high levels of negatively related factors).

Mr. May, a fifth-grade teacher, provided many good examples of the "best practices" identified in the HLM analyses for grades 2–5. He stressed higher-level questions and active pupil involvement. As students were about to read the next chapter in the *Best Christmas Pageant Ever* (Robinson, 1972), Mr. May challenged them to think about what was happening in the story. "Do you think the children should be in the play?" He took a vote, which fostered active pupil involvement, and he had students defend their opinions. In their response journals, students were to answer the following higher-level questions: Do you think the Herdmans will do a good job? Why or why not? How do you think the audience will react? Give evidence from the story.

Mr. May provided small-group instruction to struggling readers while the rest of the class was reading independently. He coached the small group in the use of word-recognition strategies as they read the chapter, and he helped prepare them for the questions he would be asking in the whole-class setting.

Mr. May reconvened the whole class and they discussed their journal entries, after which students participated in small work groups. He asked the groups to write the meanings for two vocabulary words and answer a question they had been given. They recorded their answers on an overhead transparency in anticipation of showing their work to the rest of the class.

In contrast, Mr. Burns, a fourth-grade teacher, asked mostly lower-level questions, and students were engaged in considerable passive responding. During one lesson, students were divided into three groups. One worked with an aide, a second worked at their seats on worksheets, and a third met with Mr. Burns. The introduction to the story lasted 30 minutes. During that introduction, the teacher asked students about the meaning of words they would come across when they read: building, marketplace, celebration, foolishness, snickered, vow. When students could not answer, he told them what the word meant. Once the introduction had been completed, Mr. Burns began a round robin reading session. When a student got stuck on a word, the teacher told her/him the word. Questioning was a rapid fire of low-level questions: "Why did Mom miss Merritt? Did she use her cane when she had her guide dog with her? What did her daughter want her to do when she went on errands? What is the name of the school she is going to?" During this lesson, the teacher did almost all of the talking. Not surprisingly, pupil engagement was low.

Characteristics of teachers who asked higher-level questions. The most consistent finding in this study was that higher-level questioning was related to student literacy growth. We decided to see if we could "unpack" this finding in our attempt to better understand effective reading instruction. To that end, we engaged in a comparative analysis of the observations and interviews of teachers who asked more higher-level questions and those who asked relatively few.

First, we selected the observations and interviews of teachers who were relatively high, which we operationalized as one standard deviation above the mean, on the higher-level question index. Three teachers in grade 1, four in grade 3, two in grade 4, and four in grade 5 met this criterion, for a total of 13. We also selected a comparison group of teachers who were at or above the mean in terms of asking lower-level questions but below the mean in terms of asking higher-level questions. One grade 1 teacher, four grade 2, two grade 3, four grade 4, and

one grade 5 teacher met this dual criterion, for a total of 12.

We reread all of the classroom observations and coded activities related to talk or writing about text to add greater depth to the analysis than had been possible in the initial coding. Several promising subcategories of teacher-student interactions emerged from this analysis, that is, they occurred with enough frequency in the corpus of 25 teachers to provide some insights about the correlates of higher- versus lower-order questions. The list included questions that emphasized (*a*) theme, (*b*) character interpretation, (*c*) relating the text to one's life, (*d*) story events, (*e*) retelling or summarizing the text, and (*f*) making predictions before or during reading. We also found two activities that fell outside the questioning, (*g*) engaging students in a picture walk prior to reading a story or picture book and (*h*) asking students to work in pairs or small groups to discuss/answer questions about a text.

Table 10 reports the descriptive data for these analyses. The basic data reported were the percentage of teachers in the two groups—those who asked more higher-level questions (HLQ) and those who asked more lower-level questions (LLQ)—who engaged in the eight practices identified in the second, deeper reading of the observational and interview data. Chi-square tests were conducted to determine whether the relative percentages of teachers from the two groups differed for each of the eight practices. These analyses revealed that more HLQ teachers than LLQ teachers asked about theme and asked students to discuss text in a small group or with a partner. More LLQ than HLQ teachers asked about events from a story and engaged students in a picture walk.

We also read through the three teacher interviews, coding responses to questions that were related to teachers' reading comprehension instruction. When asked to relate the three critical components of their classroom reading program, 92% of the HLQ teachers mentioned reading comprehension, whereas 33% of the LLQ teachers mentioned comprehension. In contrast, 50% of the LLQ teachers but only 15% of the HLQ teachers mentioned small-group guided reading as a critical component of their classroom reading program. When asked how they knew if their students were learning in reading, 62% of the HLQ teachers but only 20% of the LLQ teachers reported that they looked at students' responses to the questions they asked.

Discussion

One purpose of this study was to investigate the extent to which reading instruction maximizing students' cognitive engagement enhanced elementary students' growth in reading and writing. In general, results suggest that such an approach to reading instruction is effective. Below we discuss the relation of specific findings to this framework.

A second, related purpose of this study was to determine which aspects of reading instruction had the largest effect on students' reading growth. One consistent finding is that higher-level questioning matters. The more a teacher asked higher-level questions, the more growth the nine target students in her class experienced on a variety of measures. The teachers who asked more higher-level questions appeared to understand the importance of challenging their students to think about what they had read. In the process of asking more higher-level questions, at least-two thirds of the HLQ teachers emphasized character interpretation and connections to experience, and they focused more on thematic elements and student leadership in discussions than did LLQ teachers. In doing so, they implicitly implemented elements of the framework of cognitive engagement, especially in encouraging students to focus on higher-level thinking about what they had read.

Another set of findings suggests that routine practice on skills is not beneficial. The more that explicit phonics skill instruc-

TABLE 10. Percentage of High-Level Questioning and Low-Level Questioning Teachers Engaging in Story-Questioning Practices

Practice	High-Level Questioning Teachers (N = 13)	Low-Level Questioning Teachers (N = 12)
Theme[a]	46	8
Character	92	67
Relate to life	69	33
Events[a]	46	92
Retell	77	75
Predict	54	75
Picture walk[a]	8	42
Student discussion[a]	46	0

[a]Significant differences based on chi-square test.

tion was observed in grades 2–5, the lower the growth in reading achievement. Given that two-thirds of the students tested at these grades (e.g., average and above-average readers) were decoding well and thus probably had little to gain from phonics practice activities, the finding is not surprising. Furthermore, this finding is compatible with the recommendation in the NRP Report that phonics instruction should be concentrated in the earliest stages of schooling, mainly grades K–1, when it would provide novel information that might be cognitively engaging to students attempting to grasp the alphabetic principle.

The more that routine, practice-oriented approaches to teaching important comprehension processes were observed, the lower the growth in reading comprehension. A framework of cognitive engagement in literacy learning would predict that this more mechanical approach to comprehension processes (e.g., rehearsal related to comprehension skills) would not foster active thinking on the part of students and would not enhance their reading growth. In contrast, a strategic approach to the development of comprehension processes was found to be related to writing growth in grade 1. Interestingly, strategic comprehension instruction was explicitly acknowledged in the report of the National Reading Panel (2000). This finding suggests that how one teaches comprehension, mechanistically or strategically, is a key factor in determining the efficacy of comprehension instruction.

Other teaching variables such as passive responding (negatively related) and high engagement (positively related) were found to be associated with students' growth in reading comprehension. High levels of coaching and involving students in active reading enhanced students' growth in fluency. High levels of telling (negatively related) and modeling (positively related) predicted students' writing growth. Taken together, these findings suggest that two additional components of our framework of cognitive engagement—student support and active involvement—are important strategies to consider to improve reading instruction.

In the years following the current project, as we provide teachers and schools with data on teaching practices related to students' performance and as teachers continue to learn about effective reading instruction in study groups, we plan to investigate the degree to which classroom teaching practices shift over time toward reading instruction that maximizes students' cognitive engagement in literacy lessons. Additionally, we will evaluate the effects of those shifts on student achievement.

Limitations

This study involved schools that were engaged in a reform project that emphasized implementing researched-based read-

ing practices of effective schools and teachers, thus limiting the generalizability of the findings. Teachers received data on their school's reading practices and their own reading practices that they compared to research on effective teaching of reading, and this may have affected the results. Another limitation is that we were only able to investigate classroom practices in nine schools. Finally, classroom information was gathered from three 1-hour observations per classroom per year. At best, we have only a snapshot of the reading instruction in these classrooms. These limitations are important to appropriately qualify our findings; what is perhaps more remarkable is that with a limited sample of teacher behavior, we were able to explain a great deal of the between-classroom variation in student growth.

Conclusions

The improvement of U.S. students' reading achievement is a national goal (Bush, 2001). Schools know that a wealth of information exists to help them move toward this goal, but knowing where to focus is often difficult for teachers, especially when they are bombarded with so many options. Our findings suggest that elements of a framework of reading instruction that maximizes students' cognitive engagement are important to consider when attempting to improve reading instruction. In addition to the reading curriculum, or what teachers teach, how teachers teach reading is of paramount importance.

The description of effective reading instruction emerging from our work encompasses teachers who challenge students with higher-level thinking and the application of reading strategies to their reading and writing. Effective teachers' questioning for texts is purposeful, and they assess students' learning through their answers to challenging questions. They actively involve students in literacy activities, often giving them responsibility for holding their own discussions about text, and they maintain high pupil involvement. Effective teachers use coaching and modeling to help students learn as well as to help them assume responsibility for their own learning. A challenge that remains is to help teachers translate research on effective reading instruction into practice through ongoing, quality professional development within their schools. We hope that the schools in our continuing project on school reform in reading will improve both the "what" (the curricular elements) and "how" (the teaching processes) of their classroom reading instruction with the end result of enhanced reading growth for all students.

Appendix A

TABLE A1. Assessments, Year 2

Ability Assessed	Assessment Tool/Description	Grades Tested: Fall	Grades Tested: Spring
Letter names	Letter Name subtest from Pikulski Emergent Literacy Survey, Houghton Mifflin. Students identified lowercase and uppercase letters.	1	
Phonemic awareness	Rhyming subtest from Pikulski Emergent Literacy Survey. Students were given a word and asked to say a word that rhymed with this word. Nonsense words were acceptable. Total of 8 points.	1	
	Classroom Segmentation and Blending Test (Taylor, 1991). Children were given six words to blend, "What is /c/ /a/ /t/?" Then they were given six words and asked to identify the first, middle, and last sound they heard in each word. "What sound do you hear first in 'sad'? What sound do you hear in the middle of 'sad'? What sound do you hear at the end of 'sad'?"		
Writing from word-level dictation	Graded lists from Right Start Project (Longmont, CO). Students were asked to write 15 dictated words. If they could write at least seven words correctly from the first list, they went on to a second list of 15 words. (Administered in a group.)	1	
Writing to a prompt	Michigan Writing Assessment. Students were asked to write to a prompt (e.g., Tell about a favorite place). The same prompt was used in the fall and spring. A scoring rubric was used to score papers from high (4) to low (1). (Administered in a group.)	2–5	1–5
Fluency	Words correct per minute on passage from Johns's Basic Reading Inventory (BRI; 1997). In fall, students read for 1 minute from a passage that was one level below grade level. In spring they read from a passage that was at grade level.	2–5	1–5
Comprehension:			
Basal	Retelling of Johns's BRI passage read in fall.	2–5	1–5
	Houghton Mifflin Baseline Test. After reading a three-page story, students answered five short-answer questions and five multiple-choice questions. Possible score is 20 (0, 1, or 2 points for each short-answer question and 2 points for each multiple-choice question correct). A score of 0–10 is considered low, 11–15 average, and 16–20 high. Narrative only, administered in a group.)		
Standardized test	Gates-MacGinitie Reading Test. Only the passage comprehension subtest was administered. Students read short passages and answered multiple-choice questions. (Administered in a group.)	2–5	1–5

TABLE A2. Codes for Classroom Observations

Level	Code	Level	Code
Level 1—Who:		Individual	i
Classroom teacher	c	Other	o
Reading specialist	r	Not applicable	9
Special education	se	Level 3—General focus:	
Other specialist	sp	Reading	r
Student teacher	st	Composition/writing	w
Aide	a	Spelling	s
Volunteer	v	Handwriting	h
No one	n	Language	l
Other	o	Other/not applicable	o/9
Not applicable	9	Level 4—Specific focus:	
Level 2—Grouping:		Reading connected text	r
Whole class/large group	w	Listening to text	l
Small group	s	Vocabulary	v
Pairs	p	Meaning of text, lower:	

TABLE A2. Codes for Classroom Observations *(continued)*

Level	Code	Level	Code
m1 for talk	m1	Student writing	w
m2 for writing	m2	Board/chart	b
Meaning of text, higher:		Worksheet	s
m3 for talk	m3	Oral presentation	op
m4 for writing	m4	Pictures	p
Comprehension skill	c	Video/film	v
Comprehension strategy	cs	Computer	c
Writing	w	Other/not applicable	o/9
Exchanging ideas/oral production	e/o	Level 6—Teacher interaction:	
Word ID	wi	Tell/give info.	t
Sight words	sw	Modeling	m
Phonics:		Recitation	r
p1 = letter sound	p1	Discussion	d
p2 = letter by letter	p2	Coaching/scaffolding	c
p3 = onset/rime	p3	Listening/watching	l
p4 = multisyllabic	p4	Reading aloud	ra
Word-recognition strategies	wr	Check work	cw
Phonemic awareness	pa	Assessment	a
Letter ID	li	Level 7—Expected pupil response:	
Spelling	s	Reading	r
Other	o	Reading turn-taking	r-tt
Not applicable	9	Orally responding	or
Level 5—Material:		Oral turn-taking	or-tt
Textbook, narrative	tn	Listening	l
Textbook, informational	ti	Writing	w
Narrative trade book	n	Manipulating	m
Informational tradebook	i	Other/not applicable	o/9

Appendix B

Sample of Observational Notes

9:40. A small group is discussing *The Best Christmas Pageant Ever* (Robinson, 1972). The teacher asks, "What happened in the barn? Who else takes part in this? What time of year is it? Let's predict how the family will be part of the pageant." Students offer predictions. Students then read and answer the questions on a sheet: What is a pageant? What is a bad reputation? Teacher circulates. 9:45 6/7 OT (On-task)

c/s/r m1/n/r/ortt c/n/r/ortt r/n/l/r v/s/l/w

levels 1/2/3 4/5/6/7 4/5/6/7 4/5/6/7 4/5/6/7

Appendix C

Description of Classroom Observation Categories Used in Data Analysis (Taylor & Pearson, 2000)

Percentage of time (5-minute segments) coded:

Whole class or large group: All of the children in the class (except for one or two or individuals working with someone else), or a group of more than 10 children. If there are 10 or fewer in the room, code this as a small group.

Small group: Children are working in two or more groups. If there are more than 10 children in a group, call this whole group.

Narrative text: The number of segments in which a narrative textbook (tn) or narrative trade book (n) was coded out of the total number of segments coded.

Informational text: The number of segments that an informational textbook (ti) or information trade book (i) was coded as being used out of the total number of segments coded.

Telling: Telling or giving children information, explaining how to do something.

Recitation: The teacher is engaging the students in answering questions, or responding, usually low-level q-a-q-a. The purpose primarily appears to be getting the children to answer the questions asked rather than engaging them in a formal discussion or fostering independence in terms of answering questions with more complete thinking.

Modeling: The teacher is showing/demonstrating how to do something or how to do a process as opposed to simply explaining it.

Coaching: The teacher is prompting/providing support that will transfer to other situations as students are attempting to answer a question or to perform a strategy or activity. The teacher's apparent purpose is to foster in-

dependence, to get a more complete thought or action rather than to simply get a student to answer a question.

Percentage of all reading segments coded:

Phonemic awareness instruction: Students are identifying the sounds in words or blending sounds together (an oral activity). The purpose is to develop phonemic awareness, not letter-sound knowledge (e.g., Sound Box technique would be coded as "pa" because the focus is on learning the sounds in words).

Phonics instruction: Students are focusing on symbol/sound correspondences (p1) or letter-by-letter decoding (p2) or decoding by onset and rime or analogy (p3), but this is not tied to decoding of words while reading. If students are decoding multisyllabic words, code as p4. We calculated the total number of phonics activities divided by the total number of times reading, coded at level 3.

Word-recognition strategies: Students are focusing on use of one or more strategies to figure out words while reading, typically prompted by the teacher.

Lower-level text comprehension (talk or writing about text): Students are talking (m1) or writing (m2) about the meaning of text that is at a lower level of thinking. The writing may be a journal entry about the text or a fill-in-the-blank worksheet on the text meaning (rather than on a comprehension skill or vocabulary words). The total number of "low-level text comprehension" activities at level 4 out of the total number of times reading, coded at level 3, was calculated.

Higher-level text comprehension (talk or writing about text): Students are talking (m3) or writing (m4) about the meaning of text that is engaging them in higher-level thinking. This is talk or writing about text that is challenging to the children and is at either a high level of text interpretation or goes beyond the text: generalization, application, evaluation, aesthetic response. Needless to say, a child must go beyond a yes or no answer (e.g., in the case of an opinion or aesthetic response). The total number of "high-level text comprehension" activities at level 4 out of the total number of times reading (as the major focus) at level 3 was coded.

Comprehension skill instruction: Students are engaged in a comprehension activity (other than a comprehension strategy) that is at a lower level of thinking (e.g., traditional skill work such as identifying main idea, cause-effect, fact-opinion).

Comprehension strategy instruction: Students are using a comprehension strategy that will transfer to other reading and in which this notion of transfer *is* mentioned (e.g., reciprocal teaching, predicting. If predicting was done but transfer was not mentioned, this would be coded as c).

Vocabulary instruction: Students are discussing/working on word meaning(s).

Active reading practice: Students are reading (not reading turn-taking) at level 7.

Percentage of all codes for student responding:

Active responding: Children are engaged in one or more of the following level 7 responses: reading, writing, oral responding, manipulating. The total number of "active responding" codes coded out of the total number of level 7 responding codes coded was calculated.

Passive responding: Children are engaged in one or more of the following level 7 responses: reading turn-taking, oral responding turn-taking, listening. The total number of "passive responding" codes coded out of the total number of level 7 responding codes coded was calculated.

Time on-task: At the end of the 5-minute note-taking segment, the observer counted the number of children in the room who appeared to be engaged in the assigned task out of all the children in the room. If a child was quiet but staring out the window or rolling a pencil on his desk, this was not counted as on-task.

Note

This research was conducted as part of CIERA, the Center for the Improvement of Early Reading Achievement, and supported under the Educational Research and Development Centers Program, PR/Award No. R305R70004, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. However, the contents of the article do not necessarily represent the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the National Institute on Early Childhood development, or the U.S. Department of Education, and readers should not assume endorsement by the federal government.

References

Baker, L., Dreher, M. J., & Guthrie, J. T. (2000). Why teachers should promote reading engagement. In L. Baker, M. J. Dreher, & J. T. Guthrie (Eds.), *Engaging young readers: Promoting achievement and motivation* (pp. 1–16). New York: Guilford.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models.* Newbury Park, CA: Sage.

Bush, G. W. (2001). *No child left behind.* Washington, DC: Office of the President of the United States.

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

Chin, C. A., Anderson, R. C., & Waggoner, M. A. (2001). Patterns of discourse in two kinds of literature discussion. *Reading Research Quarterly,* **30,** 378–411.

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.

Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children,* **52,** 219–232.

Duffy, G. G., Roehler, L. R., Sivan, E., Rackliffe, G., Book, C., Meloth, M. S., Vavrus, L. G., Wesselman, R., Putnam, J., & Bassiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly,* **20,** 347–368.

Frank, K. A. (1998). Quantitative methods for studying social context in multilevels and through interpersonal relations. *Review of Research in Education,* **23,** 171–216.

Greenwood, C. R., Carta, J. J., Kamps, D., & Delquadri, J. (1995). *Ecobehavioral assessment systems software (EBHASS) practioner's manual* (Version 3.0). Kansas City: University of Kansas, Juniper Garden Children's Project.

Guthrie, J. T. (1996). Educational contexts for engagement in literacy. *Reading Teacher,* **49**(6), 432–445.

Guthrie, J. T., Cox, K. E., Knowles, K. T., Buehl, M., Mazzoni, S. A., & Fasulo, L. (2000). Building toward coherent instruction. In L. Baker, M. J. Dreher, & J. T. Guthrie (Eds.), *Engaging young readers: Promoting achievement and motivation* (pp. 1–16). New York: Guilford.

Houghton Mifflin Invitations to Literacy. (1996). Boston: Houghton Mifflin.

Johns, J. (1997). *Basic reading inventory* (7th ed.). Dubuque, IA: Kendall Hunt.

Kamil, M., Mosenthal, P., Pearson, P. D., & Barr, R. (Eds.). (2000). *Handbook of reading research* (Vol. 3). Mahwah, NJ: Erlbaum.

Knapp, M. S., Adelman, N. E., Marder, C., McCollum, H., Needels, M. C., Padilla, C., Shields, P. M., Turnbull, B. J., & Zucker, A. A. (1995). *Teaching for meaning in high-poverty classrooms.* New York: Teachers College Press.

Kreft, I., & DeLeeuw, J. (1998). *Introducing multilevel modeling.* London: Sage.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie Reading Tests* (4th ed.). Itasca, IL: Riverside.

Michigan Literacy Progress Profile 2000. (1998). Lansing: Michigan Department of Education.

National Reading Panel. (2000). Report of the National Reading Panel. Washington, DC: National Institute for Child Health and Human Development.

No Child Left Behind Act of 2001. Pub. L. No. 107–110, §115 Stat, 1425 (2002).

Pressley, M., Wharton-McDonald, R., Allington, R., Block, C. C., Morrow, L., Tracey, D., Baker, K., Brooks, G., Cronin, J., Nelson, E., & Woo, D. (2001). A study of effective first-grade literacy instruction. *Scientific Studies of Reading,* **5,** 35–58.

Pressley, M., Wharton-McDonald, R., Raphael, L. M., Bogner, K., & Roehrig, A. (2002). Exemplary first-grade teaching. In B. M. Taylor & P. D. Pearson (Eds.), *Teaching reading: Effective schools, accomplished teachers* (pp. 73–88). Mahwah, NJ: Erlbaum.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2000). *HLM: Hierarchical linear and nonlinear modeling* (Version 5) [computer software]. Chicago: Scientific Software International.

Robinson, B. (1972). *The best Christmas pageant ever.* New York: Harper.

Roehler, L. R., & Duffy, G. G. (1984). Direct explanation of comprehension processes. In G. G. Duffy, L. R. Roehler, & J. Mason (Eds.), *Comprehension instruction: Perspectives and suggestions* (pp. 265–280). New York: Longman.

Scanlon, D. M., & Gelzheiser, L. M. (1992). *Classroom observation manual.* Unpublished manuscript, University of Albany, State University of New York, Child Research and Study Center.

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

Taylor, B. M. (1991). *A test of phonemic awareness for classroom use.* Minneapolis: University of Minnesota.

Taylor, B. M., & Pearson, P. D. (2000). The CIERA School Change Classroom Observation Scheme. Minneapolis: University of Minnesota.

Taylor, B. M., Pearson, P. D., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: Lessons about primary-grade reading instruction in low-income schools. *Elementary School Journal,* **101,** 121–166.

Taylor, B. M., Peterson, D. P., Pearson, P. D., & Rodriguez, M. C. (2002). Looking inside classrooms: Reflecting on the "how" as well as the "what" in effective reading instruction. *Reading Teacher,* **56,** 270–279.

Ysseldyke, J., & Christenson, S. (1993–1996). *TIES: The instructional environment system—II.* Longmont, CO: Sopris West.

Copyright © 2003 EBSCO Publishing