Contract No.: 6136-550
MPR Reference No.: ED01CO0039/0010

MATHEMATICA
Policy Research, Inc.

The National Evaluation of Reading Comprehension Interventions: Design Report

Final Report

May 1, 2006

Mathematica Policy Research, Inc.
Susanne James-Burdumy
David Myers
John Deke
Wendy Mansfield

RG Research Group
Russell Gersten
Joseph Dimino

College of Education at the University of Utah
Jan Dole
Lauren Liang

The Vaughn Gross Center for Reading and Language Arts at the University of Texas at Austin
Sharon Vaughn
Meaghan Edmonds

Submitted to:

U.S. Department of Education
Institute of Education Sciences
555 New Jersey Ave., NW
Room 308
Washington, DC 20208

Project Officer:
Audrey Pendleton

Submitted by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director:
David Myers

CONTENTS

CONTENTS *(continued)*

Chapter **Page**

TABLES

FIGURES

Figure **Page**

EXECUTIVE SUMMARY

Title I of the No Child Left Behind Act of 2002 (NCLB) is the largest funding vehicle of compensatory elementary and secondary education programs for disadvantaged children in the United States. This legislation calls on educators to close the gap between low and high achievers by using instructional approaches that scientifically based methods have shown to be effective. Because of limited knowledge about the effectiveness of most instructional approaches, however, it has been difficult for educators to decide how to best use Title I funds to improve the educational outcomes of economically disadvantaged students.

This report lays out the design of a study to evaluate the impact of supplemental reading comprehension interventions for students in upper elementary grades of Title I schools. The process of finalizing the study's design—and its focus on reading comprehension interventions—occurred over a roughly three-year period, beginning in October 2002. Planning for the study took its lead largely from three sources: (1) the Institute of Education Sciences (IES); (2) the Title I Independent Review Panel (IRP), which Congress set up to provide the U.S. Department of Education (ED) with recommendations on Title I evaluation; and (3) specialized technical working groups (TWGs) made up of technical experts in reading comprehension and evaluation design. Below, we list the key design parameters that came out of these efforts. The report itself provides the rationale for these decisions and discusses design and measurement issues, including sample sizes necessary to address the study's key questions.

EVALUATION FOCUS

- The evaluation will focus on testing strategies to improve reading comprehension—particularly interventions designed to improve the reading comprehension of students in upper elementary grades—so those students can make progress in content areas (such as social studies and science) that involve a large amount of expository text.

- The study is designed primarily so that reliable inferences can be made about the effects of each intervention relative to the control group. A secondary focus is on the effects of each intervention relative to the other interventions and the effects of the interventions on selected groups of students, such as those with limited English proficiency (LEP) or low baseline reading skills.

THE PROCESS FOR SELECTING INTERVENTIONS AND SCHOOLS

- A three-stage process was used to select interventions. The first stage involved soliciting proposals from the field. The second stage involved winnowing down the proposals according to a set of initial criteria, which primarily related to whether the submissions included requested materials, such as teacher training materials or samples of classroom materials. In the third stage, a panel of experts assessed the finalists more critically based on factors related to their promise as effective interventions.

- Schools will be selected such that the evaluation sample is geographically diverse and representative of schools with different concentrations of LEP students. In general, schools also will have a high concentration of economically disadvantaged students.

KEY RESEARCH DESIGN PARAMETERS

- *Schools,* rather than classrooms or individual students, will be randomly assigned to receive either an intervention (treatment) or no intervention (control).
- Our power analysis indicates that, under reasonable assumptions, a sample of about 100 schools can achieve the evaluation objectives. These objectives include the ability to test four interventions to determine whether they are effective and whether some are more effective than others.
- We recommend spreading the 100 schools across approximately 10 districts.
- The language in NCLB requires that state assessments be used to gauge student performance. To ensure the study can assess the programs' impacts on students' reading skills, the study will concentrate on assessments that specifically measure reading comprehension.

I. STUDYING READING COMPREHENSION INTERVENTIONS UNDER TITLE I

A. STATEMENT OF THE PROBLEM

Title I of the No Child Left Behind Act of 2002 is the largest funding vehicle of compensatory elementary and secondary education programs for disadvantaged children in the United States. The legislation calls on educators to close the gap between low and high achievers by using instructional approaches that scientifically based methods have shown to be effective. Because knowledge about the effectiveness of most instructional approaches is limited, however, it has been difficult for state and local educators to decide how to best use Title I funds to improve the educational outcomes of economically disadvantaged students.

This report lays out the design of a study to evaluate the impact of supplemental reading comprehension interventions for students in upper elementary grades of Title I schools. The process of finalizing the study's design—and its focus on reading comprehension interventions and their effects on student comprehension of expository text (both in general and in science and social studies)—occurred over a roughly three-year period. It began in October 2002, when the Institute of Education Sciences (IES) contracted with Mathematica Policy Research, Inc. (MPR) to help identify issues relevant to the evaluation of Title I and to propose feasible evaluation design strategies. The process of finalizing design decisions continued after IES contracted with MPR to conduct the evaluation in October 2004.

National statistics on reading achievement provide more insight into the decision to focus the study on reading. The most recent (2005) National Assessment of Educational Progress indicates that 36 percent of the nation's fourth-grade children have difficulty reading (U.S. Department of Education [ED] 2006). Other estimates suggest that as many as 30 percent of

elementary, middle, and high school students have reading problems that severely curtail their educational progress and ultimate educational attainment (Moats 1999).

Educators and policymakers seeking to improve this situation, however, must wade through a massive body of literature to assess what is known and not known about children's reading. Although the National Reading Panel Report (2000), for example, provides some guidance for educators and policymakers, it does not tell us what interventions are best to teach students to comprehend the information in the difficult textbooks they will encounter in school and to learn from that information. This is a particularly vexing problem for children from low-income households, because these children may well be below grade level in their reading even as they enter the intermediate grades. Thus, in establishing best practices for reading in connection with subject matter learning, it is crucial to identify and test intervention programs designed to help students comprehend and learn content from expository texts.

Planning the design for the study took its lead largely from three sources: (1) IES; (2) the Title I Independent Review Panel (IRP), which Congress set up to provide ED with recommendations on Title I evaluation; and (3) specialized technical working groups (TWGs) made up of technical experts in reading comprehension and evaluation design.[1] The decision to focus on reading comprehension, specific content areas, and students in upper elementary grades reflects discussions of the three groups noted above. It also reflects the fact that (1) IES was already devoting considerable effort to understanding the effectiveness of the Reading First program, which targets younger children; and (2) even if the approaches funded under Reading First are effective, many disadvantaged children may still be struggling readers as they enter the

[1]Appendix A of Glazerman and Myers (2004) lists the members of the original design task TWG. Appendix A of this report lists the members of the TWGs for the current evaluation.

higher elementary grades. Additional details of the original design effort are provided in Glazerman and Myers (2004) (hereafter, referred to as the original design report).

B. CONCEPTUAL FRAMEWORK

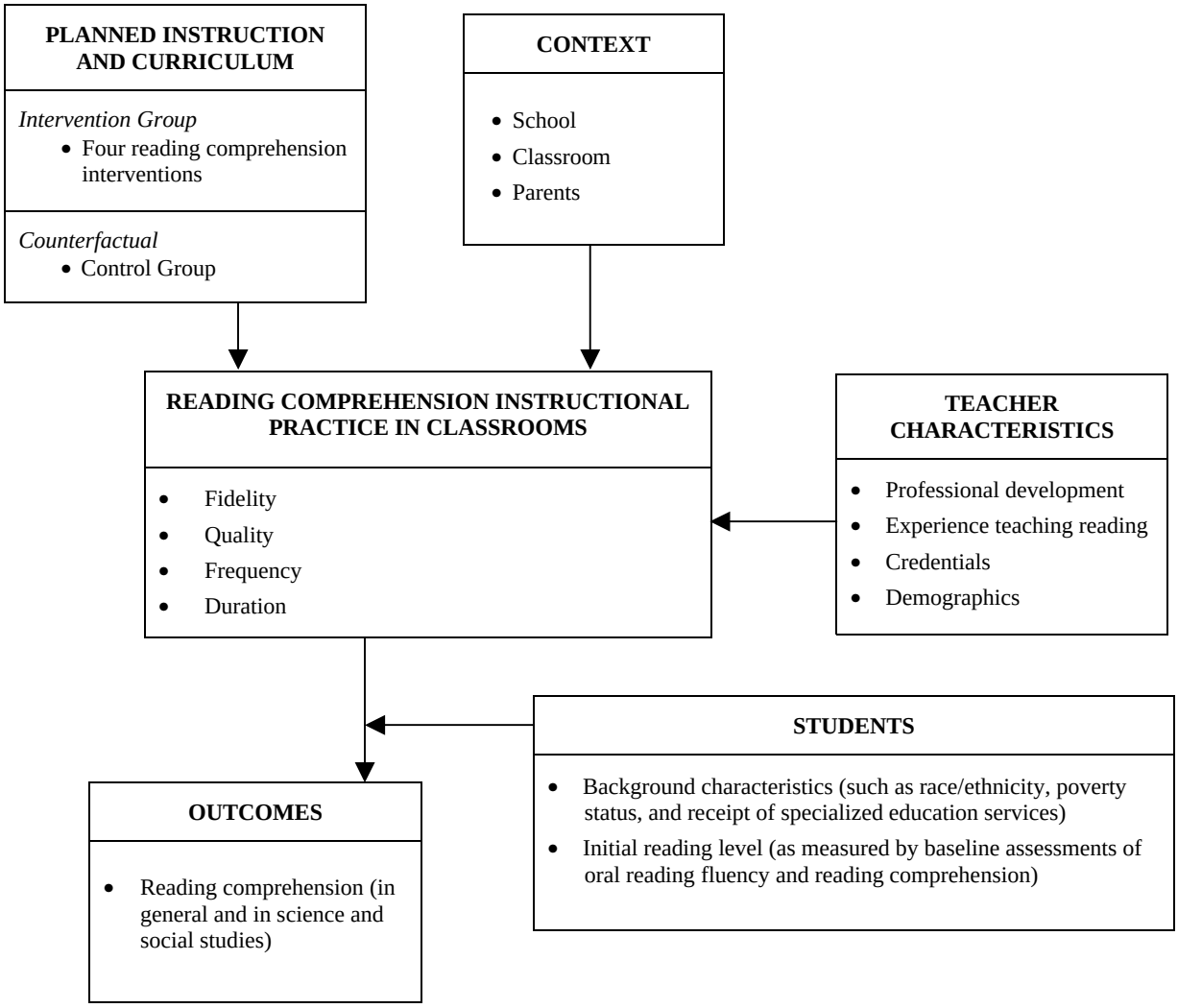
This study provides a unique opportunity to address questions critical to understanding the effectiveness of reading comprehension interventions. We formulated a conceptual framework for the study to help organize our approach to selecting and implementing interventions, setting up the experimental design, collecting data, and computing intervention impacts. The framework (Figure I.1) illustrates what we believe are the major concepts that are relevant for the evaluation of reading comprehension interventions and the pathways through which students' general reading comprehension and comprehension of expository text in content areas, such as science and social studies, will be influenced during the evaluation period. The framework suggests that the reading comprehension curriculum and instruction as it is designed (Planned Instruction and Curriculum box) indirectly affects student outcomes because its effect is mediated by how teachers implement it.[2] For example, whether they use the instructional materials provided by the curriculum developers or adapt other materials for their classes will potentially mediate the effect of the intervention.

Furthermore, we hypothesize that the school and classroom context, as well as teacher characteristics, will affect how the reading curricula and instruction are implemented. For example, as teachers implement the curricula, the level of support they receive from school principals may affect that implementation.

[2] We have not distinguished the curriculum and instruction being offered as part of the reading comprehension interventions from the approaches that would normally be used in the absence of the interventions. The distinction is made later, when the experimental design is set up and schools are randomly assigned to interventions or a control group (nonintervention instruction).

FIGURE I.1

A CONCEPTUAL FRAMEWORK FOR MEASURING THE EFFECTIVENESS OF READING COMPREHENSION INTERVENTIONS



The quality of teacher professional development may also affect the implementation of the curricula. For example, impacts of the curricula may be greater for students whose teachers have participated in larger amounts of professional development. The framework also suggests that students' initial reading skills will affect their outcomes. For example, students who enter the experiment with stronger initial reading skills in either fluency or comprehension (both of which

will be measured at baseline) may develop comprehension skills at different rates than those with weaker skills in either domain.

C. RELEVANT RESEARCH

Research has not helped answer the question of how best to teach students to comprehend the information in the textbooks they will encounter in school. Students typically develop their reading skills in the early elementary grades using narrative texts. When students enter the upper elementary grades, they need to develop ways to understand informational (expository) text. Making this transition can be difficult for some students, particularly those who have not fully developed their reading skills. The text below describes what the existing research can tell us about reading.

The National Reading Panel (NRP) (2000) identified five areas of reading in which research has been conducted to allow for some instructional recommendations for teachers.[3] These areas of reading include (Armbruster and Osborn 2001):

- ***Phonemic Awareness.*** Phonemic awareness is the ability to think about and notice the individual sounds in spoken words; for example, the word "cat" is made up of the sounds /c/ /a/ /t/. Phonemic awareness involves understanding that spoken words are made up of individual speech sounds.
- ***Phonics.*** Phonics is the understanding of the relationship between spoken sounds and written letters. To be able to read words, children must understand this relationship.
- ***Fluency.*** Fluency is the ability to read a text effortlessly and with expression. Fluency consists of accurate reading, at a reasonable rate and with appropriate expression.

[3] The RAND Reading Study Group (2000) also mentioned several other critical elements of comprehension instruction, such as fluency and the need for specific instruction within the context of subject matter learning (for example, in social studies or science).

- ***Vocabulary.*** Vocabulary refers to word knowledge. Children must know and understand many words to be able to comprehend what they hear and read. Greater vocabulary knowledge leads to increased comprehension.
- ***Comprehension.*** Comprehension is making meaning from text. The goal of all reading is understanding what is read.

The NRP (2000) review of the research on comprehension identified specific comprehension strategies, or procedures and routines, that have been demonstrated to be effective for improving comprehension.[4] However, little is known about the role of these strategies in the overall comprehension curriculum or how these strategies may help students learn subject matter content (RAND Reading Study Group 2000).

While little is known about the extent to which comprehension strategies can improve content area learning, research does indicate that all readers find expository text more difficult to comprehend than narrative text (Graesser et al. 1991). There are two main reasons why readers have difficulty with expository text: (1) the organization of expository text is often unfamiliar, and (2) the content of expository text is often unfamiliar. The combination of an unfamiliar organization and unfamiliar content makes information in expository texts, especially textbooks, difficult to comprehend. In fact, researchers have hypothesized that the legendary "fourth-grade slump" may well be due to students' inexperience with expository texts (Chall et al. 1990).

Below, we review the recent research to illustrate which reading comprehension strategies may be effective. This information is critical in designing an evaluation, which includes the selection of interventions that could affect children's ability to extract meaning and understanding from text. Toward this end, we present the following:

[4] Some of the strategies are learner strategies, such as using prior knowledge, generating questions, and summarizing, that readers use to make sense of text. Other strategies are teaching strategies, such as cooperative learning and graphic organizers, that teachers use in the classroom to help students make sense of text.

- Key reading comprehension theories and concepts
- Findings from several syntheses that present information on the effectiveness of reading comprehension strategies
- An assessment of the limitations of the current research on reading comprehension

1. Reading Comprehension Theories and Concepts

Research on comprehension is based on a few theories and concepts:

- ***Schema Theory.*** This theory suggests that what students know about a topic or a construct influences how much they can or will learn by reading a passage that addresses that topic (Anderson and Pearson 1984). Schema theory, which was the building block for reading comprehension research beginning in the 1970s, implies that the more students read and learn about a topic, the easier it will be for them to understand that topic the next time it comes up.
- ***Metacognition.*** Metacognition refers to students' conscious awareness of the cognitive processes they use and anything related to those processes (Flavel 1976). In reading, this concept explains students' awareness of whether they understand what they read and their ability to change and modify the strategies they use in order to help them comprehend better.
- ***Vygotsky's Theory of Cognitive and Early Social Constructivism.*** This theory suggests that teachers serve as models and facilitators of verbal interactions that lead to internal understandings about comprehension processes (1964, cited in Palincsar 1986). Palincsar and Brown (1984) built upon this theory by postulating that students would become more adept at using comprehension strategies through the use of shared group dialogue among small groups of readers who jointly build meaning from text.

Each of these theories led to different lines of research on comprehension instruction.[5] Eventually, a large body of work was generated, and researchers began to ponder what the collective evidence had to say and whether broader lessons could be drawn. To try to answer these questions, researchers generated syntheses of the existing research. Next, we present the

[5] For example, research based on schema theory often involved testing whether an intervention to build students' knowledge on a given topic would help students' reading comprehension (see, for example, Au and Crowell, 1979 and Langer 1981 and 1984). Research based on Vygotsky's theory involved testing whether the use of focused, group dialogue would help students make sense of text.

findings from these syntheses and related research that support the reading comprehension interventions included in the study.

2. Review of Research Syntheses

The review of research syntheses suggests six techniques and approaches for teaching reading comprehension: (1) teach strategies; (2) use proven methods for instructional delivery; (3) use embedded instruction; (4) teach with highly engaging, interesting texts; (5) use cooperative learning; and (6) participate in ongoing professional development. In discussing these techniques and approaches, we address some unresolved issues in the comprehension instruction research that may be informed by this evaluation. We end by mentioning some limitations in the research.

Teach Strategies. One approach linked to teaching reading comprehension effectively is to teach students to use comprehension strategies to help them actively make meaning out of the texts they read. Research has established that teaching students to use these strategies helps them better understand the text they read. Two recent reviews of the literature—the NRP report and a review of intervention research for students with reading disabilities (Gersten et al. 2001)—indicate that, in general, interventions that teach flexible use of multiple strategies to improve comprehension produce large improvements in reading comprehension.

There is also general agreement in the literature on the types of strategies that should be included in strategy instruction (Pearson et al. 1992; Pressley 2002; National Reading Panel 2000; RAND Reading Study Group 2000; Pressley et al. 1989). These strategies generally fall into one of three groups: (1) summary and prediction, (2) question generation, and (3) understanding of text structure and use of graphic organizers. We discuss each of these groups next:

1. ***Summary and Prediction.*** In *summarizing*, students condense information they have read into the essential or main points. Summarizing encompasses multiple strategies for comprehending text, such as determining what is important, categorizing, and organizing information (Brown and Day 1983). The NRP report found that comprehension of expository text is enhanced when students are taught to record their ideas about what they read and to summarize these ideas after reading long passages. In *prediction*, students make predictions about what they are about to read based on subtitles or material in the preceding paragraphs. For example, before reading a passage, students make predictions about what they are about to read and, after reading the passage, they evaluate the accuracy of their predictions. Hansen (1981) and Hansen and Pearson (1983) have shown that teaching students to use prediction while reading improves their comprehension.

2. ***Question Generation.*** *Question generation* involves asking questions while reading and attempting to answer those questions. This strategy makes readers more active in the comprehension process and focuses the readers' attention, particularly on the information that will answer the self-generated questions. This makes the text content easier to understand and to remember (Martin and Pressley 1991). Question generation seems especially useful for learning material from expository text (Pressley et al. 1992; Wood et al. 1990; and Rosenshine et al. 1996). The NRP report also found support for interventions that provided opportunities for students to ask and answer their own questions about the text.

3. ***Understanding of Text Structure and Use of Graphic Organizers.*** *Text structure* refers to how writers and readers organize important information in a text (Meyer et al. 1980). Four basic text structures in expository text are (1) compare-contrast, (2) cause-effect, (3) explanation, and (4) sequencing.[6] Readers who are aware that authors typically use these text structures to organize information tend to recognize and retain the important information more easily (Snow 2002). There have been few studies on the use of text structure to understand expository text because many of the expository texts students read in school have a "mixed" structure (for example, some cause-effect with a good deal of explanation or sequence). The text structure studies that do exist most often involve the use of *graphic organizers* (visual presentations of information to help students understand text). The NRP report suggests that text structures and graphic organizers may be promising strategies for understanding expository text—however, because there are few studies on this issue, it does so tentatively.

[6] A compare-contrast structure requires readers to compare and contrast one event or object with another. For example, tornadoes and hurricanes are alike in that they both can cause damage to the environment, but they are different in that they arise from different weather patterns. A cause-effect structure requires readers to infer one event causing another (for example, hurricanes are caused by …). An explanation structure provides an explanation for something (for example, there are many different kinds of precipitation. One kind of precipitation is…). A sequencing structure lays out a linear sequence of events (for example, first, second, next, finally).

Use Proven Methods for Instructional Delivery. A second approach linked to teaching reading comprehension effectively is to teach strategies using proven methods of instructional delivery. Here, we outline three recommended methods of delivering instruction:

1. ***Direct, or Explicit, Instruction.*** *Direct, or explicit, instruction* of comprehension strategies involves modeling how the comprehension strategy or skill is used, guiding practice with feedback from a teacher, and providing opportunities for students to independently practice using the strategy or skill on various reading materials (Rosenshine and Stevens 1986). Early randomized trials using explicit instruction and targeting only one specific skill or strategy demonstrated significant improvements in students' understanding of expository text, ability to critically analyze arguments, and retention of content (Adams et al. 1982; Darch and Gersten 1986; Darch and Kame'enui 1987; Lloyd et al. 1980; Patching et al. 1983).
2. ***Direct Explanation of Strategies.*** *Direct explanation of strategies* is similar to explicit instruction in that the three most important elements of explicit instruction—modeling, guided practice, and feedback—are included, however, there is more focus on the methods used by teachers to explain the strategies. Teachers first *name and explain*, or define, the particular strategy for the students, describe *when* and *how* it might best be used, and tell students *why* the strategy is important for improving reading. They next engage in a significant amount of explanation and cognitive modeling to show the students how to use the strategy. The teacher engages the students in practicing the strategy in teacher-mediated activities until students are able to use the strategy independently. The direct explanation of strategies has been shown to be particularly effective in comprehension strategy instruction (Duffy et al. 1987; Duke and Pearson 2002; National Reading Panel 2000; RAND Reading Study Group 2000).
3. ***Differentiated Instruction.*** *Differentiated instruction* is instruction that is individualized in an attempt to meet the specific needs of each student (Fuchs and Fuchs 2005). Students differ in their instructional needs, not only in reading level, but also in the strategies and skills they need to be taught. Interventions that attempt to target individual student needs, rather than group needs, should result in improved comprehension. For example, computer programs can differentiate instruction by identifying particular skills in which students are weak and providing targeted practice in those skills.

Use Embedded Instruction. A third approach linked to teaching reading comprehension effectively is to embed strategy instruction into the reading of text in different academic content areas. The idea behind this approach is that if students are taught the strategies using the

expository texts they read during their content area classes, they will more likely transfer their use of the strategies to texts they read in other content areas, and on their own.[7]

Interventions that embed strategy instruction into content area reading activities still include the direct explanations and teacher cognitive modeling associated with the direct explanation approach to teaching comprehension strategies. However, these periods of explicit instruction are generally shorter, individual strategies may be introduced more closely together, and the occasion to teach a strategy is often more natural, arising from a spontaneous classroom event or interaction.

The most well known of the embedded strategy instruction interventions is transactional strategies instruction (Pressley et al. 1998; Pressley 2002). In this approach, the teacher and students jointly develop meaning by strategic reasoning through a text together. Studies suggest this approach may be highly effective for improving reading comprehension (Brown et al. 1996; Collins 1991; Anderson 1992; Anderson and Roit 1993).

Teach with Highly Engaging, Interesting Texts. A fourth approach linked to teaching reading comprehension effectively is to use highly engaging, interesting texts when teaching reading comprehension. Interesting texts are defined as texts that have vivid details, are relevant to the task, are easily accessible to students, and have colorful photographs and illustrations (Schraw et al. 1995). Many studies have demonstrated the value of using interesting texts to teach reading comprehension (Anderson et al. 1987; Asher 1980; Guthrie et al. 1998; Guthrie et al. 2000a; Guthrie et al. 2000b).

[7] Some researchers argue that, when strategy instruction is taught in isolation (for example, on workbook pages during reading instruction), students do not make the transfer from workbook pages to their own reading of expository texts, especially to what they read for social studies or science (Pearson and Fielding 1991; Pressley 2000). This may occur because students perceive the learning activity to be completing the workbook pages, not learning strategies to use with real texts.

Use Cooperative Learning. A fifth approach linked to teaching reading comprehension effectively is to use cooperative learning strategies in the classroom. In cooperative learning, students can interact with peers when discussing text. Research has shown that interactions between peers are critical in getting students to express their thoughts and, ultimately, to internalize strategies. Both the NRP report and Gersten et al. (2001) noted that giving children the chance to practice a strategy in a small group contributed to the success of most interventions.

Participate in Ongoing Professional Development. A sixth approach linked to teaching reading comprehension effectively is to provide teachers with ongoing professional development in the teaching of reading comprehension strategies. According to the NRP report, ongoing professional development is important for successful implementation of instructional plans related to reading comprehension strategies.[8] Several studies demonstrate that, when teachers are provided with sufficient professional development, their teaching of comprehension strategies improves (Anderson 1992; Brown et al. 1996). In addition, Duffy et al. (1987) found that ongoing professional development consisting of one-on-one coaching, collaborative sharing, and lesson observation and feedback helped teachers learn how to teach comprehension strategies.

3. Limitations in the Research

Many studies in the syntheses and related research on comprehension instruction have one or more serious methodological flaws that limit their scientific merit and, therefore, their value in program and policy development. The most consistent flaws include the use of (1) quasi-

[8] The report concluded that teachers *can* learn to use effective comprehension instruction practices in the classroom.

experimental designs to estimate impacts, (2) materials the researchers developed themselves that were sometimes quite different from those that students typically encountered, (3) outcomes that were closely aligned to the strategy being taught and thus could not be used to assess general comprehension ability, and (4) short-term follow-up data on outcomes. In addition, many studies did not assess fidelity of implementation—the extent to which the intervention as implemented resembled the intervention model. Finally, other than Fuchs et al. (1997) and Vaughn et al. (2001), the majority of studies were based on instruction delivered to students by well-trained graduate students or teachers personally trained by the researchers. Thus, we know little about how useful the interventions would be in classrooms in which teachers were not exposed to such training. Because of these limitations, one should interpret many of the findings as suggestive of the effects we might expect to obtain in the context of a well-designed experiment with random assignment of subjects to intervention or control groups. For these reasons, we are limited in what we can infer from this body of research.

D. DESIGN OVERVIEW

The experimental design for the study reflects (1) careful consideration of the questions posed by IES, the Title I IRP, and TWGs that worked with MPR; (2) our knowledge that the interventions would be implemented in classrooms; (3) IES's suggestion that up to four interventions should be considered; (4) the study team's suggestion that the focus of the study would be upper elementary grades (fifth graders); and (5) concerns about spillover effects. To address these issues, we will use a cluster randomized experiment with four intervention groups and a control condition. This straightforward design is based on the random assignment of schools to the intervention groups or a control group.

Because the study involves estimating impacts of reading comprehension interventions relative to what students would have experienced without the intervention, we believe it is

important that all aspects of the schools other than the intervention, including the allocation of students to classrooms by school principals, remain the same as they were before the intervention. Accordingly, we will suggest that school principals make classroom assignments as they normally would. Therefore, we recommend that students from all fifth-grade classrooms be included in the interventions and evaluation.

The study includes baseline data collection during fall 2006 and follow-up data collection during spring 2007, with the potential for a second year of data collection, depending on the study's first-year results. Thus, the implementation of the interventions (and the measurement of the impacts of those interventions) will cover roughly one school year. This data collection schedule implies that these curricula have the potential to affect outcomes during a single school year.

When estimating sample size requirements for the design, we factored in the following four statistical considerations: (1) the need to account for multiple comparisons when conducting tests of statistical significance, (2) whether we can estimate district fixed effects in the impact analysis, (3) the benefits of a baseline test to increase precision, and (4) IES's desire to detect differences between each intervention group and the control group equal to an effect size of 0.25 or larger with a high probability. Furthermore, we worked from key assumptions concerning the expected number of fifth-grade classes within schools, the number of students within classes, and the expected correlation between pre- and post-test measures of reading achievement.[9]

[9] To estimate the number of schools needed for the evaluation, we assumed the typical elementary school would have three fifth-grade classrooms and about 26 students in each classroom. We also assumed that (1) 50 percent of the variance in test scores could be accounted for with knowledge of the pretest scores, (2) the desired probability of detecting a significant impact if there was an impact was 0.80, and (3) the probability of rejecting the hypothesis of no impact was 0.05 (two-sided). Furthermore, we assumed that 10 percent of the total variance in test scores was attributable to between-school differences.

The statistical considerations outlined above suggest an experimental design that includes 100 schools (20 in each of four intervention groups and 20 in the control group). With a total sample of 100 schools, it is important to include multiple school districts in the evaluation. In addition, it will be valuable both to have the districts geographically dispersed and to minimize the burden on the districts by not selecting too many schools within any one district. Given these considerations, we are planning on a design with roughly 10 school districts, with each district contributing about 10 schools to the study. A natural representation of an experimental design in which schools are randomly assigned within districts to intervention and control groups is the randomized blocks design (Kirk 1968). Blocking on school districts not only ensures a representation of schools in districts and interventions but also could increase the precision with which impacts can be estimated.

In addition to using the school district as a blocking factor in the experimental design, we will consider using additional blocking factors when implementing the design. The blocking factors we will consider include the percentage of students within schools eligible for free and reduced-price lunches and average reading achievement for fifth graders. As the recruiting of districts and schools progresses, we will assess the need for blocking on these factors, depending on the extent to which these characteristics vary across the participating schools in each district.

E. PLAN OF THE REPORT

Chapter II outlines the procedures for selecting interventions and schools for the study. Chapter III discusses the random assignment design, sample size requirements, and data collection and data analysis plans.

II. SELECTING INTERVENTIONS AND SCHOOLS

A. INTERVENTION SELECTION AND IMPLEMENTATION

1. Intervention Selection

To select the best mix of reading comprehension programs, we designed a competitive process that drew on the expert judgments of a panel of nationally recognized reading researchers. The competition was announced in *Education Week,* and potential candidates were notified by email. Interested parties could contact the study team and request that a hard copy of the Request for Proposals (RFP) be mailed to them. Shortly afterward, we posted the RFP on the study's website. The RFP invited researchers and partnerships of researchers, publishers, and organizations involved in professional development for instruction in reading comprehension to apply to participate in the study by submitting a proposal within the next 30 days.

Seven candidates submitted a proposal in response to the January 12, 2005, web posting of the RFP. We then began a three-stage review process. First, we reviewed all proposals for the following minimum qualifications:

- Submission includes a technical proposal.
- Submission includes samples of teacher training materials.
- Submission includes samples of classroom materials (including classroom support materials).
- Submission includes a budget.

Second, we forwarded the proposals that met the minimum qualifications (all seven did so) to the panel of experts, who reviewed and evaluated the proposals using the criteria in the table below:

Criteria	Points
I. Summary description of intervention, theoretical and empirical support for the intervention content, and evidence of the intervention's efficacy or effectiveness a. Underlying theory with empirical support (25 points) b. Evidence to support the effectiveness of the intervention (10 points)	35 points
II. Quality of the proposed intervention design a. Objectives of intervention, including description of teacher practices and skills that comprise the intervention (10 points) b. Intensity and quality of teacher training design and follow-up support design (10 points) c. Quality of training and support materials, quality of classroom activity materials, and quality of any intervention-specific assessments (10 points)	30 points
III. Institutional capability to provide training and follow-up support (staff qualifications, capacity to schedule and manage training)	20 points
IV. Appropriateness of intervention a. For target population (grade 5, Title I schools) (5 points) b. For content (comprehension of expository text in social studies or science) (10 points)	15 points

We also considered the reasonableness of proposed costs for program materials, training, and follow-up support, in conjunction with the proposed training design and the justification provided for costs.

Third, the five bidders with the highest ratings from the reading experts were selected for in-person interviews on March 10–11, 2005. The finalists presented an overview of their proposals and responded to questions from the expert panel, which then recommended three programs to participate in the pilot study. (The next section contains more information about the pilot study.) IES approved these recommendations.

After completing this competition, we decided to hold a second competition, for two reasons: (1) the evaluation plans called for up to four reading comprehension programs and (2) we believed that a second solicitation would bring forward additional promising programs. The second competition followed procedures identical to those used in the first competition. We reissued the RFP on April 29, 2005, and another six proposals were submitted. The expert panel recommended three program developers for interviews, which were conducted on June 30, 2005.

Based on the interviews, the expert panel and project staff selected two additional interventions for the pilot study. As with the first competition, IES approved the recommendations.

2. Intervention Implementation

Each of the five program developers selected for the pilot identified three or more schools in which it would implement its intervention in the 2005–2006 academic year. This pilot allowed developers to augment and refine their teacher training design and materials, as well as their classroom materials, and to implement their interventions in a small number of classrooms (nine classrooms per intervention) under conditions consistent with the upcoming full evaluation. The reading experts reviewed all implementation plans, designs, and materials before they were piloted. The study team observed (1) developers' training of teachers, and (2) teachers' classroom instruction using the developers' curricula. They then provided feedback to the developers to guide further refinements and improvements in their designs and materials before the full implementation year. Developers were required to respond to the comments but could decide what revisions, if any, they wanted to make.

In the spring of the 2006–2007 school year, we will select up to four of the five interventions that constitute the best mix of programs to be continued after the pilot-test year. The selection of interventions for the full implementation will be based on the criteria in Table II.1 and will follow these steps:

1. We will provide the Intervention TWG with (1) expert reviews of developer materials, (2) developer responses to the reviews, (3) reports on teacher training observations and the first round of classroom observations, and (4) a summary of the TWG's evaluation of the developers' initial proposals.

2. Each Intervention TWG member will independently review the above materials and assess each intervention based on the criteria provided (Table II.1). Each TWG member will then provide a preliminary assessment of the curricula to the study team.

3. We will convene a teleconference of the TWG to discuss the interventions and identify key strengths and weaknesses and distinguishing features of each intervention.

4. We will send IES recommendations of which reading programs should continue into the full implementation.

TABLE II.1

SELECTION CRITERIA FOR FULL IMPLEMENTATION

1. Meets contractual requirements for pilot-test year.
2. Ease of use for teacher. a. Materials and activities are readily integrated into classroom routines (e.g., teacher's guide provides lesson plans that are easy to follow; student materials have a wraparound teacher's guide; activities, including computer applications, are functional). b. Teacher-friendly materials (e.g., lessons follow similar format; use of color or graphics makes lesson plans or scripts appealing and easy to follow).
3. Intensity/duration of teacher professional development. a. Duration of initial training and follow-up support are commensurate with (or adequate for) program complexity. b. Initial training and follow-up support are sufficient in motivating teachers to implement program as intended. c. Initial training and follow-up support are well specified.
4. Program is well specified and robust. a. Program activities are clearly outlined and tied to expository reading comprehension objectives. b. Program activities can be satisfactorily implemented by teachers with a range of teaching skill or experience.
5. Developer has the capacity to support large-scale implementation. a. Developer has sufficient staff to support up to 20 schools. b. Training and support model is adequate to ensure fidelity of implementation.
6. Theoretical and empirical support for the program's content and effectiveness. a. Effectiveness of program's strategies based on prior theory or research. b. Effectiveness of program based on program-specific empirical research.

The full implementation of up to four selected interventions in 100 schools will occur in the 2006–2007 school year. During this year, the intervention team will continue to monitor the

implementation of each intervention but will not provide formative feedback on designs or materials. Chapter III describes how the information the study team collects about the implementation of the curricula will be used.

B. SELECTING AND RECRUITING DISTRICTS AND SCHOOLS

Recruiting districts and schools is one of the most formidable challenges the study faces because (1) the schools for the evaluation must meet specific criteria and (2) there is no federal requirement to participate in the evaluation. Our recruiting approach deals with this challenge by (1) initially targeting more school districts and schools than will actually end up in the study, (2) seeking support from high-level education officials, (3) appealing directly to school staff who may be affected, and (4) offering modest cash payments to defray the potential burden associated with participation.

1. Identifying Eligible Districts and Schools

The criteria we used for selecting eligible districts reflect three main concerns: (1) the recommendation of the Title I national assessment IRP to focus on high-poverty Title I schools, (2) the need to include districts with enough schools to make it possible for the study to assign two schools to each of the treatment groups (four experimental and one nonexperimental), and (3) the desire to select schools and districts that would provide face validity to the study.

We first created a list of eligible districts. To create the list, we used data from the 2003 Common Core of Data (CCD) school- and district-level files. To be included, each district had to have:

- At least 12 schoolwide Title I schools[1]

[1] Individual public schools with poverty rates above 40 percent may use Title I funds, along with other federal, state, and local funds, to operate a schoolwide program to upgrade the instructional program for the whole school.

- At least 40 percent of students in the schools noted above who were eligible for free or reduced-price lunch
- At least 60 fifth-grade students in the schools noted above

A total of 157 school districts met the criteria listed above.[2]

To begin our recruiting efforts, we selected a subset of 28 districts from the pool of 157 eligible districts. The process used to identify the 28 districts, which we will approach first about participating in the study, reflects the desire to include districts that are geographically dispersed throughout the United States. We randomly selected 7 districts from each of four regions (south, west, midwest, and northeast) for a total of 28 randomly selected districts.[3,4,5,6] Table II.2 lists the randomly selected districts.

Our goal was *not* to create a statistically representative sample of districts. Instead, we used the stratified random selection to help us select a diverse set of districts to approach about participating in the study. While the set of randomly selected districts is fairly well balanced on the key characteristics shown in the table, there were a few areas in which the sample could have been more balanced. In particular:

[2] Schools that participated in the pilot year cannot participate in the full implementation of the study, because teachers in those schools would have a year of experience with the reading programs and would thus not be comparable to teachers at nonpilot schools (either experimental or nonexperimental) selected for the study.

[3] The south region contains Alabama, Arkansas, the District of Columbia, Delaware, Georgia, Florida, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia.

[4] The west region contains Arizona, Alaska, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming.

[5] The midwest region contains Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin.

[6] The northeast region contains Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont.

- The set of districts would benefit from more representation among suburban and rural schools. Among the 157 eligible districts, there were seven suburban (and one rural) schools on average, while the randomly selected set of districts included just three suburban (and no rural) schools on average.[7]

- The set of districts would benefit from more geographic diversity in the south and west regions (where many Texas and California districts were selected).

TABLE II.2

LIST OF RANDOMLY SELECTED DISTRICTS

District Name	City	State	Number Urban Schoolwide Title I Schools with at Least 60 Fifth Graders	Number Suburban Schoolwide Title I Schools with at Least 60 Fifth Graders	Number Rural Schoolwide Title I Schools with at Least 60 Fifth Graders
Midwest					
City of Chicago School District	Chicago	IL	196	1	0
Cincinnati City School District	Cincinnati	OH	17	2	0
Cleveland Municipal City Schools	Cleveland	OH	54	0	0
Columbus Public Schools	Columbus	OH	41	2	0
Milwaukee School District	Milwaukee	WI	45	0	0
Gary Community Schools	Gary	IN	15	0	0
Omaha Public Schools	Omaha	NE	18	1	0
Northeast					
Philadelphia City School District	Philadelphia	PA	104	0	0
Jersey City Public Schools	Jersey City	NJ	0	17	0
Bridgeport School District	Bridgeport	CT	16	0	0
Rochester City Schools	Rochester	NY	20	0	0
Springfield Public Schools	Springfield	MA	15	0	0
Providence School District	Providence	RI	21	0	0
Brockton	Brockton	MA	12	0	0
South					
Goose Creek School District	Baytown	TX	12	0	0
Richardson Independent School District	Richardson	TX	15	0	0
Fort Worth Independent School District	Fort Worth	TX	41	2	0
McAllen Independent School District	McAllen	TX	13	0	0
Birmingham City	Birmingham	AL	22	0	0
Ector County Independent School District	Odessa	TX	20	0	0
Fulton County	Atlanta	GA	1	14	3
West					
Sacramento City Unified	Sacramento	CA	18	5	0
Compton Unified	Compton	CA	15	4	0
Granite School District	Salt Lake City	UT	0	13	0
Oxnard Elementary	Oxnard	CA	13	0	0
Alhambra City Elementary	Alhambra	CA	4	8	0
Adams-Arapahoe (Aurora)	Aurora	CO	14	0	0
Albuquerque Public Schools	Albuquerque	NM	24	8	0

[7] The urban, suburban, and rural groups were created using location information from the CCD. Large city (population greater than 250,000) and mid-size city (population less than 250,000) categories are included in the urban group. Fringe of large city, fringe of mid-size city, and large town (population greater than 25,000) are included in the suburban group. The small town (population less than 25,000) and rural categories are included in the rural group.

To address these issues, we recommended that IES make the following changes to the districts randomly selected:

- Replace Philadelphia, Pennsylvania, with Paterson, New Jersey, because Paterson would provide suburban schools.
- Replace Goose Creek, Texas, with Marion County, Florida, because Marion County would provide suburban and rural schools and more geographic diversity in the southern region.
- Replace Ft. Worth, Texas, with Robeson County, North Carolina, because Robeson County would provide rural schools and give us more geographic diversity in the southern region. In addition, the list already includes another district (Richardson, Texas) in the Dallas area.
- Replace Compton, California, with Tacoma, Washington, because Tacoma would provide more geographic diversity in the west region. In addition, the list already includes two other districts (Oxnard and Alhambra) in the Los Angeles area.

Table II.3 reflects these recommendations.

TABLE II.3

LIST OF RECOMMENDED DISTRICTS

District Name	City	State	Number Urban Schoolwide Title I Schools with at Least 60 Fifth Graders	Number Suburban Schoolwide Title I Schools with at Least 60 Fifth Graders	Number Rural Schoolwide Title I Schools with at Least 60 Fifth Graders
Midwest					
City of Chicago School District	Chicago	IL	196	1	0
Cincinnati City School District	Cincinnati	OH	17	2	0
Cleveland Municipal City Schools	Cleveland	OH	54	0	0
Columbus Public Schools	Columbus	OH	41	2	0
Milwaukee School District	Milwaukee	WI	45	0	0
Gary Community Schools	Gary	IN	15	0	0
Omaha Public Schools	Omaha	NE	18	1	0
Northeast					
Paterson City Schools	Paterson	NJ	0	20	0
Jersey City Public Schools	Jersey City	NJ	0	17	0
Bridgeport School District	Bridgeport	CT	16	0	0
Rochester City Schools	Rochester	NY	20	0	0
Springfield Public Schools	Springfield	MA	15	0	0
Providence School District	Providence	RI	21	0	0
Brockton	Brockton	MA	12	0	0
South					
Marion County Schools	Ocala	FL	7	8	12
Richardson Independent School District	Richardson	TX	15	0	0
Robeson County Schools	Lumberton	NC	0	0	14
McAllen Independent School District	McAllen	TX	13	0	0

TABLE II.3 *(continued)*

District Name	City	State	Number Urban Schoolwide Title I Schools with at Least 60 Fifth Graders	Number Suburban Schoolwide Title I Schools with at Least 60 Fifth Graders	Number Rural Schoolwide Title I Schools with at Least 60 Fifth Graders
Birmingham City	Birmingham	AL	22	0	0
Ector County Independent School District	Odessa	TX	20	0	0
Fulton County	Atlanta	GA	1	14	3
West					
Sacramento City Unified	Sacramento	CA	18	5	0
Tacoma School District	Tacoma	WA	14	0	0
Granite School District	Salt Lake City	UT	0	13	0
Oxnard Elementary	Oxnard	CA	13	0	0
Alhambra City Elementary	Alhambra	CA	4	8	0
Adams-Arapahoe (Aurora)	Aurora	CO	14	0	0
Albuquerque Public Schools	Albuquerque	NM	24	8	0

These changes improve the geographic diversity of the districts in the southern and western regions and the likelihood of having suburban and rural schools included in the study. IES agreed with the above recommendations, and recruiting has started with the districts recommended in Table II.2.

Another school characteristic on which we will attempt to achieve some balance is English-language proficiency. Geographic distribution and a low-income student population may produce a reasonably diverse study sample. It is important to ensure that students with limited English proficiency (LEP) are also represented, however, because the Title I IRP singled out LEP students as a group that should be represented in the study. This implies that, to support subgroup analysis, a sizable fraction of schools should have some representation of LEP students. If possible, it would be desirable to have some schools with a small percentage of LEP students and some with a high percentage of LEP students so that we could observe whether interventions are effective for LEP students in each type of setting. We will consider this as we talk with districts and schools about participating in the study.[8]

[8] One factor that MPR considered as a possible criterion was whether the school was implementing Reading First, the federal government's program to promote core reading skills in the early elementary grades. We believe that, if possible, it would be useful to observe the effectiveness of comprehension interventions in both types of settings—those with and without Reading First programs. For those schools that have Reading First programs, the

If we need to select back-up districts (beyond the initial 28 districts identified above) to reach our target number of schools, we will do so from the eligible pool of 157 districts noted above. We will attempt to select back-up districts to help ensure distribution both geographically and in key district characteristics.

2. Recruiting Districts and Schools

To make it easier to negotiate with districts and schools, we will begin recruitment efforts at the state level. State education officials can provide important contextual information on the school districts we plan to contact about participating in the evaluation. In addition, their explicit support may be able to sway district- or school-level officials and increase the likelihood that the districts and schools will participate. We will telephone the chief state school officers in the relevant states and discuss with them (or their designees) (1) the purpose of the study, (2) the fact that the burden of participation will be minimal, (3) the data to be collected and how they will be used, (4) our plans for protecting the confidentiality of school and student data, and (5) the potential benefits of participation.

Some of the benefits we will stress, first to state officials and later to district and school officials, are (1) the opportunity for districts and schools to try new, promising approaches for improving the reading comprehension of low-income children; (2) teachers who are trained in these approaches with no out-of-pocket expenses for the school districts; and (3) the chance to

(continued)
proposed comprehension interventions may be seen as a follow-on, to continue with innovative practices that bring reading skills developed in early elementary grades to enhance those skills in the middle elementary grades.

contribute to a knowledge base with rigorous evidence about what works in teaching reading comprehension—information that policymakers and educators across the country can use.[9]

After we talk with state officials and gain support from some or all of them, we will contact school districts. In the initial telephone calls with school districts, we will introduce the study and discuss its requirements. One important issue to discuss with district officials is parental consent. The Family Educational Rights and Privacy Act (FERPA) exempts studies conducted under contract to ED (which acts as a representative of the Secretary of Education) from needing parental consent for the collection of school records. The study was ruled exempt from Institutional Review Board (IRB) review through the provision of Title 45 of the Code of Federal Regulations (CFR), Part 46 Protection of Human Subjects, Section 101(B)(1), which exempts "research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods." However, our experience indicates that some districts may nevertheless require parental consent. Plans for gaining parental consent will be one of several issues to be made explicit in the work plan that MPR will prepare for each school district and school. The work plan, which will be agreed upon by local school officials, will outline the roles of the research team and the school district and schools.

After gaining district-level approval to move forward and determining which schools to target, we will set up an in-person meeting with the district and school staff to provide district and school administrators with much the same information about the study as presented to state

[9] For example, information gleaned from the evaluation of reading comprehension interventions will fit the requirements for the What Works Clearinghouse funded by IES. Our reports will present information that meets the requirements of data reporting as outlined by the Clearinghouse.

officials. A reading expert from the study team will also attend these meetings to provide information to districts and schools about the curricula. To participate in the evaluation, school officials must be receptive to random assignment and understand how it will be implemented. We base this statement on the approach MPR used in the Closing the Reading Gap study, a similar evaluation in which 50 schools across 27 school districts were randomly assigned to four intervention groups. That experience suggested that a critical aspect of the negotiation process with school staff was to clearly explain the random assignment procedures and the interventions.

Our plan for implementing the experimental design, described in the next chapter, is based on that experience. First, we plan to maintain the integrity of random assignment by explaining to the school and district administrators why schools assigned to the control group should do nothing more than they would normally have done in the absence of the evaluation. They should neither adjust their current reading comprehension curriculum nor implement a new curriculum unless all schools in the district are doing so. This means that participating schools and fifth-grade teachers should continue implementing their current reading curricula, even if that means that some teachers are currently implementing another supplemental reading curriculum or using techniques that are part of the interventions the study is testing. The contrast the study is making involves the comparison of each intervention to the control group, which may include teachers who are teaching strategies similar to those being taught in the interventions the study is testing. Because random assignment of a school to a reading comprehension program is a necessary criterion for participating in the study, schools that refuse to participate in random assignment will be dropped from consideration.

School officials must sometimes be persuaded of the benefits of participating in a study like this one. One potentially relevant benefit has to do with school accountability. Although not listed as an explicit criterion in the legislation authorizing the proposed Title I research, a

school's failure to meet its Adequate Yearly Progress (AYP) targets could be an incentive for the school to adopt changes to its educational practices.

Using information collected during the initial visits to districts and schools, we will develop a semifinal list of the districts and schools that are able and willing to participate in the evaluation. We expect the list will include about 13 to 15 districts. Out of these 13 to 15 districts, we expect to reach final agreements to participate in the study with about 10 districts. If more than 10 districts agree to participate, we will assess the extent to which the districts contribute to the geographic diversity and face validity of the study, and make recommendations to IES on the 10 districts we recommend including in the study.

III. EXPERIMENTAL DESIGN AND ANALYSIS

The Evaluation of Reading Comprehension Interventions will use an experimental design in which schools are randomly assigned to selected reading comprehension interventions or to a control group. As part of the earlier reading comprehension intervention study design task that IES awarded to MPR, we made several recommendations concerning the design of the experiment to assess the impacts of up to four reading interventions. In this chapter, we expand on those recommendations by describing our general approach to meeting the major requirements of the evaluation, including indicating the research questions the study will address and how we will implement the experimental design and collect and analyze the data.

A. RESEARCH QUESTIONS

Table III.1 lays out the main questions this study will address. The upper portion of the table shows the questions about the impact of reading comprehension interventions on students' reading comprehension (both in general and when using expository text, such as that encountered in science and social studies). The lower portion of the table shows questions about conditions and practices (implementation and context) and their association with variation in impacts. The many questions related to the correlational analyses (associations) should not divert attention from the proposed evaluation's most important goal—to document the impact of up to four interventions on students' reading comprehension.

We have also added a research question regarding the effect of the average intervention to the questions requested by IES in the original Statement of Work (SOW). We added this

TABLE III.1

ANALYSIS QUESTIONS, DATA ITEMS AND SOURCES, AND ANALYSIS METHODS

Analysis Questions	Data Items	Data Source	Analysis Method
Can reading comprehension interventions improve student reading comprehension? What are the most effective reading comprehension interventions for improving student reading comprehension?			
1. What is the effect of each reading comprehension intervention on reading comprehension (both in general and when using expository text)?	Test scores in general reading comprehension and comprehension of expository text (such as the text students may encounter in science and social studies)	Student tests (for scores) and student records (for background data)	Simple differences of means and hierarchical linear models to adjust for clustering and to increase sample precision.
2. What is the average effect of supplemental reading comprehension instruction on reading comprehension (both in general and when using expository text)?			
3. How are student characteristics related to effects?			
Which conditions and practices are related to effects?			
1. What school characteristics are related to effects?	Location, urbanicity, proportion of low-income students, proportion of LEP students, average reading achievement score	CCD and school records and student tests	Simple correlations of conditions and practices with alternative measures of effectiveness (e.g., gains, average achievement). Hierarchical linear models of relationship between outcomes for the intervention groups and measures of conditions and practices.
2. How is teacher training related to the effects?	Adequacy of teacher training	Observation of teaching training and student tests	
3. Which teacher characteristics are related to effects? Is a school's professional culture related to effects?	Teacher's educational background, qualifications, experience, prior experience teaching reading, measure of school's professional culture	Teacher questionnaire and student tests	
4. Is fidelity of implementation associated with variation in impacts?	Intervention-specific fidelity measures	Classroom observations and student tests	
5. Are duration and frequency of reading comprehension instruction associated with variation in impacts?	Exposure to intervention	Classroom observations, student tests, and student records	
6. Is quality of implementation associated with variation in impacts?	Quality measures	Classroom observations and student tests	

question because educators may want to know whether supplemental reading comprehension interventions, on average, can improve students' reading comprehension.[1]

B. THE EXPERIMENTAL DESIGN

The experimental design of this study reflects (1) careful consideration of the questions posed by IES, the Title I IRP, and the TWG that worked with MPR on the design task; (2) our knowledge that all the interventions would be implemented in classrooms; (3) IES's suggestion that up to four interventions should be considered; (4) our recommendation that the study focus on the fifth grade; and (5) our concerns about contamination. To address these issues, we will implement a cluster randomized experiment consisting of about 100 schools divided evenly among four intervention groups and a control group.[2]

We plan to include multiple school districts in the evaluation to reduce burden and achieve geographic diversity. Though this study is not designed to be nationally representative, including a geographically diverse sample of school districts provides some assurance that impacts measured by this study are not driven by idiosyncrasies in a few districts. We propose a design with 10 school districts, each contributing about 10 schools to the study. Next, we discuss the unit of random assignment, interpretation of the counterfactual, implementation of random assignment, and sample size and statistical power.

[1] An alternative approach would be to conduct a global test of the joint significance of the four interventions. This test would be similar to testing the effect of the average intervention, because it would tell us if at least one of the interventions is significant without being able to identify which intervention is significant.

[2] Glazerman and Myers (2004) had proposed that the control group include 36 schools and that each intervention group include 16 schools. However, that proposal was based on the assumption of a large number of schools in a small number of districts. Because we have decided to target a larger number of districts (with fewer schools per district) to increase the external validity of the study, it is not possible to divide schools into groups using this ratio of treatment to control schools. For example, with 10 schools per district, we would need to assign 3.6 schools per district to the control group and 1.6 schools per district to each treatment group, which is clearly not possible.

1. Unit of Random Assignment

We will randomly assign *schools*, instead of classrooms or students, to treatment and control groups, for three reasons. First, randomly assigning interventions within schools would require that class formation occur before assignment of teachers to interventions. This could be impractical because training will occur before class rosters are finalized. Second, implementing more than one intervention in the same school would impose the unusual constraint on teachers that they cannot collaborate. This would create an unrealistic context for the implementation of the interventions, and it could also dissuade schools from participating in the study. Third, random assignment of classrooms within schools could lead to contamination.[3] If classrooms or students in a single school were assigned to treatment and control groups, students in the control group could benefit from one of the treatments. For example, control group students might accidentally be included in treatment group activities, might talk with treatment group students about what they are learning in class, or could be reassigned to treatment group classrooms in an attempt to provide them with the intervention. Control group teachers might talk with treatment group teachers about the methods they are using to teach reading comprehension. By conducting random assignment at the school level, we greatly reduce these risks.

2. The Counterfactual

When interpreting the effect of the reading interventions, it is useful to consider the question, "Compared to what?" That is, what would students in each treatment group have experienced if they did not have access to their assigned reading intervention? Due to random assignment, the treatment groups and control group are, on average, statistically similar before

[3] In consultation with IES, we will explore the feasibility of conducting a small ethnographic study of the extent to which contamination might occur if random assignment were conducted within schools. Subsection 5 describes this small study.

the intervention starts. Thus, we can infer what students in each treatment group would have experienced, on average, in the absence of their assigned intervention by examining the experiences of students in the control group. To accurately understand the counterfactual, we will use classroom observations to assess what students in the control group experience. Section C provides details on the classroom observations we plan to conduct.

3. Implementing Random Assignment of Schools

In principle, the random assignment of schools to interventions and a control group is straightforward. Figure III.1 illustrates how the experimental design, which includes random assignment of schools within districts to interventions, will be set up. We will use a randomized block design, which is analogous to stratification techniques used to make statistical sampling more efficient. The primary blocking factor will be the school district. That is, we will conduct random assignment of schools within districts to hold district policies such as teacher hiring, compensation, and professional development constant. If feasible, we will also form two blocks of five schools within each district based on school characteristics (schools will be matched on reading test scores from the previous year).[4,5] We will then randomly assign schools within those blocks to increase balance among intervention groups.

[4] When more schools are eligible to participate than can be included in the study from a given district, we will work with district administrators (who are most knowledgeable about the schools) to identify schools to invite to participate in the study.

[5] We will also examine other school characteristics such as the race and ethnicity of students in the schools, the percentage of students in the schools that are LEP, and whether schools are making AYP to determine whether those may be important blocking factors. For example, if all schools are roughly similar in terms of race and ethnicity, test scores, and percentage of LEP students, but one group of schools is making AYP and another group is not, we could form two blocks of five schools (making AYP and not making AYP) and conduct random assignment within those two blocks.

Random assignment will ensure that, when the data are pooled across districts, the only expected difference between the groups, on average, is that they will have been exposed to different interventions or to no intervention. Figure III.1 also shows that we expect to include all fifth-grade students in each of the 100 schools in the study (we are assuming three fifth-grade classes with 26 students each in each school).

FIGURE III.1

RANDOM ASSIGNMENT OF SCHOOLS WITHIN DISTRICTS

	Intervention Group				
	A	B	C	D	Control
Within-district total of:					
Schools	2	2	2	2	2
Classrooms	6	6	6	6	6
Students	156	156	156	156	156
Studywide total of:					
Districts	10	10	10	10	10
Schools	20	20	20	20	20
Classrooms	60	60	60	60	60
Students	1,560	1,560	1,560	1,560	1,560

Protecting the Integrity of the Random Assignment Design. To protect the integrity of the random assignment design, evaluators must always guard against threats to internal validity. A key threat to the internal validity of the reading comprehension evaluation is "compensatory equalization." This may occur if, for example, districts give additional resources for reading instruction to schools in the control group because they were not selected to participate in one of the four interventions, and those schools use these resources to modify reading, science, or social studies curricula or instructional practices. These modifications to the curricula in the control schools could boost students' reading comprehension to levels above what would normally be expected in the absence of the interventions operating in other schools in the district. As a result, the evaluation could underestimate the impact of the reading interventions.

To help protect against the possibility that schools assigned to the control group will make compensatory adjustments that would threaten the integrity of the random assignment design, we will (1) carefully document the prevailing reading comprehension instruction and curricula for fifth graders before we announce the schools' assignment status, and (2) monitor curriculum instruction to ensure that it does not substantially deviate from the prevailing curriculum. Specifically, as part of the implementation analysis, we will cross-check the data from our observations of experimental *and* control classrooms with our knowledge of what the curriculum would have been without the interventions. Limited changes in curricula that would have occurred in the absence of the study are permissible. If we observe more substantial changes in curricula, we will contact the school district to explain the importance of maintaining the integrity of the control group. We will also work with district and school administrators (both before the evaluation begins and again when the interventions start) to ensure they understand the consequences of making these types of adjustments.

4. Sample Size and Statistical Power

Due to the study's size and use of rigorous methods, educators and policymakers may draw heavily on the findings of this study when choosing whether to adopt one of the interventions under consideration. Because of the policy relevance of the study, it is particularly important to make clear distinctions between effects that are likely to be real and those that are likely due to chance.

When estimating sample size requirements for the design, we factored in four considerations: (1) the need to account for multiple comparisons when conducting tests of statistical significance, (2) whether we can estimate district fixed effects in the impact analysis, (3) the benefits of a baseline test to increase precision, and (4) IES's desire to detect differences between each intervention group and the control group equal to an effect size of 0.25 or larger

with a high probability. Furthermore, we worked from key assumptions concerning the expected number of fifth-grade classes within schools, the number of students within classes, and the expected correlation between pre- and post-test measures of reading achievement. These statistical considerations and assumptions suggest an experimental design that includes 20 schools in each of the intervention groups (4 interventions x 20 schools = 80 total intervention schools) and 20 schools in the control group.

a. Multiple Comparisons

In this study, making clear distinctions between effects that are real and those that are due to chance is complicated by the issue of multiple comparisons. By comparing multiple intervention groups to a control group, and multiple treatments with each other, the probability that one of those differences will appear to be statistically significant is greater than the probability that any single difference will appear statistically significant. Intuitively, this is similar to the difference between the probability of a *single* toss of a coin yielding heads and the probability that *at least one of several* coin tosses will yield heads.

The literature suggests a variety of techniques for addressing the issue of multiple comparisons. However, no clear consensus exists on how to address the issue of multiple comparisons. To calculate the MDE of the study, we will adjust for multiple comparisons using methods that strongly control the family-wise error rate (FWE). The FWE is the probability of falsely declaring at least one of the effects under consideration to be statistically significant. Specifically, we will use the Dunnett critical values for MDE calculations involving comparisons of the intervention groups to a control group, and the Tukey Honest Significant Difference (HSD) critical values for calculations involving all pair-wise comparisons among the

intervention groups.[6] Appendix B describes these and other procedures for adjusting for multiple comparisons.

b. District Fixed Effects

Districts included in this study will not be chosen to be representative of any larger population of school districts. If the study were repeated, the same districts would be included, meaning that it may be appropriate to treat district effects as fixed rather than random.[7] By treating district effects as fixed, the variance of the impact estimate may be substantially lower than if district effects were treated as random, which will allow us to detect smaller impacts.

To calculate district fixed effects, we will need at least two schools per treatment condition in each district. If possible, we will recruit 10 schools from each district, yielding 2 schools in each of the 4 intervention groups and 2 schools in the control group. If we can attain this sample target, we will be able to treat district effects as fixed in our impact analysis.

However, if we cannot recruit 10 or more schools in every district, we will have insufficient sample to calculate district fixed effects. If this occurs, we will attempt to approximate district fixed effects by pooling schools across similar districts. Under this scenario, we could form blocks consisting of matched pairs of districts with fewer than 10 schools. Districts would be matched based on geographical considerations and average baseline test scores. Each district with 10 or more schools would form its own block. In the impact analysis, we would calculate fixed effects for each block and random effects for the schools in each block. In this case, the

[6] Adjusting for multiple comparisons using the Dunnett procedure increases the MDE by approximately 0.03 standard deviations relative to no adjustment for multiple comparisons, and adjusting using the Tukey procedure increases the MDE by approximately 0.04 standard deviations relative to no adjustment for multiple comparisons.

[7] When treated as fixed, the effects of districts are estimated using district-indicator variables. The explanatory power of these variables reduces the amount of unexplained variation, thereby increasing the precision of impact estimates.

block fixed effects would not explain as much of the total variation as district fixed effects would have, because schools from two different districts will be included in some of the blocks.

c. Using a Baseline Test to Increase Precision

A large proportion of the variance in the test score that will be administered at the end of the study can be explained by students' prior achievement. By estimating impacts with a regression model that uses prior achievement as one of the covariates, we can substantially increase the precision of the impact estimates, allowing us to detect smaller effects. However, the first test administered to students will not be a true baseline test - it will be administered shortly after the intervention has begun. Consequently, it might not be appropriate to include this early test in the analysis (see section D), which will adversely affect the study's MDEs.

An alternative approach to increasing the precision of the impact estimates is to use prior achievement at the school level as a covariate in the analytic model. Bloom et al. (1999 and 2005) demonstrate that using prior achievement data aggregated to the school level can be almost as effective as individual student data in reducing the estimated MDEs for a study.[8] Specifically, they found only a 0.01 difference in the MDE between using student-level test score data and school-level data. The MDE calculations presented below assume that student level data will be used. If we determine that it is only appropriate to use school-level data on achievement (such as the test score data gathered from school records) due to the timing of the baseline test, we expect the MDEs will be slightly higher than those reported here.

[8] Bloom et al. (2005) used data from Atlanta, Georgia; Columbus, Ohio; Houston, Texas; Newark, New Jersey; and Rochester, New York, on students in grades 3, 8, and 10. Bloom et al. (1999) used data from Rochester, New York, on students in grades 3 and 6.

d. Detecting the Desired Effect Size

To assess appropriate sample sizes for the evaluation, we adopted the precision standard specified in the statement of work of an MDE of 0.25 standard deviations for the impact that is the primary focus of this study—the difference in student outcomes between each intervention group and the control group. Table III.2 presents the MDE for three types of comparisons:

1. ***Effect of Each Intervention.*** In the first column, we calculate the MDE when comparing *each individual* intervention group to the control group. We use the Dunnett procedure to adjust for comparing multiple intervention groups to a control group. This comparison answers the question that is the primary focus of this study—"Which interventions are effective?"
2. ***Average Effect of All Interventions.*** In the second column, we calculate the MDE when comparing the *average* treatment group to the control group. This is a single comparison and therefore is not adjusted for multiple comparisons.[9] This comparison answers the question, "What is the average effect of the supplemental reading interventions offered in this study on reading comprehension?"
3. ***Differences in the Effects of Interventions.*** In the third column, we calculate the MDE when comparing *differences* between treatment groups. We use the Tukey HSD procedure to adjust for all pair-wise comparisons among intervention groups. This comparison answers the question, "Which intervention is most effective?"

If we achieve the target sample size of 100 schools distributed evenly among 10 districts, we will be able to detect, with a high probability, an effect size of 0.25 or larger for each individual intervention. The MDE corresponding to the average effect of all interventions will be 0.17, and the MDE corresponding to all pair-wise comparisons among intervention groups will be 0.26 (see panel A of Table III.2). Analyses of district subgroups (for example, districts with a high or low percentage of LEP students) will only be able to detect effects larger than the target of 0.25,

[9] As noted earlier, although the original SOW did not specifically request this comparison, we propose that the comparison be made, because educators may be interested in whether supplemental reading comprehension instruction, in general, can improve students' reading comprehension.

except in the case of the effect of the average intervention, where we are likely to detect an effect size of 0.24.[10]

If 100 schools can only be included by recruiting 5 schools from each of 20 districts, it will not be possible to reduce variance by controlling for district fixed effects. If this occurs, we will approximate district fixed effects by pooling schools across similar districts. Consequently, the MDE associated with comparing each individual intervention group to the control group is 0.31. The MDE corresponding to the average effect of all interventions will be 0.21, and the MDE corresponding to all pair-wise comparisons among intervention groups will be 0.32 (see panel B of Table III.2).

TABLE III.2

MDEs ON STUDENT TEST SCORES

	MDE, by Type of Comparison		
Sample Size	Effect of Each Intervention	Average Effect of All Interventions	Differences in the Effects of Interventions
A. 100 schools, 10 school districts, [a] divided evenly among 4 intervention groups and a control group			
Full sample	0.25	0.17	0.26
50 percent subgroup of districts	0.36	0.24	0.38
B. 100 schools, 20 school districts, [b] divided evenly among 4 intervention groups and a control group			
Full sample	0.31	0.21	0.32
50 percent subgroup of districts	0.45	0.30	0.47

[a]With two schools per treatment condition per district, district fixed effects can be calculated.
[b]With only one school per treatment condition per district, fixed effects are calculated for each of 10 blocks of two districts.

Note: These calculations take into account clustering effects at the school level, as well as adjustments for multiple comparisons. A Dunnett adjustment is made for column 1, while a Tukey HSD adjustment is made for column 3. Because school districts are purposefully selected, we treat their effects as fixed. Because all fifth-grade classrooms at each school are included in the study, classroom clustering does not affect the variance of the estimated impact. The equation used to calculate the MDE between two curriculum groups is:

$$Factor(\alpha, \beta, df) * \sqrt{1-R^2} * \sqrt{\frac{2\rho}{S} + \frac{2(1-\rho)}{N}} \ ,$$

where:

S is the number of schools in each group, N is the number of students, α=0.05, β = 0.20 (that is, 80 percent power), ρ (=0.10 in panel A, = 0.16 in panel B) is the between-school variance as a percentage of the total variance of the outcomes based on previous studies, and df is the degrees of freedom, which is equal to the number of schools minus the number of strata minus the number of intervention and control groups. Previous impact evaluations have found that an R^2 of about 50 percent may be an appropriate assumption when baseline measures of test scores are available. Parameter assumptions are based on authors' calculations using data from the LESCP and the analysis of past studies in Schochet (2005).

[10] Because the sample of students is large, analyses of 50 percent student subgroups will be able to detect effects almost as small as the main impacts.

5. An Ethnographic Study of Classroom Random Assignment

The primary motivation for randomly assigning schools, rather than classrooms within schools, to intervention groups is the possibility of contamination of the control group. However, the extent to which contamination might actually occur has not been tested. If contamination does not occur, the statistical precision of random assignment studies could be increased considerably by assigning classrooms, rather than schools, to intervention and control groups. In the present study, randomly assigning classrooms rather than schools would reduce the MDE, when comparing multiple intervention groups to a common control group, from 0.25 to 0.20. Alternatively, we could maintain an MDE of 0.25 but reduce the size of the study from 100 schools to 70 schools.

Because of the potential benefits of classroom random assignment, we propose to conduct, in consultation with IES, a limited study of contamination when classrooms within schools are randomly assigned to intervention and control groups. The study would involve a small number of schools in addition to the 100 included in the main study. We recommend identifying the schools by identifying an additional school in 8 of the 10 participating districts in the study. This process will result in eight schools in the ethnographic study, two schools in each treatment group. Fifth-grade classrooms in these schools would be randomly assigned to the intervention and control groups.[11]

To examine contamination, we could use teacher interviews, student focus groups, and limited student testing. We would ask teachers and students about specific details of the intervention being offered in their school, to determine whether control group students have been exposed to that intervention. It is important to talk with both treatment and control groups to

[11] It will not be necessary to field all four interventions in a single school since the purpose of this smaller study is to study contamination, not to calculate impacts.

allow us to better understand whether contamination occurred. For example, the interpretation of a finding of no control group students having knowledge of the intervention strategies would be affected if the study were to also find that no treatment group students report knowledge of the strategies. In addition to determining whether the students have been exposed to the programs, we will try to understand how the exposure occurred. For example, separate focus groups of control group and treatment group students would:

- Ask about the extent to which students talk with students in other classes about what they are learning in class.
- Ask about the extent to which students work with students from other classes on projects or homework.
- Assess whether students have seen the materials associated with the intervention being implemented in their school. If students have seen the materials, focus group facilitators will ask students how they came in contact with the materials (for example, through their teacher or other students).
- Assess whether the students have heard of the techniques being taught as part of the intervention. If students have heard of the specific techniques being taught, focus group facilitators will ask how students heard of the techniques (for example, through their teacher or other students).

Interviews of treatment group and control group teachers will assess:

- The extent to which they generally collaborate with their colleagues on the teaching of reading.
- The extent to which they have talked with colleagues about the specific intervention being tested in that school, including which aspects of the intervention were discussed.
- The extent to which they have looked at or used materials from the intervention, including which materials were used.
- Whether they believe restrictions on collaboration for the sake of a study would be acceptable to them.

Students in both intervention and control classrooms could also be administered a brief test in spring 2007 (for example, a worksheet asking questions about the specific strategies being

taught as part of the intervention). If students in the control group do as well on the test as students in the treatment group, contamination would be suspected. At the request of IES, we can produce a more complete design for this smaller study, including survey instruments, interview protocols, and cost estimates.

C. DATA COLLECTION

To address the central research questions, we will obtain several types of information from a variety of sources during the 2006–2007 school year.[12] Data collection activities will include (1) administering reading tests to students; (2) extracting school records data; (3) observing classrooms; and (4) obtaining information from principals, teachers, and curricula developers. These sources will provide information on reading comprehension instruction and curriculum, the context in which the instruction occurs, the implementation of the interventions, teacher characteristics, student characteristics, and student outcomes. Table III.3 presents a general overview of the study's data collection plan. It shows some of the major variables or concepts to be addressed under each domain shown in the conceptual framework in Chapter I. The rest of this section describes each of the data collection activities.

1. Reading Tests

The evaluation's key outcomes are reading comprehension test scores in general reading and in expository text, such as the text students encounter in social studies and science. We will administer tests at the beginning and end of the 2006–2007 school year. Fall test administration will begin about three weeks after the start of the year (allowing approximately one week for enrollment to stabilize and one week for the dissemination of passive consent forms and parents'

[12] After considering the preliminary impacts detected during the initial year of data collection and the percentage of students continuing in the same school or district in the 2007–2008 academic year, we may collect data for a second year.

TABLE III.3

OVERVIEW OF DATA COLLECTION PLAN

	Reading Tests	Student Records	Classroom Observations	Review of Materials	School Information Form	Teacher Survey	Developer Survey
Reading Comprehension Instruction and Curriculum							
Instructional components			X	X			X
Types of material used			X	X			X
Fidelity			X	X			X
Quality			X	X			X
Frequency and duration of use			X	X			X
Context							
Teacher demographics, background characteristics						X	
Teacher experience and training before study period						X	
Professional development for selected interventions			X			X	
Presence of other staff (paraprofessional, literacy coach, reading specialist) to assist teacher or students			X			X	
School-level student characteristics					X		
Professional culture at the school						X	
Adequate Yearly Progress (AYP) status					X		
Types of materials used for reading and social studies				X			
Implementation							
Fidelity			X				
Quality and focus of student reading materials			X	X			
Cost				X			X
Teachers' knowledge and use of reading comprehension instructional practices			X			X	
Student Characteristics							
Demographics		X					
English-language ability (e.g., LEP status)		X					
Use of learning support services		X					
Eligibility for free or reduced-price lunch		X					
Reading fluency	X						
Attendance		X					
Outcomes							
Reading comprehension (in general and when using expository text)	X						

reply).[13] The fact that this fall test is administered *after* the intervention has begun has implications for the design and analysis (discussed in Section D). The spring test administration will begin about six weeks before the end of the school year (dates may vary slightly, depending on districts' schedules for administering their own tests).

For general reading comprehension, we will administer the Passage Comprehension subtest of the Group Reading Assessment and Diagnostic Evaluation (GRADE). The GRADE is a multiple-choice, paper-and-pencil, group-administered test that can be used to measure baseline skills and student improvement in critical reading areas. The Passage Comprehension subtest measures the student's ability to comprehend an extended text as a whole. We will administer the GRADE to students at the beginning and end of the 2006–2007 school year.

To assess students' reading comprehension of expository text in science and social studies, we will use a test that either the Northwest Evaluation Association or the Educational Testing Service will develop from its item bank and from other newly created items. Half of the students will take the test in science, and half will take it in social studies. For this assessment, we are considering a computerized adaptive test (a test that adapts the questions that individual test-takers will receive based on whether they answered previous questions correctly), which requires fewer questions to determine a test-taker's score. Students will take this test near the end of the 2006–2007 school year, generally within the same week (but not on the same day) in which they take the GRADE test.

We will also administer a brief fluency test—PRO-ED's Test of Silent Contextual Reading Fluency (TOSCRF). This group-administered test is used to assess students' silent reading fluency—it measures skills such as word identification, word meaning, and sentence structure,

[13] If a district requires active consent, fall test administration will be delayed somewhat.

all of which are important skills for reading comprehension. It will provide more refined data for the subgroup analysis than would sorting students on the baseline reading comprehension test alone. We will administer this test along with the baseline administration of the GRADE test.

2. Student Records

The student records form will contribute useful background information on student characteristics and reading support services received (a draft is provided in Appendix C). We will use several methods to collect school records data. Some schools or districts may be able to extract the information electronically from their databases. At other schools, administrative staff will complete the records forms manually. In some cases, where confidentiality policies do not preclude access to student records by nonschool personnel, we will train local field staff to collect the student records information.

3. Classroom Observations

Our main approach for measuring implementation of the reading comprehension interventions in the treatment group schools and students' reading instruction in control group schools is to observe classroom instruction. To provide context for interpreting the observations, we will also collect data on the nature of the professional development activities for each intervention and will ask reading experts to review each intervention's curriculum for adherence to research-supported principles. Specifically, we will examine:

- Variations in fidelity of implementation in treatment classrooms (across interventions and across schools and/or districts)
- Variations in quality and content of reading comprehension instruction (across the four interventions and the control conditions)

We will conduct two rounds of classroom observations—the first from November through mid-December, the second from mid-January through February. We selected the fall dates to

allow teachers to become familiar with the program before being observed. We selected the spring dates to avoid conducting observations during periods when schools typically administer standardized tests. During each round, we will observe all classrooms once and a randomly selected half of the classrooms a second time the next day. This approach—using clustered observations in the fall and spring, with repeated measurements at each point for half of the classrooms—will allow us to evaluate the reliability of estimates from classroom observations and assess temporal stability. These factors are particularly critical for the Likert scales being used for this study, as they require observers to make judgments about the quality of teacher behaviors. Because observer judgment is involved in these scales, they require higher levels of inference than scales that involve simply quantifying occurrences of relevant teacher behaviors, such as measures developed by Foorman and Scatschneider (2003) (see, for example, Gersten et al. 2005).

We will observe reading and either science or social studies instruction in the classroom. Observations will last about 120 minutes. Staff will schedule observations of intervention classrooms on a day when teachers are conducting a science or history lesson using a textbook (as opposed to conducting assessments, science experiments, American history projects, etc.). This will let us observe treatment teachers using their assigned intervention and control teachers providing their regular text instruction. In both intervention and control classrooms, observers will complete a quality-of-instruction form that collects data on the use of research-based reading comprehension and vocabulary development techniques (see below for more information on this form). A new form module will be completed for each 15-minute interval of the observation (forms will be completed during the time treatment teachers are using the interventions, as well as during other aspects of their lessons). In treatment classrooms only, observers will also complete a second form that collects information on the fidelity of the implementation of the

intervention—that is, the extent to which teachers implement the curriculum approach as intended by developers (see below for more information on the fidelity forms).[14]

The instruments to be used were adapted from other instruments that were based on current reading research (for example, Edmonds and Briggs, 2003). Extensive testing is being done to ensure that the measures are reliable and sensitive to the quality and content of reading comprehension. When we train staff to complete the observation forms, we will assess inter-rater reliability to ensure that forms will be completed consistently across observers (preliminary testing found 0.90 inter-rater reliability). For staff who are unable to provide ratings consistent with other observers' ratings, we will either provide additional training or train new staff. Next, we discuss the two forms staff will complete during classroom observations.

Fidelity of Implementation. We will collect data on the fidelity of implementation in each intervention classroom, because a description of the fidelity of implementation is essential for understanding the relationship between an intervention and its impacts. Fidelity measures do not purport to assess the *quality* of instruction, but they do suggest why impacts may vary across districts and indicate if an intervention's components are too difficult or confusing to be implemented widely.

In developing these fidelity measures (current drafts are provided in Appendix C), we solicited extensive input from the developers to ensure that the measures reflect the critical elements of each intervention. To the extent possible, we are designing these measures to have a parallel structure across approaches—that is, they are similar in the number of items, the range of the Likert-type scale, and how observations of teachers and students are recorded. The measures

[14] Fidelity of implementation (also known as treatment fidelity or treatment integrity) refers to the extent to which an intervention is implemented as intended across the entire duration of the evaluation (Gersten et al. 2000; Gersten et al., in press; Gresham et al. 2000).

address such questions as: Was the implementation carried out for the required amount of time? Were the required materials used? Were all the key features of the intervention (for example, teacher modeling, teacher/student think-alouds, peer groupings, peer discussions) implemented? For each statement, the observer will indicate either the presence or absence of an element or the extent to which the teacher implemented that aspect of the intervention (or the extent to which students were actively engaged in a task) on a Likert-type scale (for example, little, some, most; few—less than 25 percent, many—25 percent to 75 percent, most—more than 75 percent).

Quality Measure. We will collect data on the quality of instruction in all classrooms (treatment and control). The quality measure documents the quality of instruction based on current research on effective practice in reading comprehension. The measure (a current draft is provided in Appendix C) collects data on instruction procedures and strategies that are associated with, or are hypothesized to be linked to, enhanced outcomes in reading comprehension (Baker et al., in press; National Reading Panel 2000; Snow 2002; Gersten et al. 2001). Items address (1) instructional quality, (2) frequency of instructional practices associated with improved reading comprehension outcomes, (3) student-generated questions, and (4) natural use of feedback and/or quality of prompting provided to support student learning. Most items are measured by frequency (counting occurrences), and several items are measured by Likert scales (such as measures of teachers' responsiveness to students, which are rated as poor, fair, good, or excellent).

4. School Information Form, Teacher Survey, and Developer Survey

School Information Form. We will ask principals to complete a brief form on school-level characteristics. This form will provide background information on each participating school (see Appendix C). Items include school-level test scores for fifth graders for the most recent year, fifth-grade reading curricula (core and supplemental), and student characteristics (such as the

percentage of students with LEP and the percentage eligible for free or reduced-price lunch). We will mail this short, one-page preliminary information form to principals in spring 2006; data from this form will be used to plan the random assignment process and in some analyses. We will ask schools to complete a more comprehensive, two-page school information form as part of the student records collection process in spring 2007. In both cases, we will conduct telephone followup for nonresponse.

Teacher Survey. The teacher survey will provide classroom- and school-level context for interpreting the reading assessment results (see Appendix C for the draft survey). The survey includes items from teacher professional culture scales (The Consortium on Chicago School Research 1999; Carlisle 2003), which can rapidly portray conditions affecting quality of instruction; an abbreviated version of a teacher efficacy scale (Hoy and Woolfolk 1993; Gibson and Dembo 1984), which correlates with the ability to benefit from professional development (Sparks 1988); and items on the teacher's educational and professional background. If fifth-grade students have more than one teacher during the school day, we will ask all teachers providing instruction using the intervention—such as teachers of English/language arts, science, or social studies—to complete a survey. We will mail the teacher surveys in October and conduct telephone followup for nonresponse.

Developer Survey. We will collect information from developers on the costs of their programs for schools using an ingredients approach (Levin 1983). In this approach, we will identify all the items schools would need to purchase to implement and obtain support for the interventions. We will ask developers to specify the unit costs for each of the items, and we will calculate total costs per reading comprehension program based on the quantities needed of each unit. Analyses of these data will provide uniform information on the costs of large-scale implementation of each reading intervention.

D. ANALYSIS

Next, we describe our general approach for conducting analyses of the quantitative data and synthesizing the quantitative and qualitative data. We discuss five topics: (1) estimating impacts, (2) estimating impacts for subgroups, (3) addressing anticipated complications, (4) examining long-term effects and duration of interventions, and (5) synthesizing quantitative and qualitative data.

1. Estimating the Impacts of Reading Comprehension Interventions

Because the evaluation will use an experimental design, the estimator of the effect of reading comprehension interventions is the difference between the average test scores for students in schools using the selected interventions and the average test scores for students in the control schools. For each of the test scores of interest (for example, scores on the GRADE comprehension test, and scores on the tests of comprehension of expository text in social studies and science), we will compute four separate impact estimates: (1) intervention A versus control, (2) intervention B versus control, (3) intervention C versus control, and (4) intervention D versus control.[15] We will also compare the average test scores for students in all the intervention groups to those for students in the control group.[16]

[15] In general, it is assumed that the interventions affect only the average achievement of students within the group. Some interventions may also affect the spread of achievement scores. For example, an intervention may have a larger impact on students at the lower end of the achievement distribution and thereby change the variance of the group and influence the mean. Such a finding can have both substantive and statistical implications. Substantively, it gives us more information about the effectiveness of an intervention. Statistically, it creates issues about how best to conduct tests of statistical significance. If we find differences in variances across interventions and control groups, we will need to use more complex analytic procedures to estimate the correct standard errors of the impact estimates. With large numbers of schools, such as proposed for the evaluation, it is appropriate to use the Huber-White or sandwich estimator for standard errors (Raudenbush and Bryk 2002).

[16] As noted earlier, although the original SOW did not specifically request this comparison, we propose that the comparison be made, because whether supplemental programs in general can affect reading comprehension may be of interest to educators.

In practice, the estimation of impacts will use hierarchical linear modeling (HLM), which allows us to take into account the clustering of students that occurs with random assignment of schools to interventions and the blocking that will be used when randomly assigning schools within districts to interventions (Raudenbush and Bryk 2002). Failure to take into account the clustering of students within schools can lead to overstating the statistical significance of impacts. We will also use techniques that take into account stratification and clustering using Taylor series linearization for simple comparisons of means, or generalized estimating equations (GEE) for regression models, because these approaches can be more stable when sample sizes are small.[17]

In addition to accounting for the clustering and the randomized blocks design used to define the experiment, we will include baseline test scores as a covariate in the statistical models.[18] Baseline test scores are strongly correlated with later test scores, and their inclusion as a covariate in the model will increase the precision with which we can estimate impacts. In fact, the use of baseline scores is critical in the design of the evaluation because it allows us to use a smaller number of schools than would be the case in a design where pretest scores were not available.

The basic HLM is a two-level model. The first level corresponds to the student; the second, to the school. As expressed below:

[17] The GEE method is a generalization of generalized linear models (GLM) that takes into account within-group correlations between observations.

[18] Because the fall 2005 test will be administered *after* the intervention has begun, it may be preferable to use baseline achievement at the school level, which will be collected using the school records form. Bloom et al. (1999 and 2005) show that school-level data is nearly as effective as student-level data in reducing the MDE of a study when random assignment occurs at the school level. In Section 3, we discuss the trade-offs involved in using baseline achievement data from the school records form versus data from the fall test administration.

Level One: Students (i) within schools (j):

$$y_{ij} = \beta_{0j} + \beta_{1j} y_{0ij} + r_{ij}$$

Level Two: Schools (j):

$$\beta_{0j} = \gamma_{00} A_j + \gamma_{01} B_j + \gamma_{02} C_j + \gamma_{03} D_j + \gamma_{04} CONT_j + \sum_{m=1}^{10} \lambda_{0m} DIST_{jm} + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \sum_{m=1}^{10} \lambda_{1m} DIST_{km} + \mu_{1j}$$

The Level One model relates students' post-intervention test scores to their fall test scores, a constant term, and a residual term (unexplained variation).[19] The Level Two model relates the school-level average test score (β_{0j}) to the four treatment indicators (A, B, C, and D), the indicator for the control group for schools (*CONT*), fixed block effects (*DIST*), and a residual term that accounts for random variation in the school means.[20] The Level Two model also allows the effect of baseline test scores (β_{1j}) to vary by district (λ_{1m}) and by a deviation (μ_{1k}) from the overall effect (γ_{01}).

The parameters of interest to the evaluation are the ones that pertain to intervention impacts. The impacts are defined as the difference between the average achievement score for the intervention group and the control group. For example, the impact of reading comprehension intervention A on reading achievement is defined as the average achievement for the reading test for intervention A minus the average achievement for the reading test for the students in the

[19] For the proposed analysis in the two-level HLM framework, we would "mean-center" preintervention achievement around the grand mean.

[20] We have used a no-intercept model specification for Level Two so that the parameters on intervention and control group indicators can be directly interpreted as the average outcome in each group.

control group. Using the HLM, we see that the impact is estimated from the following expression:

$$\delta_A = \gamma_{00} - \gamma_{04}.$$

Similar expressions can be constructed to estimate the impacts of the other interventions. Differences in the effects of interventions can also be computed. For example, the difference in effects between intervention A and B is:

$$\delta_{A-B} = \gamma_{00} - \gamma_{01}.$$

The HLM estimates the model parameters and their associated standard errors by using maximum likelihood and related methods. Statistical significance will be calculated using those standard errors and by taking into account the issue of multiple comparisons. The primary multiple comparison adjustment will be a stepwise technique that controls the FWE while providing more statistical power than the Dunnett or Tukey methods used to calculate the study's MDE. See Appendix B for a complete discussion of the issue of multiple comparisons. In tables, we will denote impacts that are different from zero at a 5 percent level of statistical significance using a single asterisk.

2. Student Subgroup Analyses

Subgroup analyses will be important for revealing who benefits most from specific reading interventions. When developing the data collection instruments, we identified groups of children that may be of particular interest to policymakers or educators to inform future decisions about the implementation of the reading interventions. Some of these groups may center on students' reading abilities (for example, students who have weaker initial reading comprehension skills and lower levels of fluency, data that the study will be collecting at baseline). Other groups may

center on students' background characteristics (for example, LEP students, students from low-income families). The subgroup analyses may show, for example, whether students with weak reading foundations—those with limited fluency and weak reading comprehension skills at baseline—benefit as much as students who enter the interventions with stronger foundations.

One limitation of this specific subgroup analysis is that the administration of the fall test used to determine students' existing abilities will occur *after* the intervention has already begun. If the intervention has large, immediate impacts on students' *rank* in the distribution of reading comprehension ability (large enough to change the ability group in which the student is classified), this analysis might not be valid. However, the impact on students' rank would have to be considerable to invalidate subgroups based on the early test score, particularly if the subgroups are large (for example, defined by test score tertiles).

3. Complications to Be Anticipated in Impact Analyses

The analysis of the impact data is likely to give rise to three complications. First, students might transfer from one intervention school to another, or between an intervention school and a control school. Second, students may drop out of the evaluation and generate missing data.[21] Third, baseline test scores are likely to be collected shortly *after* the interventions have begun, rather than before.

Student Mobility. Some students attending schools in the control group may move to one of the intervention schools, or students in an intervention school could transfer to a control school. The statistical literature refers to such a problem as noncompliance with the

[21] Students may also change classrooms within a school. Since all classrooms in the same grade will be in the study, this type of mobility will not complicate most analyses. An exception is the analysis of the relationship between implementation fidelity and outcomes, because it will not be clear in which classroom to include students. We will address this issue by investigating the sensitivity of findings to whether students are included in their original classroom or the classroom to which they transfer.

experiment's protocols (see, for example, Angrist et al. 1996). The effects of students' actual intervention status will be estimated using the instrumental variables (IV) method, where initial assignment to intervention groups will be used as an instrument for the group that students actually participate in.

The IV method is a two-step procedure. In the first step, students' final intervention status is regressed on their initial intervention assignment, and the coefficient estimates from this regression are used to calculate students' *predicted* intervention status. In the second step, the outcome is regressed on the *predicted* intervention status calculated in the first step. The coefficient on predicted intervention status in the second regression provides the estimate of the program's impact on the outcome. This approach is analogous to dividing an impact by the participation rate of students assigned to a treatment group to generate the impact on participants.

Missing Data. Although we anticipate high response rates to both surveys and student tests, it is reasonable to expect that some students will be "lost" from the evaluation, particularly if they transfer to a school outside of the study. Fortunately, we can adjust for "nonresponse" by using statistical procedures such as the estimation of nonresponse models. These models predict the probability of remaining in the study and responding to the survey or of being tested on the basis of student characteristics such as earlier test scores, sex, and race and ethnicity. Nonresponse models are used to create statistical weights that allow the remaining students to "stand in" for students who left the sample. These weights are applied while conducting the analyses and will produce unbiased estimates of intervention effects under the assumption that the observed characteristics of students adequately account for the processes that led to students leaving the sample.

Baseline Test Scores. As noted above in Section B on the experimental design, the administration of the GRADE test in fall 2006 will take place *after* the interventions have begun.

Consequently, this will not be a true baseline test, and might already reflect an impact of the interventions on students' scores on the GRADE test. If this early test does reflect an impact of the interventions, including it as a regressor in the impact analysis could understate the impact of the interventions.

Consistent with the options laid out for handling this issue in the study design, two solutions to addressing this complication in the analysis are possible. First, we can include school-level average test scores from previous years (Bloom et al. 1999 and 2005). One drawback to this approach is that the tests administered by the state may not measure the same skills as the GRADE test that we will administer at followup, thereby reducing its ability to explain variation in the GRADE test administered at followup.[22] Another drawback is that we would not be able to form subgroups based on students' prior reading ability, nor can we use the baseline test to assess change in scores on the GRADE.

A second solution to this problem is to include the early test scores in our analysis, despite the problem with the test's timing. Although the early test will be administered after the intervention has begun, the administration will still take place early in the intervention period, and the test might not yet exhibit effects of the intervention. If we find that including the baseline test increases the precision of our impact estimates without attenuating the size of the impact, we will deem it appropriate to include the early test in the analysis. Alternatively, if including the early test reduces the size and significance of our impact estimates, we will conclude that it is not appropriate to include in the analysis. We will consult with IES to determine how we will address this issue.

[22] As described in Section C, we will also administer assessments of students' reading comprehension of expository text in science and social studies in the spring, but not in the fall. Whether the GRADE would be a better predictor of the science and social studies tests than the state tests is unknown. However, since the GRADE test is designed to assess reading comprehension in general, students' scores on the fall GRADE test may be a better predictor of their performance on the spring science and social studies tests than the state tests.

4. Long-Term Effects of Reading Comprehension Interventions and Duration of Interventions

In the SOW for this project, IES stated two possible options for future data collection and analysis. Option 1 would address whether reading comprehension interventions in the fifth grade produce impacts that last beyond the intervention year by conducting additional student testing in spring 2008. As described in the original SOW, the study may examine second-year effects if three conditions are satisfied. The interventions must have been well implemented, they must yield substantially meaningful impacts after one year, and student mobility must be reasonably low. The rationales behind these criteria are straightforward. First, assuming that an intervention showed an impact after the first year but was poorly implemented, its replication would be difficult in other districts because it failed to follow a clearly specified approach. Second, if impacts are insignificant after one year, it is hard to imagine that meaningful impacts would occur in later years without an additional year of intervention. Third, without an adequate sample size or with strong differential attrition between the treatment and control groups, there may be too little statistical power to detect meaningful impacts, or the analysis may produce biased impact estimates.

Option 2 would address issues related to the duration of the reading comprehension interventions. In this option, the study would follow some fifth graders into sixth grade and assess the impact of adding a second year of reading comprehension instruction. This option calls for us to randomly assign fifth graders (rising sixth graders) in the intervention schools to an intervention or control group within those schools. However, this approach runs the risk of contamination, where students in the control classrooms are indirectly influenced by the intervention, either through interaction with peers or through collaboration among teachers. Direct contamination effects are also possible if some teachers provide the intervention to students in the control group who the teachers view as being in particular need of help. The

pressure on teachers to provide students with whatever help they can may be particularly great in schools that are not meeting their AYP requirements.

We will assess the feasibility of these options in consultation with IES over the course of the study. At the request of IES, we can provide a more detailed design for these options.

5. Synthesizing Quantitative and Qualitative Data

To help with the interpretation of impact estimates, we will describe, using data from the classroom observations, the basic characteristics of the four interventions and the reading instruction provided to control group students. We will also analyze how these characteristics correlate with impacts. First, we will use summary tables to describe (1) the school conditions in each intervention group and the control group, (2) the practices of teachers in each intervention group and the control group, and (3) the implementation fidelity in each intervention group. Second, we will relate these conditions and practices to intervention effectiveness using three different analytic approaches. We do not plan to relate implementation fidelity to intervention effectiveness, because there is no corresponding measure of implementation fidelity for the control group.[23] An important limitation of all these analyses is that the conditions, practices, and fidelity measures that we observe might be correlated with many important factors that we do *not* observe. Therefore, these analyses are all potentially biased and will be interpreted cautiously.

[23] It would be possible to calculate the correlation between student test scores and implementation fidelity using only students in the intervention group. However, because implementation fidelity is a choice of teachers, it would not be possible to distinguish the effects of fidelity from the effects of unobserved teacher characteristics. Put another way, we cannot provide separate answers to the questions, "What is the effect of the intervention when it is implemented well?" and "What is the effect of a highly motivated teacher?," because we observe both effects simultaneously.

a. Descriptive Analyses

Descriptive analyses will draw on the data sources described in Section C, and the specific data items will be collected using the forms in Appendix C. Examples of the conditions, practices, and implementation fidelity measures are:

- ***Conditions.*** Average test scores for fifth graders at each school in the previous year, fifth-grade reading curricula, percentage of students eligible for free or reduced-price lunch, race and ethnicity of students in the schools, school size, professional culture in the school, and the percentage of students classified as LEP.
- ***Practices.*** Whether the teacher managed student behavior effectively, gave accurate and clear explanations, asked students to justify or elaborate their responses to questions, and modeled particular reading comprehension strategies before, during, or after reading.
- ***Implementation Fidelity.*** Whether implementation was carried out for the required amount of time, correct materials were used, and key features of the intervention (for example, teacher modeling, teacher/student think-alouds, peer groupings, peer discussions) were used.

We will compare the average conditions and practices in each intervention group and the control group using descriptive tables. The columns of the tables will correspond to each intervention and control group, and the rows will correspond to specific conditions and practices. The cells in each table will report the average condition or practice for each group. Similar tables reporting implementation fidelity will be prepared for each of the four intervention groups. Combined with the main impact findings, these tables will inform the development of hypotheses regarding why some interventions appear more effective than others.

b. Correlational Analyses

We propose three approaches for relating conditions and practices to impacts. The first approach relates the district-level impacts computed as part of the main impact analysis to district-level conditions and practices. The second approach calculates impacts in each district for the classrooms or schools that belong to subgroups defined by conditions or practices. The

third approach compares average outcomes across intervention and control groups for schools or classrooms that belong to subgroups defined by conditions or practices, regardless of school district. These approaches, and the trade-offs of the approaches, are described below.

The first approach draws on the design of the experiment by relating district-level impacts to district-level characteristics. Each district in the study can be regarded as a mini-experiment with its own impact estimate. We can calculate correlations between those impact estimates and district-level characteristics, including student, classroom, and school characteristics aggregated to the district level. For example, we can find the correlation between impacts and the proportion of students in a district who are poor. District-level impacts are the most precise because they can be calculated accounting for district fixed effects and because they involve all schools, classrooms, and students in a district. District-level conditions and practices are the least refined, however, because of the high level of aggregation.

A limitation of this approach is that a multivariate analysis will not be feasible due to the small number of districts, meaning that any correlation between impacts and conditions could be due to the influence of other factors that are correlated with impacts and the condition of interest. Furthermore, variation in conditions and practices aggregated to the district level could be limited and may mask important variation within districts.[24]

A second approach involves calculating impacts for the subgroups of schools in each district that meet the condition of interest. For example, within each district, we could calculate the impact of an intervention for schools with a high proportion of low-income students by calculating the difference in average student test scores for students in high-poverty schools in the intervention group and students in high-poverty schools in the control group. These impacts

[24] Aggregation of data could lead to biased estimates of the true associations (Freedman 1999).

can then be averaged across districts for an overall impact of the intervention in schools with a high concentration of students from low-income families. The same approach would be used to estimate impacts for schools with a lower concentration of low-income students. We would then compare these two impacts to assess whether they are significantly different. This approach is appealing because it ensures that all districts are represented in both the treatment and control groups. However, this approach requires sufficient variation in school characteristics in each intervention group, within each district. In the above example, this would require a high-poverty school in each intervention group and in the control group in each district. This data requirement may be too stringent for analysis of school-level conditions, because there are likely to be only two schools per intervention group per district. Analyses of classroom-level conditions and practices might be feasible, however, because there will be, on average, six classrooms per intervention group in each district.

The third approach allows schools and classrooms in different intervention groups to be compared across districts. For example, there might not be a school with a high proportion of low-income students in every intervention group in every district. Using this method, the average outcome for high poverty schools in the intervention group will be compared to the average outcome for high poverty schools in the control group, regardless of which districts these schools are in. Put another way, the impact for the high poverty subgroup will *not* be computed within district. This approach would be repeated for low-poverty schools, and then the impacts of the two groups would be compared to determine whether they are significantly different. Though this approach allows a more refined analysis of conditions and practices, the fact that impacts are calculated across districts reduces the precision of the impacts, because district effects must be treated as random. Furthermore, if the analysis of conditions and

practices is too refined (for example, if very small subgroups are analyzed), large spurious differences between intervention groups could threaten the face validity of this analysis.

We will use all three of these methods to examine the relationship between program effects, conditions, and practices. When interpreting findings in study reports, findings that are robust to the choice of method will be reported with more confidence. All these findings will be interpreted cautiously, however, because all three approaches are fundamentally nonexperimental and may be influenced by correlations between unobserved factors and the conditions and practices we can observe.

REFERENCES

Adams, A., D. Carnine, and R. Gersten. "Instructional Strategies for Studying Content Area Texts in the Intermediate Grades." *Reading Research Quarterly*, vol. 18, 1982, pp. 27–55.

Anderson, R.C., and P.D. Pearson. "A Schema-Theoretic View of Basic Processes in Reading Comprehension." In *Handbook of Reading Research*, edited by P.D. Pearson, R. Barr, M.L. Kamil, and P. Mosenthal. White Plains, NY: Longman, 1984.

Anderson, R.C., L.L. Shirey, P.T. Wilson, and L.G. Fielding. "Interestingness of Children's Reading Material." In *Aptitude, Learning, and Instruction: Cognitive and Affective Process Analyses*, edited by R. Snow and M. Farr. Hillsdale, NJ: Erlbaum, 1987.

Anderson, V. "A Teacher Development Project in Transactional Strategy Instruction for Teachers of Severely Reading-Disabled Adolescents." *Teaching and Teacher Education*, vol. 8, 1992, pp. 391–403.

Anderson, V., and M. Roit. "Planning and Implementing Collaborative Strategy Instruction for Delayed Readers in Grades 6-10." *Elementary School Journal*, vol. 94, 1993, pp. 121–137.

Angrist, J.D., G.W. Imbens, and D.B. Rubin. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*, vol. 91, 1996, pp. 444–455.

Armbruster, B.B., and T.H. Anderson. "Structures of Explanations in History Textbooks or So What if Governor Stanford Missed the Spike and Bit the Rail?" *Journal of Curriculum Studies*, vol. 16, no. 2, 1984, pp. 181–194.

Armbruster, B.B., and J. Osborn. "Put Reading First: The Research Building Blocks for Teaching Children to Read." Washington, DC: National Institute for Literacy, 2001.

Armbruster, B.B., T.H. Anderson, and J. Ostertag. "Does Text Structure/Summarization Instruction Facilitate Learning from Expository Text?" *Reading Research Quarterly*, vol. 22, 1987, pp. 331–346.

Asher, S. "Topic Interest and Children's Reading Comprehension." In *Theoretical Issues in Reading Comprehension*, edited by R.J. Spiro, B.C. Bruce, and W.F. Brewer. Hillsdale, NJ: Erlbaum, 1980.

Au, K.H., and D.C. Crowell. "Using the Experience-Text-Relationship Method with Minority Children." *Reading Research Quarterly*, vol. 32, 1979, pp. 677–679.

Baker, L., and A.L. Brown. "Metacognitive Skills and Reading." In *Handbook of Reading Research, Volume I*, edited by P.D. Pearson, R. Barr, M.L. Kamil, and P. Mosenthal. New York: Longman, 1984, pp. 353–394.

Baker, S., R. Gersten, J. Dimino, and R. Griffiths. "The Sustained Use of Research-Based Instructional Practice: A Case Study of Peer-Assisted Learning Strategies in Mathematics. Contemporary Research." *Remedial and Special Education*, vol. 25, no. 1, 2004, pp. 5–21.

Baker, S., R. Gersten, D. Haager, C. Goldenberg, and M. Dingle. "The Relationship Between Observed Teaching Practice and Growth in Reading in First Graders Who Are English Learners." *Elementary School Journal* (in press).

Barr, Rebecca. "Interventions for Children Experiencing Early Reading Difficulties." In *Successful Reading Instruction*, edited by Michael L. Kamil, JoAnn B. Manning, and Herbert J. Walberg. Greenwich, CT: Information Age Publishing, 2002.

Beck, I.L., M.G. McKeown, C. Sandora, L. Kucan, and J. Worthy. "Questioning the Author: A Yearlong Classroom Implementation to Engage Students with Text." *Elementary School Journal*, vol. 96, 1996, pp. 385–414.

Blakely, C.H., J.P. Mayer, and R.G. Gottschalk. "The Fidelity-Adaptation Debate: Implication for the Implementation of Public Sector Social Programs." *American Journal of Community Psychology*, vol. 15, 1987, pp. 253–268.

Block, Cathy Collins, and Michael Pressley (eds.). *Comprehension Instruction: Research-Based Best Practices.* New York: Guilford Press, 2002.

Bloom, Howard, Johannes Bos, and Suk-Won Lee. 1999. "Using Cluster Random Assignment to Measure Program Impacts: Statistical Implications for the Evaluation of Education Programs." *Evaluation Review*, vol. 23, no. 4, 1999, pp. 445–469.

Bloom, Howard, Lashawn Richburg-Hayes, and Alison Rebeck Black. "Using Covariates to Improve Precision: Empirical Guidance for Studies That Randomize Schools to Measure the Impacts of Educational Interventions." MDRC Working Paper, November 2005.

Box, George P., William G. Hunter, and J. Stuart Hunter. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building.* New York: John Wiley, 1978.

Brown, A.L., and J.D. Day. "Macrorules for Summarizing text: The Development of Expertise." *Journal of Verbal Learning and Verbal Behavior*, vol. 22, 1983, pp. 1–14.

Brown, R., M. Pressley, P. Van Meter, and T. Schuder. "A Quasi-Experimental Validation of Transactional Strategies Instruction with Low-Achieving Second-Graders." *Journal of Educational Psychology*, vol. 88, 1996, pp. 18–37.

Carlisle, J. *Teacher's QUEST: Self-Administered Questionnaire.* Ann Arbor, MI: Regents of the University of Michigan, 2003.

Carlisle, J., and M. Rice. Improving Reading Comprehension: Research-Based Principles and Practices. Baltimore, MD: York Press, 2002.

Chall, K.S., V.A. Jacobs, and I.E. Baldwin. *The Reading Crisis: Why Poor Children Fall Behind.* Cambridge, MA: Harvard University Press, 1990.

Chen, H. *Theory-Driven Evaluations*. Thousand Oaks, CA: Sage, 1990.

Cochran, W.G., and D.B. Rubin. "Controlling Bias in Observational Studies: A Review." *Sankhya*, series A, vol. 35, 1973, pp. 417–446.

Collins, C. "Reading Instruction That Increases Thinking Abilities." *Journal of Reading*, vol. 34, 1991, pp. 510–516.

Consortium on Chicago School Research. "Improving Chicago's Schools: The Teachers' Turn, 1999; Elementary School Teacher Survey, 1999." Chicago, IL: Consortium on Chicago School Research, 1999. Available online at [www.consortium-chicago.org].

Cook, Thomas, and Donald Campbell. *Quasi-Experimentation: Design and Analysis Issues for Field Settings.* Hopewell, NJ: Houghton Mifflin, 1979.

Dallali, Gerard. "Multiple Comparison Procedures." *The Little Handbook of Statistical Practice.* Available online at [http://www.tufts.edu/~gdallal/LHSP.HTM].

Darch, C., and R. Gersten. "Direction Setting Activities in Reading Comprehension: A Comparison of Two Approaches." *Learning Disabilities Quarterly*, vol. 9, no. 3, 1986, pp. 235–243.

Darch, C., and E. Kame'enui. "Teaching LD Students Critical Reading Skills: A Systematic Replication." *Learning Disability Quarterly*, vol. 10, 1987, pp. 82–91.

Denton, C.A., S. Vaughn, and J.M. Fletcher. "Bringing Research-Based Practice in Reading Intervention to Scale." *Learning Disabilities Research and Practice*, vol. 18, no. 3, 2003, pp. 201–211.

DeWitz, P. "Comprehension Instruction: A Research Agenda for the 21st Century: Understanding Expository Texts." Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, 1997.

Dole, J.A., and T.L. Jetton. "Improving Literacy Through Professional Development: Success and Sustainability in a Middle School." In *Bridging the Gap: Improving Literacy Learning for Pre-Adolescent and Adolescent Learners Grades 4-12*, edited by D.S. Strickland and D.E. Alvermann. New York: Carnegie Corporation, 2004, pp. 164–182.

Dole, J.A., G. Duffy, L.R. Roehler, and P.D.P. Pearson. "Moving from the Old to the New: Research on Reading Comprehension Instruction." *Review of Educational Research*, vol. 61, 1991, pp. 239–264.

Duffy, G.G. "The Case for Direct Explanation of Strategies." In *Comprehension Instruction: Research-Based Best Practices*, edited by C.C. Block and M. Pressley. New York: Guilford Press, 2002, pp. 28–41.

Duffy, G.G., et al. "Effects of Explaining the Reasoning Associated with Using Reading Strategies." *Reading Research Quarterly*, vol. 23, 1987, pp. 347–386.

Duffy, G.G., and L.R. Roehler. "Why Strategy Instruction Is So Difficult and What We Need to Do About It." In *Cognitive Strategy Research: From Basic Research to Educational Applications,* edited by C.B. McCormick, G. Miller, and M. Pressley. New York: Springer-Verlag, 1989.

Duffy, G.G., L.R. Roehler, E. Sivan, G. Rackliffe, C. Book, M. Meloth, L.G. Vavrus, R. Wesseman, J. Putnam, and D. Bassiri. "Effects of Explaining the Reasoning Associated with Using Reading Strategies." *Reading Research Quarterly,* vol. 22, 1987, pp. 347–368.

Duke, N.K., and P.D. Pearson. "Effective Practices for Developing Reading Comprehension." In *What Research Has to Say About Reading Instruction* (Third edition), edited by A.E. Farstrup and S.J. Samuels. Newark, DE: International Reading Association, 2002, pp. 205–242.

Duke, Nell K. "For the Rich It's Richer: Print Environments and Experiences Offered to First-Grade Students in Very Low- and Very High-SES School Districts." *American Education Research Journal,* vol. 37, no. 3, 2000, pp. 456–457.

Dunnett, C.W. "A Multiple Comparison Procedure for Comparing Several Treatments with a Control." *Journal of the American Statistical Association,* vol. 50, 1955, pp. 1096–1121.

Durkin, D. "What Classroom Observations Reveal About Reading Comprehension Instruction." *Reading Research Quarterly,* vol. 14, 1978–1979, pp. 481–533.

Edmonds, M.S., and K.L. Briggs. "Instructional Content Emphasis Instrument." In *Reading in the Classroom: Systems for Observing Teaching and Learning,* edited by S.R. Vaughn and K.L. Briggs. Baltimore, MD: Paul H. Brookes, 2003, pp. 31–52.

Flavel, J. "Metacognitive Aspects of Problem Solving." In *The Nature of Intelligence,* edited by L.B. Resnick. Hillsdale, NJ: Erlbaum, 1976, pp. 231–235.

Foorman, B., and C. Schatschneider. "Measurement of Teaching Practices During Reading/Language Arts Instruction and Its Relationship to Student Achievement." In *Reading in the Classroom: Systems for Observing Teaching and Learning,* edited by S.R. Vaughn and K.L. Briggs. Baltimore, MD: Paul H. Brookes, 2003, pp. 1–30.

Freedman, D.A. "Ecological Inference and the Ecological Fallacy." Technical report no. 549. Prepared for the *International Encyclopedia of the Social & Behavioral Sciences.* Berkeley, CA: University of California, Department of Statistics, 1999.

Fuchs, D., and L.S. Fuchs. "Response to Intervention: A Blueprint for Practitioners, Policymakers, and Parents." *Teaching Exceptional Children,* vol. 38, 2005, pp. 57–61.

Fuchs, D., L.S. Fuchs, P.H. Mathes, and D.C. Simmons. "Peer-Assisted Strategies: Making Classrooms More Responsive to Diversity." *American Educational Research Journal,* vol. 34, 1997, pp. 174–206.

Gersten, R., S. Baker, and J.W. Lloyd. "Designing High-Quality Research in Special Education: Group Experimental Design." *Journal of Special Education,* vol. 34, no. 1, 2000, pp. 2–18.

Gersten, R., S. Baker, D. Haager, and A. Graves. "Exploring the Role of Teacher Quality in Predicting Reading Outcomes for First Grade English Learners: An Observational Study." *Remedial & Special Education*, vol. 26, 2005, pp. 197-206.

Gersten, R., D. Carnine, L. Zoref, and D. Cronin. "A Multifaceted Study of Change in Seven Inner City Schools." *Elementary School Journal*, vol. 86, no. 3, 1986, pp. 257–276.

Gersten, R., L.S. Fuchs, D. Compton, M. Coyne, C. Greenwood, and M. Innocenti. "Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education." *Exceptional Children*, vol. 71, 2005, pp. 149–164.

Gersten, R., L. Fuchs, J. Williams, and S. Baker. "Teaching Reading Comprehension Strategies to Students with Learning Disabilities." *Review of Educational Research*, vol. 71, 2001, pp. 279–320.

Gersten, Russell, Lynn S. Fuchs, Joanna P. Williams, and Scott Baker. "Teaching Reading Comprehension to Students with Learning Disabilities: A Review of the Research." *Review of Reading Research*, vol. 71, no. 2, summer 2001, pp. 279–320.

Gersten, R., S. Vaughn, D. Deshler, and E. Schiller. "What We Know About Using Research Findings: Implications for Improving Special Education Practice." *Journal of Learning Disabilities*, vol. 30, no. 5, 1997, pp. 466–476.

Gibson, S., and M.H. Dembo. "Teacher Efficacy: A Construct Validation." *Journal of Educational Psychology*, vol. 76, 1984, pp. 569–582.

Glazerman, Steve, and David Myers. "Assessing the Effectiveness of Education Interventions: Issues and Recommendations for the Title I Evaluation." Washington, DC: Mathematica Policy Research, May 17, 2004.

Graesser, A.C., J.M. Golding, and D.L. Long. "Narrative Representation and Comprehension." In *Handbook of Reading Research, Volume II*, edited by R. Barr, M.L. Kamil, P. Mosenthal, and P.D. Pearson. Mahwah, NJ: Lawrence Erlbaum, 1991, pp. 171–205.

Graves, M.F., C. Juel, and B.B. Graves. *Teaching Reading in the 21st Century* (Third edition). Boston, MA: Allyn and Bacon, 2004.

Gresham, F.M., D.L. MacMillan, M.E. Beebe-Frankenberger, and K.M. Bocian. "Treatment Integrity in Learning Disabilities Intervention Research: Do We Really Know How Treatments Are Implemented?" *Learning Disabilities Research & Practice*, vol. 15, 2000, pp.198–205.

Guthrie, J.T. "Engagement and Motivation in Reading Instruction." In *Successful Reading Instruction*, edited by Michael L. Kamil, JoAnn B. Manning, and Herbert J. Walberg. Greenwich, CT: Information Age Publishing, 2002.

Guthrie, J.T., K.E. Cox, E. Anderson, K. Harris, S. Mazzoni, and L. Rach. "Principles of Integrated Instruction for Engagement in Reading." *Educational Psychology Review*, vol. 10, no. 2, 1998, pp. 177–199.

Guthrie, J.T., A., Wigfield, and C. Von Secker. "Effects of Integrated Instruction on Motivation and Strategy Use in Reading." *Journal of Educational Psychology,* vol. 92, no. 2, 2000b, pp. 331–341.

Guthrie, J.T., W.D. Shafer, C. Von Secker, and T. Alban. "Contributions of Integrated Reading Instruction and Text Resources to Achievement and Engagement in a Statewide School Improvement Program." *Journal of Educational Research,* vol. 93, 2000a, pp. 211–226.

Guthrie, J.T., P. Van Meter, G.R. Hancock, S. Aloa, E. Anderson and A. McCann. "Does Concept Oriented Reading Instruction Increase Strategy Use and Conceptual Learning from Text?" *Journal of Educational Psychology,* vol. 90, no. 2, 1998, pp. 177–199.

Hansen, J. "The Effects of Inference Training and Practice on Young Children's Reading Comprehension." *Reading Research Quarterly*, vol. 16, 1981, pp. 391–417.

Hansen, J., and P.D. Pearson. "An Instructional Study: Improving the Inferential Comprehension of Good and Poor Fourth-Grade Readers." *Journal of Educational Psychology*, vol. 75, 1983, pp. 821–829.

Hochberg, Y., and A.C. Tamhane. *Multiple Comparison Procedures.* New York: Wiley, 1987.

Hohmann, A.A., and M.K. Shear. "Community-Based Intervention Research: Coping with the 'Noise' of Real Life in Study Design." *American Journal of Psychiatry*, vol. 159, 2002, pp. 201–207.

Holland, P.W. "Statistics and Causal Inference (with Discussion)." *Journal of the American Statistical Association,* vol. 81, 1986, pp. 945–970

Hoy, W.K., and A.E. Woolfolk. "Teachers' Sense of Efficacy Scale." *Elementary School Journal,* vol. 93, 1993, pp. 355–372.

Kamil, Michael. "Desirable Characteristics of Comprehension Interventions." Paper commissioned by the Institute of Education Sciences, December 2003.

Keogh, B.K. "Research in Learning Disability: A View of Status and Need." In *Theory and Research in Learning Disabilities,* edited by J.P. Das, R.F. Mulchahy, and A.E. Wall. New York: Plenum Press, 1982.

Kirk, R.E. *Experimental Design: Procedures for the Behavioral Sciences.* Belmont, CA: Brooks-Cole, 1968.

Klahr, D., Z. Chen, and E. Toth. "Cognitive Development and Science Education: Ships That Pass in the Night or Beacons of Mutual Illumination?" In *Cognition and Instruction: Twenty-Five Years of Progress,* edited by S.M. Carver and D. Klahr. Mahwah, NJ: Lawrence Erlbaum Associates, 2001.

Kucan, L., and I.L. Beck. "Thinking Aloud and Reading Comprehension Research: Inquiry, Instruction, and Social Interaction." *Review of Educational Research,* vol. 67, 1997, pp. 271–299.

Lane, David. *HyperStat Online Textbook*. Available online at [http://davidmlane.com/hyperstat/].

Langer, J.A. "Examining Background Knowledge and Text Comprehension." *Reading Research Quarterly*, vol. 19, 1984, pp. 486–481.

Langer, J.A. "From Research to Practice: A Prereading Plan." *Journal of Reading*, vol. 24, 1981, pp. 152–156.

Learning First Alliance. *Every Child Reading: An Action Plan of the Learning First Alliance.* Washington, DC: LFA, 1998.

Levin, H.M. *Cost-Effectiveness: A Primer*. London: Sage, 1983.

Levin, H.M., and P.J. McEwan. *Cost-Effectiveness Analysis* (Second edition). Thousand Oaks, CA: Sage, 2001.

Lloyd, J., D. Cullinan, E. Heins, and M. Epstein. "Direct Instruction: Effects on Oral and Written Language Comprehension." *Learning Disabilities Quarterly*, vol. 3, 1980, pp. 70–76.

Macias, C., R. Propst, C. Rodican, and J. Boyd. "Strategic Planning for ICCD Clubhouse Implementation: Development of the Clubhouse Research and Evaluation Screening Survey (CRESS)." *Mental Health Services Research*, vol. 3, 2001, pp. 155–167.

Martin, V.L., and M. Pressley. "Elaborative-Interrogation Effects Depend on the Nature of the Question." *Journal of Educational Psychology*, vol. 83, 1991, pp. 113–119.

McKeown, M.G., I.L. Beck, and M.J. Worthy. "Grappling with Text Ideas: Questioning the Author." *The Reading Teacher*, vol. 46, no. 7, 1993, pp. 560–566.

Meyer, B.J.F., D.M. Brandt, and G.J. Bluth. "Use of Top-Level Structure in Text: Key for Reading Comprehension of Ninth-Grade Students." *Reading Research Quarterly*, vol. 16, 1980, pp. 72–103.

Moats, L. *Teaching Reading Is Rocket Science*. Washington, DC: American Federation of Teachers, 1999.

Mowbray, C.T., M.C. Holter, G.B. Teague, and D. Bybee. "Fidelity Criteria: Development, Measurement, and Evaluation." *American Journal of Evaluation*, vol. 24, no. 3, 2003, pp. 315–340.

Myers, David, and Mark Dynarski. *Random Assignment in Program Evaluation and Intervention Research: Questions and Answers.* Princeton, NJ: Mathematica Policy Research, June 2003.

National Reading Panel. *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction.* Washington, DC: National Institute of Child Health and Human Development, December 2000.

Palincsar, A.S. "The Role of Dialogue in Providing Scaffolded Instruction." *Educational Psychologist*, vol. 21, 1986, pp. 73–98.

Palincsar, A.S., and A.L. Brown. "Reciprocal Teaching: A Means to a Meaningful End." In *Reading Education: Foundations for a Literate America,* edited by J. Osborn, P.T. Wilson, and R.C. Anderson. Lexington, MA: D.C. Heath, 1985.

Palincsar, A.S., and A.L. Brown. "Reciprocal Teaching of Comprehension-Fostering and Comprehension-Monitoring Activities." *Cognition and Instruction*, vol. 1, 1984, pp. 117–175.

Patching, W., E. Kame'enui, D. Carnine, R. Gersten, and G. Colvin. "Direct Instruction in Critical Reading." *Reading Research Quarterly*, vol. 18, 1983, pp. 406–418.

Pearson, P. D., and L. Fielding. "Comprehension Instruction." In *Handbook of Reading Research, Volume II*, edited by R. Barr, M.L. Kamil, P. Mosenthal, and P. Mosenthal. Mahwah, NJ: Lawrence Erlbaum, 1991, pp. 815–860.

Pearson, P.D., L.R. Roehler, J.A. Dole, and G.G. Duffy. "Developing Expertise in Reading Comprehension." In *What Research Has to Say About Reading Instruction* (Second edition), edited by S.J. Samuels and A.E. Farstrup. Newark, DE: International Reading Association, 1992, pp. 145–199.

Pressley, M. "Effective Beginning Reading Instruction." *Journal of Literacy Research,* vol. 34, 2002, pp. 165–188.

Pressley, M. "Comprehension Strategies Instruction: A Turn-of-the-Century Status Report." In *Comprehension Instruction: Research-Based Best Practices,* edited by C.C. Block and M. Pressley. New York: Guilford Press, 2002, pp. 11–27.

Pressley, M. "What Should Comprehension Instruction Be the Instruction of?" In *Handbook of Reading Research*, *Volume III*, edited by M. Kamil, P. Mosenthal, P.D. Pearson, and R. Barr. Mahwah, NJ: Erlbaum, 2000, pp. 545–562.

Pressley, M. *Reading Instruction That Works: The Case for Balanced Teaching.* New York: Guilford, 1998.

Pressley, M., and P. Afflerbach. *Verbal Protocols of Reading: The Nature of Constructively Responsive Reading*. Hillsdale, NJ: Lawrence Erlbaum, 1995.

Pressley, M., P.B. El-Dinary, R. Wharton-McDonald, and R. Brown. "Transactional Instruction of Comprehension Strategies in the Elementary Grades." In *Self-Regulated Learning: From Teaching to Self-Reflective Practice,* edited by D.H. Schunk and B.J. Zimmerman. New York: Guilford Press, 1998, pp. 42–56.

Pressley, M., P.B. El-Dinary, I. Gaskins, et al. "Beyond Direct Explanation: Transactional Instruction of Reading Comprehension Strategies." *Elementary School Journal*, vol. 92, 1992, pp. 511–554.

Pressley, M., K.R. Harris, and M.B. Marks. "But Good Strategy Instructors Are Constructivists!" *Educational Psychology Review*, vol. 4, 1992, pp. 3–31.

Pressley, M., C.J. Johnson, S. Symons, J.A. McGoldrick, and J.A. Kurita. "Strategies That Improve Children's Memory and Comprehension of Text." *Elementary School Journal*, vol. 90, 1989, pp. 3–32.

Pressley, M., E. Wood, V.E. Woloshyn, V. Martin, A. King, and D. Menke. "Encouraging Mindful Use of Prior Knowledge: Attempting to Construct Explanatory Answers Facilitates Learning." *Educational Psychologist*, vol. 27, 1992, pp. 91–110.

Pressley, Michael. "Comprehension Strategies Instruction: A Twentieth Century Report." In *Comprehension Instruction: Research-Based Best Practices*, edited by C.C. Block and M. Pressley. New York: Guilford Press, 2002.

Ramsey, Philip. "Comparison of Closed Testing Procedures for Pairwise Testing of Means." *Psychological Method*, vol. 7, no. 4, 2002.

RAND Reading Study Group. *Reading for Understanding: Toward an R&D Program in Reading Comprehension.* Washington DC: Office of Educational Research and Improvement, 2000.

Raphael, T.E., and P.D. Pearson. "The Effect of Metacognitive Awareness on Children's Question-Answering Behavior." Technical report no. 238. Urbana, IL: University of Illinois, Center for Study of Reading, 1982.

Raudenbush, S.W., and A.S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods* (Second edition). Thousand Oaks, CA: Sage, 2002.

Resnick, L.B. "Making America Smarter." *Education Week*, June 16, 1999, pp. 38–40.

Rosenshine, B., and C. Meister. "Reciprocal Teaching: A Review of the Research." *Review of Educational Research*, vol. 64, 1994, pp. 479–530.

Rosenshine, B., and R. Stevens. "Teaching Functions." In *Handbook of Research on Teaching*. New York: Macmillan, 1986, pp. 376–391.

Rosenshine, B., C. Meister, and S. Chapman. "Teaching Students to Generate Questions: A Review of the Intervention Studies." *Review of Educational Research*, vol. 66, no. 2, 1996, pp. 181–221.

Rothman, K. "No Adjustments Are Needed for Multiple Comparisons." *Epidemiology*, vol. 1, 1990, pp. 43–46.

Schraw, G., R. Bruning, and C. Zosvoboa. "Source of Situational Interest." *Journal of Reading Behavior*, vol. 27, 1995, pp. 1–17.

Schochet, Peter Z. "Statistical Power for Random Assignment Evaluations of Education Programs." Document no. PR05-36. Princeton, NJ: Mathematica Policy Research, Inc., June 2005.

Snow, C.E. *Reading for Understanding. Toward an R&D Program in Reading Comprehension.* Santa Monica, CA: RAND, 2002.

Snow, C.E., M.S. Burns, and P. Griffin (eds.). *Preventing Reading Difficulties in Young Children.* Washington, DC: National Academy Press, 1998.

Sparks, G.M. "Teachers' Attitudes Toward Change and Subsequent Improvements in Classroom Teaching." *Journal of Educational Psychology*, vol. 80, 1988, pp. 111–117.

Stoolmiller, M., J.M. Eddy, and J.B. Reid. "Detecting and Describing Preventive Intervention Effects in a Universal School-Based Randomized Trial Targeting Delinquent and Violent Behavior." *Journal of Consulting and Clinical Psychology*, vol. 68, 2000, pp. 296–306.

Tierney, R.J., and J.W. Cunningham. "Research on Teaching Reading Comprehension." In *Handbook of Reading Research, Vol. I,* edited by P.D. Pearson, R. Barr, M.L. Kamil, and P. Mosenthal. New York: Longman, 1984, pp. 609–656.

Trabasso, Tom. "Teaching of Comprehension in Content Areas: Design Criteria." Paper commissioned by the Institute of Education Sciences, December 2003.

U.S. Department of Education. "National Assessment of Educational Progress 2005 Assessment Results." Available online at [http://nces.ed.gov/nationsreportcard/]. Accessed February 6, 2006.

Vaughn, S., and K.L. Briggs (eds.). *Reading in the Classroom: Systems for the Observation of Teaching and Learning.* Baltimore, MD: Brookes, 2003.

Vaughn, S., and M. Coleman. "The Role of Mentoring in Promoting Use of Research-Based Practices in Reading." *Remedial and Special Education*, vol. 25, no. 1, 2004, pp. 25–38.

Vaughn, S., and S. Linan-Thompson. *Reading: Effective Instructional Activities for Elementary Students.* Alexandria, VA: Association for Supervision and Curriculum Development, 2004.

Vaughn, S., J.K. Klingner, and D.P. Bryant. "Collaborative Strategic Reading as a Means to Enhance Peer-Mediated Instruction for Reading Comprehension and Content Area Learning." *Remedial and Special Education*, vol. 22, no. 2, 2001, pp. 66–74.

Vaughn, S.R. "The Three-Tier Reading Intervention as a Means of Preventing Reading Difficulties." Unpublished fidelity of implementation instrument. Office of Special Education Programs, Grant #: H324X010013, 2002.

Vaughn, S.R. "Oracy/Literacy Development in Spanish-Speaking Children." Unpublished fidelity of implementation instrument. National Institute of Child Health and Human Development, National Institutes of Health, Grant #: UTA01-287, HD39521, 2001.

Vygotsky, L.S. *Mind in Society: The Development of Higher Psychological Processes.* Cambridge, MA: Harvard University Press, 1978.

Westfall, P. "Multiple Testing of General Contrasts Using Logical Constraints and Correlations." *Journal of the American Statistical Association*, vol. 92, 1997, pp. 299–306.

Westfall, P.H., and S.S. Young. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment.* New York: John Wiley & Sons, Inc., 1993.

Westfall, Peter, Randall Tobias, Dror Rom, Russell Wolfinger, and Yosef Hochberg. *Multiple Comparisons and Multiple Tests Using SAS.* Cary, NC: SAS Institute, 1999.

Wong, B.Y.L., and W. Jones. "Increasing Metacomprehension in Learning Disabled and Normally Achieving Students Through Self-Questioning Training." *Learning Disability Quarterly*, vol. 5, 1982, pp. 228–238.

Wood, E., M. Pressley, and P.H. Winne. "Elaborative Interrogation Effects on Children's Learning of Factual Content." *Journal of Educational Psychology*, vol. 82, 1990, pp. 741–748.

Yates, B.T. *Analyzing Cost, Procedures, Processes and Outcomes in Human Services.* Thousand Oaks, CA: Sage, 1996.

APPENDIX A

TECHNICAL WORKING GROUP MEMBERS

Intervention Technical Working Group Members

Dr. Donna Alvermann, University of Georgia

Dr. Timothy Shanahan, University of Illinois-Chicago

Dr. Joseph Torgesen, Florida State University

Dr. Joanna Williams, Columbia University

Impact Evaluation Technical Working Group Members

Dr. Donna Alvermann, University of Georgia

Dr. Mark Berends, Vanderbilt University

Dr. Isabel Beck, University of Pittsburgh

Dr. Thomas Cook, Northwestern University

Dr. David Francis, University of Houston

Dr. Larry Hedges, University of Chicago

Dr. Timothy Shanahan, University of Illinois-Chicago

Dr. Joseph Torgesen, Florida State University

Dr. Joanna Williams, Columbia University

APPENDIX B

THE ISSUE OF MULTIPLE COMPARISONS

The purpose of testing the statistical significance of the difference in outcomes between a treatment and control group is to determine whether the observed difference is most likely due to an effect of the treatment, rather than random chance. When more than one pair of groups is compared in this way (for example, when comparing multiple treatment groups to a single control group), the chances of falsely declaring at least one of the differences statistically significant is greater than when a single comparison is made. This is analogous to the difference in the probability of winning the lottery with a single ticket and the probability of winning the lottery with multiple tickets. This issue is known in the literature as the issue of multiple comparisons.

For this study, we will need to make multiple comparison adjustments when calculating MDEs and analyzing impacts (see Chapter III). These adjustments will be made when conducting three distinct types of comparisons (1) when comparing several treatment groups to a single control group, (2) when making all pair-wise comparisons among intervention groups (excluding the control group) and (3) when estimating impacts for subgroups of students.

The literature suggests two general approaches to addressing the issue of multiple comparisons. The first approach is to use adjustments that control the probability that at least one of the comparisons under consideration will be falsely declared significant. This probability is known as the familywise error rate (FWE). Most of the procedures developed to adjust for multiple comparisons are designed to control the FWE.

More recently, an alternative to controlling the FWE has been suggested. This second approach is to control the proportion of "discoveries" (that is, impacts that appear significant) that are not truly different from zero. This proportion is known as the false discovery rate (FDR). The difference between the FWE and the FDR is that the FWE criterion considers it unacceptable to have any effects declared falsely significant, whereas the FDR tries to keep the

number of falsely significant effects low relative to the number of correctly significant effects. Our primary analyses will be focused on controlling the more conservative FWE, however we will explore how our findings change when using a technique that controls the FDR.

In this appendix, we first describe the types of comparisons we intend to make. We then describe several methods of adjusting for multiple comparisons, weighing the advantages and disadvantages of each method. We conclude by describing the multiple comparison procedures that will be used on this study's design and analysis of impacts.

A. COMPARISONS PLANNED FOR THIS STUDY

We will make three qualitatively distinct sets of comparisons in this study that will require adjustment for multiple comparisons:

1. ***Comparing Multiple Interventions to a Single Control Group.*** Determining whether each intervention has a statistically significant impact is the primary focus of the study. Educators and policymakers need to be able to draw upon these findings with confidence that reported impacts are real, and not due to chance differences among groups. Therefore, we believe it is necessary to report the statistical significance of these effects after adjusting for multiple comparisons.

2. ***Making All Pair-Wise Comparisons Among Treatment Groups.*** Though this issue is important to educators and policymakers, the study is designed primarily to detect differences between intervention groups and a control group, not to detect differences among intervention groups. Since differences among intervention groups are likely to be substantially smaller than main impacts, a larger study would be needed to reliably detect these smaller effects. When interpreting findings in study reports, differences among treatment groups that are robust to multiple comparison adjustments will be reported with more confidence. (Differences that are significant only when no adjustments for multiple comparisons are made will be regarded as suggestive.)

3. ***Calculating Impacts Among Multiple Subgroups.*** Another circumstance in which adjustments for multiple comparisons are sometimes considered is when calculating effects for different subgroups. We will report subgroup findings with and without adjustments for multiple comparisons. For example, if we were to compare all four intervention groups to the control group for girls and all four intervention groups to the control group for boys, we would need to adjust for eight comparisons. When interpreting findings in study reports, subgroup findings that are robust to multiple comparison adjustments will be reported with more confidence. (Findings that are

significant only when no adjustments for multiple comparisons are made will be regarded as suggestive.)

In addition to these three sets of comparisons,[1] we will also make one comparison that does not require adjustment. In this comparison, we will compare the average intervention to a single control group. Education administrators and policymakers will probably want to know if supplemental reading instruction, in general, is effective at improving reading comprehension. To address this question, the study can examine whether the effect of the average intervention in this study is statistically significant. Since this is a single comparison, we will not adjust its statistical significance to account for multiple comparisons, yielding a more powerful test.

B. MULTIPLE METHODS THAT ADJUST FOR MULTIPLE COMPARISONS

We consider six different approaches to adjusting for multiple comparisons. The first five approaches are designed to control the FWE while the sixth approach controls the FDR. While the first five approaches all control the FWE, they differ in the settings in which they are designed to be applied and/or in terms of their statistical power.[2] The Bonferroni adjustment can be used in any setting but has the least statistical power. The Dunnett and Tukey adjustments are intended for more specific settings than Bonferroni and achieve greater statistical power by taking into account correlations among tests. The resampling approach can be applied in nearly

[1] It could be argued that, because we are comparing all treatments to a control group and making all pair-wise comparisons among treatment groups, we should adjust for all these comparisons in both contexts. That is, we are essentially proposing to make 10 comparisons (4 treatment groups versus a control and 6 pair-wise comparisons among the treatment groups), yet we are proposing to make separate adjustments for only the 4 comparisons and then for the next 6 comparisons. We believe this separation is appropriate because the primary focus of the study is on the difference between each treatment group and the control group. The pair-wise comparisons are of secondary interest and are intended as suggestive findings that can be used to inform future research.

[2] Useful discussions and examples of these methods are available at [http://www.tufts.edu/~gdallal/mc.htm] and [http://davidmlane.com/hyperstat/intro_ANOVA.html]. Hochberg and Tamhane (1987) provide a thorough review of multiple comparison procedures in use at that time. Westfall and Young (1993**)** and Westfall (1997) are examples of further refinements of these procedures.

as many settings as Bonferroni while also taking into account correlations between tests. Stepwise procedures can be applied in a variety of setting and increase power by taking into account logical relationships between tests.

1. A Conservative Approach: Bonferroni

The Bonferroni adjustment is a simple, conservative approach that controls the FWE but is not very powerful. The adjustment is simply to multiply the unadjusted p-value of a t-test by the total number of comparisons being made. The appeal of this approach is that it is easy to implement in any setting and with little computational cost. It can be used both when comparing multiple treatments to a single control group and when making all pair-wise comparisons among groups. It can also be used to adjust for additional comparisons that are not easily accounted for using other techniques—for example, adjusting for all pair-wise comparisons among treatment groups using multiple outcomes and multiple subgroups.

The disadvantage of this approach is that it does not take into account correlations among tests, resulting in a conservative adjustment. For example, when comparing multiple treatment groups to a single control group, the tests are correlated because they each involve the same control group. For example, the probability of falsely rejecting $H_{01}: T_A - CONT = 0$ (where T_A represents the mean outcome for the first treatment group (intervention A) and *CONT* represents the mean outcome for the control group) is not independent of the probability of falsely rejecting $H_{02}: T_B - CONT = 0$, because both tests are based on the same control group. By ignoring this correlation, we reduce the probability of detecting a real effect more than is needed in order to control the FWE. For example, when comparing four treatment groups to a single control group, the Bonferroni adjustment would be to multiply the p-value from the test by 4. If correlations among tests were taken into account (see section on Dunnett below), the adjustment would be to multiply the p-value from the test by approximately 3.4.

2. Comparing Multiple Treatments to a Control Group: Dunnett

The use of the multivariate t-distribution in the context of comparing multiple treatment groups to a common control group was first suggested by Dunnett (1955). By using the multivariate t-distribution, the Dunnett procedure controls for correlations between tests. An important assumption underlying the Dunnett test is that the test statistics actually follow the multivariate-t distribution. If the underlying distribution of the outcome of interest is heavily skewed, then this assumption is not valid (Westfall and Young 1993).

3. All Pair-Wise Comparisons: Tukey's Honest Significant Difference (HSD)

The same general approach used by Dunnett to compare multiple treatments to a control group can also be applied when making all pair-wise comparisons among multiple treatment groups. That is, we can calculate critical values for these test statistics, taking into account the correlations between the tests using a multivariate t-distribution. This approach is named after Tukey and is sometimes called Tukey's HSD. This approach is discussed in Hochberg and Tamhane (1987). As is the case with Dunnett, this approach relies on the assumption that the test statistics follow a multivariate-t distribution.

4. Resampling Techniques

Westfall and Young (1993) suggest using the bootstrap (or a permutation approach) to calculate p-values when making multiple comparisons. The primary appeal of this approach is that, instead of assuming test statistics follow a multivariate t-distribution, the joint distribution is estimated empirically. The approach is conceptually simple, requires fewer distributional assumptions than other methods, and controls the FWE.

The idea of resampling techniques in general is to simulate what would happen if an experiment were repeated a large number of times (say, 1,000). For example, to calculate the

variance of the sample mean using a resampling method, we would draw 1,000 samples (with replacement) from our actual data. For each sample, we would calculate the sample mean, yielding 1,000 estimates of the mean. The variance of the sample mean would be the variance of those 1,000 estimated means. The appeal of this approach is that we do not have to make any assumptions regarding the distribution of the sample mean. Instead, we estimate the distribution of the sample mean directly, based on our data.

In the case of multiple comparisons, we can use resampling to create 1,000 sets of test statistics corresponding to the comparisons of interest. We can then calculate the multivariate distribution of the test statistics using these 1,000 sets. This empirically estimated distribution can then be used to adjust for multiple comparisons rather than relying on the multivariate-t distribution.

This approach is particularly useful when there is good reason to believe that differences between groups do not follow a multivariate t-distribution—for example, when the outcomes of interest have a particularly skewed distribution. In general, test scores are approximately normally distributed. In this study, however, scores could be skewed due to the makeup of the sample (children in low-achieving schools). If we find that the distribution of test scores is skewed, we will investigate whether our interpretation of findings changes when using multiple comparison adjustments based on resampling.

5. Step-Wise Procedures

The Dunnett and Tukey procedures described above are single-step procedures in which adjustments are applied equally to all the tests being performed. For example, if we compare group A to B, B to C, and A to C, the same adjustment is used for each comparison.

Step-wise procedures do not apply the same adjustment to each test. Instead, they use information about whether one test, or set of tests, was rejected to alter the multiple-comparison adjustment made to subsequent tests.

Not all step-wise tests control the FWE. An example of a step-wise test that does not always control the FWE is a two-step procedure called Fisher's LSD, also sometimes called "protected LSD." In the first step, an F-test is used to evaluate the joint significance of all tests under consideration. If that test is rejected, all the individual t-tests are performed without any adjustment. Though intuitively appealing, this procedure suffers from a loophole. If all the null hypotheses are true (that is, there are no significant differences among groups), then this procedure controls the FWE. However, consider what happens if *just one* of the null hypotheses is false while *the rest are true*: the F-test is rejected due to the one false null, and all the other hypotheses are tested *with no adjustment whatsoever*.[3] In other words, the FWE for those remaining tests is not controlled. Other approaches that seem intuitively appealing, but upon closer examination do not control the FWE under all combinations of null and alternative hypotheses, are Neuman-Keuls and the Duncan Multiple Range test. See Hochberg and Tamhane (1987) for details as to why these methods fail to control the FWE.

Step-wise procedures that *always* control the FWE have also been developed. Hochberg and Tamhane discuss the theory behind these tests and explore several specific examples. Westfall (1997) and Westfall et al. (1999) provide further refinements to these procedures, as well as algorithms for implementation. For example, a step-wise procedure might rank the differences between all groups from the largest difference to the smallest difference. These procedures begin

[3] See Hochberg and Tamhane (1987), p. 3.

by applying a Tukey or Dunnett adjustment to the largest difference between groups, and then make less conservative adjustments to the remaining differences.

Though useful when calculating impacts, these step-wise procedures are less useful in the context of MDE calculations in this study. The MDE analysis for this study is focused on the largest difference between groups, because if we cannot detect the largest difference, we will be unable to detect any of the smaller differences. Therefore, MDE analyses are based on the Dunnett or Tukey adjustments.

6. A Procedure that Controls the FDR

Benjamini and Hochberg (1995) first suggested the FDR as an alternative to the FWE. They also suggested a step-wise procedure that controls the FDR. We will investigate the sensitivity of our findings to this alternative approach to adjusting for multiple comparisons. When describing findings in the study report, we will regard impacts that are significant when controlling the FWE with more confidence. (Findings that are not significant when controlling the FWE but are significant when controlling the FDR will be regarded as suggestive.)

C. CONCLUSION

To calculate MDEs, we will rely on the single-step procedures of Dunnett (comparing multiple treatment groups to a single control group) and Tukey (making all pair-wise comparisons among treatment groups). For impact analysis, we will rely on step-wise procedures that always control the FWE, though we will also explore the sensitivity of our findings to procedures that only control the FWE under the null hypothesis of no effects (Fisher's LSD) and methods that control the FDR.

APPENDIX C

DATA COLLECTION INSTRUMENTS

TEACHER SURVEY (2006-2007)

NATIONAL EVALUATION OF READING COMPREHENSION PROGRAMS

U.S. DEPARTMENT OF EDUCATION

ATTACH LABEL HERE
Teacher ID Teacher Name
School ID School Name

IF ABOVE INFORMATION IS INCORRECT,
PLEASE MAKE CORRECTIONS DIRECTLY ON LABEL.

This survey is part of the Evaluation of Reading Comprehension Programs, a national evaluation being conducted for the U.S. Department of Education. The questions ask about the training you received on the reading comprehension program, professional culture at your school, your reflections, and your background. All information you provide will be kept confidential. While you are not required to respond, your cooperation is needed to make the results of this survey comprehensive, accurate, and timely. Thank you.

Please return the completed form to:	**If you have questions, please contact:**
Mathematica Policy Research, Inc. 600 Maryland Ave., SW, Suite 550 Washington, DC 20024-2512 ATTN: Valerie Williams	Valerie Williams Phone: 888.535.0283 FAX: 202.863.1763 E-mail: vwilliams@mathematica-MPR.com

OMB NO.: ####-####
EXPIRATION DATE: ##/##/200#

SECTION I. READING COMPREHENSION PROGRAM TRAINING

This section asks about the training you recently received on the reading comprehension program you are using in your classroom as part of the Evaluation of Reading Comprehension Programs.

If you did not receive training in a reading comprehension program *for this study,* please check this box and go to Question 4 on the next page. → ☐ 1

1. Thinking about the initial training you received on the reading comprehension program you are using with your class, how would you rate the following:

IN EACH ROW, CHECK ONE BOX ONLY	POOR	FAIR	GOOD	EXCELLENT
a. Trainer's (or trainers') knowledge of reading comprehension instruction for fifth graders.....	1 ☐	2 ☐	3 ☐	4 ☐
b. Trainer's (or trainers') preparedness	1 ☐	2 ☐	3 ☐	4 ☐
c. Trainer's (or trainers') presentation style	1 ☐	2 ☐	3 ☐	4 ☐
d. Quality of content covered in training	1 ☐	2 ☐	3 ☐	4 ☐
e. Amount of content covered in training	1 ☐	2 ☐	3 ☐	4 ☐
f. Training schedule (i.e., amount of time spent on the various sessions)	1 ☐	2 ☐	3 ☐	4 ☐
g. Materials provided in training.....	1 ☐	2 ☐	3 ☐	4 ☐

2. Overall, how well did the initial training you received prepare you to use the reading comprehension program with your students?	NOT AT ALL	SOMEWHAT	VERY WELL
	1 ☐	2 ☐	3 ☐

3. If you have any other comments about the training, please note them below.

SECTION II. PROFESSIONAL CULTURE

This section asks about the professional culture within your school.[1]

4. CONVERSATIONS ABOUT TEACHING

During the past school year, how often have you had conversations with colleagues about…

IN EACH ROW, CHECK ONE BOX ONLY	LESS THAN ONCE A MONTH	2 OR 3 TIMES A MONTH	ONCE OR TWICE A WEEK	DAILY
a. The goals of this school?	1 ☐	2 ☐	3 ☐	4 ☐
b. Development of new curriculum?	1 ☐	2 ☐	3 ☐	4 ☐
c. Managing classroom behavior?	1 ☐	2 ☐	3 ☐	4 ☐
d. What helps students learn best?	1 ☐	2 ☐	3 ☐	4 ☐

5. MY GRADE LEVEL

How much do you disagree or agree with each of the following?

IN EACH ROW, CHECK ONE BOX ONLY	STRONGLY DISAGREE	DISAGREE	AGREE	STRONGLY AGREE
a. Teachers in this grade level trust each other	1 ☐	2 ☐	3 ☐	4 ☐
b. It's OK in this grade level to discuss feelings, worries, and frustrations with other teachers	1 ☐	2 ☐	3 ☐	4 ☐
c. Teachers respect other teachers who take the lead in grade level improvement efforts	1 ☐	2 ☐	3 ☐	4 ☐
d. Teachers in this grade level respect those colleagues who are expert at their craft	1 ☐	2 ☐	3 ☐	4 ☐

*PLEASE NOTICE **DIFFERENT** RESPONSE CHOICES FOR THE ITEM BELOW.*	NOT AT ALL	A LITTLE	SOME	A GREAT EXTENT
e. To what extent do you feel respected by other teachers in this grade level?	1 ☐	2 ☐	3 ☐	4 ☐

*PLEASE NOTICE **DIFFERENT** RESPONSE CHOICES FOR THE ITEM BELOW.*	NONE	SOME	ABOUT HALF	MOST	NEARLY ALL
f. How many teachers in this grade level really care about each other?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐

[1] Items in this section are from The Consortium on Chicago School Research. (1999). "Improving Chicago's Schools: The Teachers' Turn, 1999; Elementary School Teacher Survey, 1999." Chicago, IL. Available at [www.consortium-chicago.org.]

6. ACCESS TO NEW IDEAS

How often have you…

IN EACH ROW, CHECK ONE BOX ONLY	NEVER	ONCE	TWICE	3 TO 4 TIMES	5 TO 9 TIMES	10 OR MORE TIMES
a. Taken courses at a college or university relative to improving your school?	0☐	1☐	2☐	3☐	4☐	5☐
b. Participated in a network with other teachers outside your school?	0☐	1☐	2☐	3☐	4☐	5☐
c. Discussed curriculum and instruction matters with an outside professional group or organization?	0☐	1☐	2☐	3☐	4☐	5☐
d. Attended professional development activities organized by your school (include meetings that focus on improving your teaching)?	0☐	1☐	2☐	3☐	4☐	5☐
e. Attended workshops or courses sponsored by your school district (exclude required in-services)?	0☐	1☐	2☐	3☐	4☐	5☐
f. Attended professional development activities sponsored by the teachers' union?	0☐	1☐	2☐	3☐	4☐	5☐

7. MY EXPERIENCE OF CHANGE

How much do you disagree or agree with the following?

IN EACH ROW, CHECK ONE BOX ONLY	STRONGLY DISAGREE	DISAGREE	AGREE	STRONGLY AGREE
a. Most changes introduced at this school involve only a few teachers; rarely does the whole faculty become involved	1☐	2☐	3☐	4☐
b. We receive adequate professional development support for the changes we introduce at our school	1☐	2☐	3☐	4☐
c. Most changes introduced at this school gain little support among teachers	1☐	2☐	3☐	4☐

8. PROFESSIONAL DEVELOPMENT

How much do you disagree or agree with the following?

Overall, my professional development experiences over the past school year… *IN EACH ROW, CHECK ONE BOX ONLY*	STRONGLY DISAGREE	DISAGREE	AGREE	STRONGLY AGREE
a. …have included opportunities to work productively with teachers from other schools.	1 ☐	2 ☐	3 ☐	4 ☐
b. …have included enough time to think carefully about, to try, and to evaluate new ideas	1 ☐	2 ☐	3 ☐	4 ☐
c. …have deepened my understanding of subject matter	1 ☐	2 ☐	3 ☐	4 ☐
d. …have helped me understand my students better	1 ☐	2 ☐	3 ☐	4 ☐
e. …have been sustained and coherently focused, rather than being short term and unrelated	1 ☐	2 ☐	3 ☐	4 ☐
f. …have included opportunities to work productively with colleagues in my school	1 ☐	2 ☐	3 ☐	4 ☐
g. …have led me to make changes in my teaching	1 ☐	2 ☐	3 ☐	4 ☐
h. …have been closely connected to my school's improvement plan	1 ☐	2 ☐	3 ☐	4 ☐

CHECK ONE BOX ONLY	STRONGLY DISAGREE	DISAGREE	AGREE	STRONGLY AGREE
i. Most of what I learn in professional development addresses the needs of the students in my classroom	1 ☐	2 ☐	3 ☐	4 ☐

9. LEADERSHIP AND SUPPORT

How much do you disagree or agree with the following?

IN EACH ROW, CHECK ONE BOX ONLY	STRONGLY DISAGREE	DISAGREE	AGREE	STRONGLY AGREE
a. The principal at this school is strongly committed to shared decision making.	1 ☐	2 ☐	3 ☐	4 ☐
b. The principal at this school works to create a sense of community in the school.	1 ☐	2 ☐	3 ☐	4 ☐
c. The principal at this school promotes parent and community involvement in the school.	1 ☐	2 ☐	3 ☐	4 ☐
d. The principal at this school supports and encourages teachers to take risks.	1 ☐	2 ☐	3 ☐	4 ☐
e. The principal at this school is willing to make changes.	1 ☐	2 ☐	3 ☐	4 ☐
f. Most changes introduced at this school receive strong support from the principal.	1 ☐	2 ☐	3 ☐	4 ☐
g. The principal at this school encourages teachers to try new methods of instruction.	1 ☐	2 ☐	3 ☐	4 ☐

10. THOUGHTS ABOUT TEACHING READING[2]

How much do you agree or disagree with the following?

IN EACH ROW, CHECK ONE BOX ONLY	STRONGLY DISAGREE	DISAGREE	AGREE	STRONGLY AGREE
a. I feel I need to make changes in the methods I use to teach children to read and spell	1 ☐	2 ☐	3 ☐	4 ☐
b. I get help from staff members to understand some children's difficulties learning to read	1 ☐	2 ☐	3 ☐	4 ☐
c. I have benefited from opportunities to learn more about methods for teaching reading	1 ☐	2 ☐	3 ☐	4 ☐
d. The children in my class are making satisfactory progress in learning to read	1 ☐	2 ☐	3 ☐	4 ☐
e. I do not have sufficient materials to teach reading effectively	1 ☐	2 ☐	3 ☐	4 ☐
f. I do not understand why some children learn to read easily while other children struggle to learn basic reading skills	1 ☐	2 ☐	3 ☐	4 ☐
g. The literacy coach supports my efforts to teach reading effectively *IF A LITERACY COACH IS NOT AVAILABLE FOR FIFTH-GRADE STUDENTS, PLEASE SKIP THIS QUESTION AND CHECK THIS BOX* → ☐ 1	1 ☐	2 ☐	3 ☐	4 ☐
h. I have a good understanding of how children acquire language and literacy skills	1 ☐	2 ☐	3 ☐	4 ☐
i. I wish I had more opportunities to discuss how to teach reading with other teachers	1 ☐	2 ☐	3 ☐	4 ☐
j. I feel I am good at teaching reading and writing	1 ☐	2 ☐	3 ☐	4 ☐
k. The principal of my school supports my efforts to teach reading effectively	1 ☐	2 ☐	3 ☐	4 ☐
l. I would like to learn methods to help children develop their oral language	1 ☐	2 ☐	3 ☐	4 ☐
m. I look for opportunities to learn effective methods to teach reading and writing	1 ☐	2 ☐	3 ☐	4 ☐
n. I could do a better job teaching reading if I had more assistance from aides or volunteers in my class	1 ☐	2 ☐	3 ☐	4 ☐
o. I know how to assess the progress of my students in reading	1 ☐	2 ☐	3 ☐	4 ☐
p. The parents of children in my class support my efforts to teach their children to read	1 ☐	2 ☐	3 ☐	4 ☐
q. The school day is organized to maximize instructional time	1 ☐	2 ☐	3 ☐	4 ☐

[2] Items on this page were borrowed from Joanne Carlisle's "Teacher's QUEST: Self-Administered Questionnaire" (Regents of the University of Michigan: Ann Arbor, MI, 2003), with minor modifications.

SECTION III. TEACHER REFLECTIONS

This section asks for your reflections.[3]

11. TEACHER REFLECTIONS

IN EACH ROW, CHECK ONE BOX ONLY	NOTHING	VERY LITTLE	SOME	QUITE A BIT	A GREAT DEAL
a. How much can you do to control disruptive behavior in the classroom?	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
b. How much can you do to motivate students who show low interest in schoolwork?	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
c. How much can you do to get students to believe they can do well in schoolwork?	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
d. How much can you do to help your students value learning?	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
e. How much can you do to get children to follow classroom rules?	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
f. How much can you do to calm a student who is disruptive or noisy?	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
g. How much can you use a variety of assessment strategies?	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
h. How much can you assist families in helping their children do well in school?	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
*PLEASE NOTICE **DIFFERENT** RESPONSE CHOICES FOR THE ITEMS BELOW.*	NOT AT ALL	SMALL EXTENT	MODERATE EXTENT	QUITE A BIT	A GREAT EXTENT
i. To what extent can you craft good questions for your students?	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
j. To what extent can you provide an alternative explanation or example when students are confused?	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
*PLEASE NOTICE **DIFFERENT** RESPONSE CHOICES FOR THE ITEMS BELOW.*	NOT AT ALL	SLIGHTLY	MODERATELY	QUITE WELL	EXTREMELY WELL
k. How well can you establish a classroom management system with each group of students?	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
l. How well can you implement alternative strategies in your classroom?	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐

[3] Items on this page were borrowed from W.K. Hoy and A.E. Woolfolk's "Teachers' Sense of Efficacy Scale" (*Elementary School Journal*, vol. 93, pp. 355–372), with minor modifications.

SECTION IV. BACKGROUND

This section asks about your background.

12. **How many years have you taught, either full-time or part-time, at the elementary or secondary level (not counting the current school year)?** *Include years teaching in both public and private schools. Do not include time spent as a student teacher.*

☐☐ TOTAL YEARS TEACHING

13. **How many years have you been teaching in THIS school (not counting the current school year)?** *If you have had a break in service of <u>one year or more</u>, please report the year that you returned to this school. Do not include time spent as a student teacher. Include years spent teaching both full- and part-time at this school.*

☐☐ TOTAL YEARS TEACHING AT THIS SCHOOL

14. **What grade levels have you taught?** *CHECK ALL THAT APPLY*

1 ☐ 1st grade	6 ☐ 6th grade	11 ☐ 11th grade
2 ☐ 2nd grade	7 ☐ 7th grade	12 ☐ 12th grade
3 ☐ 3rd grade	8 ☐ 8th grade	13 ☐ Ungraded
4 ☐ 4th grade	9 ☐ 9th grade	14 ☐ Kindergarten
5 ☐ 5th grade	10 ☐ 10th grade	15 ☐ Prekindergarten

15. **For each degree below, please check YES or NO in <u>Column A</u> to indicate if you hold that degree. For those degrees you hold, please specify your major field of study in <u>Column B</u> and the year you received the degree in <u>Column C</u>.**

IN EACH ROW, CHECK ONE BOX IN COLUMN A. IF YOU ANSWER YES, COMPLETE COLUMNS B AND C FOR THAT ROW.	A. DEGREE HELD		B. MAJOR	C. YEAR RECEIVED
	YES	NO		
a. Associate's degree	1 ☐	0 ☐	_____	☐☐☐☐
b. Bachelor's degree	1 ☐	0 ☐	_____	☐☐☐☐
c. Master's degree	1 ☐	0 ☐	_____	☐☐☐☐
d. Educational specialist or professional diploma (at least one year beyond a master's degree)	1 ☐	0 ☐	_____	☐☐☐☐
e. Certificate of Advanced Graduate Studies	1 ☐	0 ☐	_____	☐☐☐☐
f. Doctorate (Ph.D., Ed.D.)	1 ☐	0 ☐	_____	☐☐☐☐
g. Professional (M.D., D.D.S., J.D., L.L.B)	1 ☐	0 ☐	_____	☐☐☐☐

16. **Which of the following describes the teaching certificate you currently hold in this state?** *CHECK ONE ONLY*

1 ☐ Regular or standard state certificate or advanced professional certificate

2 ☐ Probationary certificate (the initial certificate issued after satisfying all requirements except the completion of a probationary period)

3 ☐ Provisional or other type given to persons who are still participating in an "alternative certification program"

4 ☐ Temporary certificate (requires some additional college coursework and/or student teaching before regular certification can be obtained)

5 ☐ Emergency certificate or waiver (issued to teachers who do not have regular certification who need to complete a regular certification program in order to continue teaching)

17. **In what content area does the teaching certificate marked above allow you to teach in this state (e.g., elementary general, secondary general, special ed., a specific subject matter)?**

_____ CONTENT AREA

18. **Column A: Please indicate if you participated in any professional development activities listed below in the past 12 months.**

Column B: If you mark "yes" in Column A, please indicate in Column B how many hours you spent on the activities. *Include courses you have taken for recertification or advanced certification, workshops sponsored by your district, conferences, or other training that is relevant to your teaching.*

IN EACH ROW, CHECK ONE BOX IN COLUMN A. IF YOU ANSWER YES, CHECK ONE BOX IN COLUMN B.	A. PARTICIPATED?		B. NUMBER OF HOURS			
	YES	NO	8 OR FEWER	9-16	17-32	33 OR MORE
a. Reading instruction	1 ☐	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
b. Science instruction.....	1 ☐	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
c. Social studies instruction	1 ☐	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐

19. **Are you male or female?**

1 ☐ Male

2 ☐ Female

20. **Are you of Hispanic or Latino origin?**

1 ☐ Yes

0 ☐ No

21. **How do you describe yourself?** (*CHECK ALL THAT APPLY*)

1 ☐ American Indian or Alaska Native

2 ☐ Asian

3 ☐ Black or African American

4 ☐ Native Hawaiian or Other Pacific Islander

5 ☐ White

22. **What is your year of birth?**

|___|___|___|___| YEAR

CONTACT INFORMATION

Please provide your contact information and the best time to reach you in case we have questions about your responses.

MR./MS. FIRST NAME LAST NAME

STREET APT. NUMBER

CITY STATE ZIP

E-MAIL ADDRESS

( )
PHONE NUMBER (INCLUDE AREA CODE)

BEST TIME TO REACH YOU

THANK YOU FOR COMPLETING THIS SURVEY FOR THE U.S. DEPARTMENT OF EDUCATION.

PRELIMINARY SCHOOL INFORMATION FORM
National Evaluation of Reading Comprehension Programs

School_____ Principal: _____

Person completing form: _____ Phone number_____

1. How many students are enrolled:
   a. In this school? ..... |__|__|__|__| **Total enrollment**
   b. In the fifth grade? ..... |__|__|__|__| **Fifth-grade students**

2. How many fifth-grade classes do you have? ..... |__|__| **Fifth-grade classes**

3. What percentage of your school's students are:
   a. Eligible for the federally funded free or reduced-price lunch program?... |__|__|__| **% of students**
   b. Classified as limited English proficient (LEP)? ..... |__|__|__| **% of students**

4. How many students enrolled in this school are:
   a. Hispanic or Latino? ..... |__|__|__|__| **Students**
   b. Not Hispanic or Latino? ..... |__|__|__|__| **Students**

5. How many students enrolled in this school are *(please select one or more categories for each student)*:
   a. American Indian or Alaska Native? ..... |__|__|__|__| **Students**
   b. Asian? ..... |__|__|__|__| **Students**
   c. Black or African American? ..... |__|__|__|__| **Students**
   d. Native Hawaiian or other Pacific Islander? ..... |__|__|__|__| **Students**
   e. White? ..... |__|__|__|__| **Students**

6. Did your school participate in Reading First in the 2005-2006 school year? ..... 1 ☐ Yes 0 ☐ No

Please complete the other side.

7. What resources does your school use for its fifth-grade reading curriculum? *(Please specify resources for all components of the reading curriculum, including reading comprehension.)*

Core Curriculum	**Name**	**Publisher**
Textbook		
Basal reader series		
Special program		
Supplemental Curriculum	**Name**	**Publisher**
Specify topic (e.g., phonics): _____ _____		
Specify topic (e.g., phonics): _____ _____		

8. Please complete the table below for the most current average reading and math standardized test scores for this school's fourth- and fifth-grade students.

Grade Level	Test	Publisher	Month/ Year	Reading		Math	
				Standard Score*	Nat'l Percentile	Standard Score*	Nat'l Percentile
4th							
4th							
5th							
5th							
*If standard scores are not available, check here if reporting:				1 ☐ Scaled Scores 2 ☐ Raw Scores		1 ☐ Scaled Scores 2 ☐ Raw Scores	

9. Did your school make Adequate Yearly Progress (AYP) in the **2005-2006** school year in the following areas:

   a. Reading/language arts ..... 1 ☐ Yes 0 ☐ No

   b. Mathematics ..... 1 ☐ Yes 0 ☐ No

   c. Attendance rate ..... 1 ☐ Yes 0 ☐ No

10. Did your school make Adequate Yearly Progress (AYP) in the **2004-2005** school year in the following areas:

   a. Reading/language arts ..... 1 ☐ Yes 0 ☐ No

   b. Mathematics ..... 1 ☐ Yes 0 ☐ No

   c. Attendance rate ..... 1 ☐ Yes 0 ☐ No

Please return this form to Mathematica Policy Research, Inc., in the postage-paid envelope provided or by faxing it to 202-863-1763, attention Melissa Dugger. Thank you very much.

SCHOOL INFORMATION FORM (2006-2007)
National Evaluation of Reading Comprehension Programs

INSERT SCHOOL LABEL HERE

1. For what grade levels does this school offer instruction? (CHECK ALL THAT APPLY)

1 ☐ Prekindergarten	5 ☐ 3rd grade	9 ☐ 7th grade
2 ☐ Kindergarten	6 ☐ 4th grade	10 ☐ 8th grade
3 ☐ 1st grade	7 ☐ 5th grade	11 ☐ Other (specify):_____
4 ☐ 2nd grade	8 ☐ 6th grade	12 ☐ Ungraded *(including ungraded special ed. students)*

2. What was the total number of students enrolled in this school around the first of October 2006?..... |__|__|__|__| **Students enrolled**

3. How many students enrolled in this school are:
   a. Hispanic or Latino? ..... |__|__|__|__| **Students**
   b. Not Hispanic or Latino? ..... |__|__|__|__| **Students**

4. How many students enrolled in this school are:
   (PLEASE SELECT ONE OR MORE CATEGORIES FOR EACH STUDENT)
   a. American Indian or Alaska Native? ..... |__|__|__|__| **Students**
   b. Asian? ..... |__|__|__|__| **Students**
   c. Black or African American? ..... |__|__|__|__| **Students**
   d. Native Hawaiian or other Pacific Islander? ..... |__|__|__|__| **Students**
   e. White? ..... |__|__|__|__| **Students**

5. What percentage of students in the 2006-2007 academic year are:
   a. Eligible for the federally funded free or reduced-price lunch program?..... |__|__|__| **% of students**
   b. Classified as limited English proficient (LEP)? ..... |__|__|__| **% of students**

6. How many <u>fifth-grade students</u> were enrolled in this school around the first of October 2006?..... |__|__|__|__| **Fifth-grade students**

7. How many <u>fifth-grade classes</u> do you have?..... |__|__| **Fifth-grade classes**

Please complete the other side.

8. What type of school is this? (CHECK ONE)

1 ☐ Regular

2 ☐ Special Program Emphasis *(science/math school, talented/gifted school, foreign language immersion school, etc.)*

3 ☐ Special Education *(primarily serves students with disabilities)*

4 ☐ Other **(specify):** _____

9. Does this school offer a magnet program? .....1 ☐ Yes 0 ☐ No

10. Is this a charter school? .....1 ☐ Yes 0 ☐ No

11. a. Is this a Title I school? .....1 ☐ Yes 0 ☐ No

b. *If yes:* Is it schoolwide Title I? .....1 ☐ Yes 0 ☐ No

12. Is your school participating in any comprehensive school reform?

1 ☐ Yes→ **Please describe:** _____

_____

_____

0 ☐ No

13. Please complete the table below for the most current average reading and math standardized test scores for this school's fourth- and fifth-grade students.

Grade Level	Test	Publisher	Month/ Year	Reading		Math	
				Standard Score*	Nat'l Percentile	Standard Score*	Nat'l Percentile
4th							
4th							
5th							
5th							
*If standard scores are not available, check here if reporting:				1 ☐ Scaled Scores 2 ☐ Raw Scores		1 ☐ Scaled Scores 2 ☐ Raw Scores	

Please return this form to Mathematica Policy Research, Inc., in the postage-paid envelope provided or by faxing it to 202-863-1763, attention Melissa Dugger.

Thank you very much.

Student bar-coded label

STUDENT RECORDS FORM (2006-07)
NATIONAL EVALUATION OF READING COMPREHENSION PROGRAMS

1. What is this student's **date of birth**? ☐☐/☐☐/☐☐
   MONTH DAY YEAR

2. Is this student **male or female**? 1 ☐ Male 2 ☐ Female

3. What is the student's **ethnicity**?
   1 ☐ Hispanic or Latino
   0 ☐ Not Hispanic or Latino
   9 ☐ Don't know

4. What is this student's **race**? (PLEASE SELECT ONE OR MORE)
   1 ☐ American Indian/Alaska Native
   2 ☐ Asian
   3 ☐ Black or African American
   4 ☐ Native Hawaiian or other Pacific Islander
   5 ☐ White
   9 ☐ Don't know

5. How many days was this student **absent** during the 2006-07 school year? (WRITE "0" IF NO ABSENCES)
   a. ☐☐☐ **Total** days absent in the 2006-07 school year
   b. ☐☐☐ **Unexcused** days absent in the 2006-07 school year (WRITE "NA" IF NOT AVAILABLE)

6. Is this student… (CHECK ONE IN EACH ROW)

	YES	NO
a. Classified as **limited English proficient (LEP)**?	1 ☐	0 ☐
b. Eligible for the federally funded **free or reduced-price lunch program**?	1 ☐	0 ☐

7. For which of the following **disability categories** has this student been officially identified? (CHECK ALL THAT APPLY)
   1 ☐ Autism
   2 ☐ Deaf-blindness
   3 ☐ Developmental delay
   4 ☐ Emotional disturbance
   5 ☐ Hearing impairment
   6 ☐ Learning disability
   7 ☐ Mental retardation
   8 ☐ Orthopedic impairment
   9 ☐ Other health impairment
   10 ☐ Speech or language impairment
   11 ☐ Traumatic brain injury
   12 ☐ Visual impairment
   13 ☐ Other disability (SPECIFY):_____
   14 ☐ None of the above

Please complete the other side.

8. Which of the following **services** does this student receive in reading? (CHECK ALL THAT APPLY)

   1 ☐ Reading support

   2 ☐ Speech/Language support

   3 ☐ English as a Second Language (ESL)/English for Speakers of Other Languages (ESOL), English Language Development (ELD)

   4 ☐ Any other extra support or tutoring (i.e., Title I or other extra help to bring students up to grade-level expectations)

   5 ☐ None of the above

9. In what **grade** was this student enrolled in the 2006-07 school year? ☐☐ Grade

10. What was this student's **enrollment status** on the last day of the 2006-07 school year? (CHECK ONE)
If the student transferred, was expelled, or left for another reason, please fill in the box to the right.

   1 ☐ Enrolled at this school on the last day of the 2006-07 school year

   2 ☐ Transferred to another school.....

   3 ☐ Expelled .....

   4 ☐ Other (SPECIFY)_____

   _____ ....

   → Last day of attendance: ☐☐/☐☐/☐☐ (MONTH DAY YEAR)

   Name of new school: _____

   New school's address: _____ (CITY STATE)

11. Has this student been **promoted** to the next grade for the 2007-08 school year? (CHECK ONE)
If the student will attend a new school next year, please fill in the box to the right.

   1 ☐ Yes → Promoted to grade: ☐☐

   0 ☐ No

   9 ☐ Don't know

   If attending a new school next year:
   Name of new school: _____

   New school's address: _____ (CITY STATE)

Please return this form to Mathematica Policy Research, Inc., in the postage-paid envelope provided or by faxing it to 202-863-1763, attention Melissa Dugger.

Thank you very much.

Expository Reading Comprehension Classroom Observation Form

Background Information (or label)

Observer _____

School _____

District _____

Teacher _____

State _____

Grade _____

Today's Date____/____/_____
mm dd yyyy

Start time _____ a.m. p.m.

End time _____ a.m. p.m.

Subject (circle all that apply):
Social Studies Science
Reading/LA Intervention

	Number		Number
Maximum number of students observed in classroom		**Maximum number of adults observed providing instruction or educational support in the classroom (including teacher)**	

Any special circumstances that interrupted instruction? (please explain)

If this is an intervention observation please check below:

_____ Project CRISS _____Read About (Scholastic) _____Reading for Knowledge (SFA)

_____Read for Real _____SRA

Note to Rater: Focus on primary teacher for rating purposes. If student teacher is leading class, please do not observe.

Obs Initial_____ Date _____ School _____ Teacher _____ Time _____ to _____

Part I. 1st Interval (Start time:_____End Time:_____)

Comprehension

Before Reading	Models	Explains, Reviews, Names	Student Practice	Notes
1. The teacher/student activates prior knowledge and/or previews text before reading (e.g., shares background information about the title, author, content; reviews relevant content from previous lessons; picture walk; makes predictions; makes connections)				
Before, During, or After Reading	Models	Explains, Reviews, Names	Student Practice	
2. Instruction using text features (sub-heads, captions, charts, maps, graphs, sidebars, bold and italicized words) to interpret text				
3. Using text structure to teach/identify compare-contrast, cause-effect, or problem-solution (may include story grammar/elements if using informational text that has a narrative structure)				
4a. Explicit comprehension instruction that teaches students how to use strategies, such as main idea, summarizing, drawing conclusions, visualizing events, evaluating predictions, identifying fact vs. opinion, sequencing, monitoring for comprehension, other _____.				
4b. Generating questions				
5. Asks students to justify or elaborate their responses (e.g., teacher asks "why," "how did you reach that conclusion," etc.)				
During or After Reading	Models	Explains, Reviews, Names	Student Practice	
6. Teacher asks questions based on material in the text that require one of the following: (making inferences [MI], summarizing/finding main ideas [S], drawing conclusions [DC] or some other complex skill). **Please indicate in notes your best guess at type of skill (e.g., MI, S, DC, other)**				
7. Teacher elaborates, clarifies, or links concepts during text reading. May be an elaboration of student responses.				

Vocabulary (includes concepts, terminology, ideas; may be technical or complex content-area vocabulary)

	Tally	Notes
1. The teacher provides an explanation, a definition, or an example.		
2. The teacher elaborates or extends a definition. May include using multiple or contrasting examples to pinpoint a definition; further developing or paraphrasing the definition by incorporating ideas from students' responses, examples, and experiences; or discussing multiple-meanings.		
3. The teacher uses visuals, gestures related to word meaning, facial expressions, pictures, or demonstrations to discuss/demonstrate word meanings.		
4. Teaches word-learning strategies—using context clues, word parts, or root meaning.		
5. Asks students to do something that requires knowledge of words (e.g., answer questions, define words, make sentences, find words based on clues, physically demonstrate meaning).		
6. Gives students opportunity to apply word-learning strategies—using context clues, word parts, root meaning.		

Grouping arrangements and text reading (Code during each 15-minute cycle)

Teacher is working with (*Choose all that apply.)*	***Grouping arrangements*** *(Choose all that apply.)*	***Text Reading*** (*Choose all that apply.)*
1. Whole class (≥75% of class) 2. Large group (> 6 students, < 75% of class) 3. Small groups (3-6 students) 4. Pairs 5. An individual 6. No direct student contact	1. Whole class (≥75% of class) 2. Large group (> 6 students, < 75% of class) 3. Small groups (3-6 students) 4. Pairs 5. Working with an adult (other than a teacher) 6. Reading individually 7. Doing individual work (e.g., seat work/computer)	1. Supported oral reading (includes choral reading) 2. Independent silent reading 3. Independent oral reading 4. Teacher reads aloud 5. Teacher reads aloud with students following along silently 6. Text not used for comprehension instruction
1 2 3 4 5 6	1 2 3 4 5 6 7	1 2 3 4 5 6

Materials (Check materials that were used during the interval):

1. ___maps, charts, and graphs (including workbook, worksheet pages, or graphic organizers)
2. ____other visuals (with or without print)
3. ____text-basal
4. ____text-trade book, authentic text (e.g., poem, nonfiction book, song)
5. ____textbook-science
6. ____textbook-social studies
7. ____text handouts from supplementary materials, magazines
8. ____text-student made
9. ____computers
10. ____audio-tapes (e.g., books on tape)
11. ____workbook pages and worksheets
12. ____ chalk board or equivalent (dry erase, easel, overhead)
13. ____ videos
14. ____ paper and pencil
15. ____ other

Part II.

Answer the following questions at the end of your observation:

Features of Effective Instruction

During/After instruction, the teacher:

1. Gave inaccurate and/or confusing explanations or feedback.	Y	N	
2. Missed opportunity to correct or address error.	Y	N	
3. Provided opportunities for most students to participate actively during teacher-led instruction.	Y	N	
4. Paced instruction so that the length of the comprehension or vocabulary activities were appropriate for this age group.	Y	N	
5. Teaches using outlining and note taking.	Y	N	
6. Uses graphic organizers (e.g., KWL, Venn diagrams).	Y	N	
7. Keeps students thinking for 2+ seconds before calling on a student to respond to complex questions.	Y	N	
8. Gives independent/pairs/small-group practice in answering comprehension questions or applying comprehension strategy(ies) with expected product. (Can include response journals if a comprehension strategy is entailed.)	Y	N	
9. Uses writing activities in response to reading (does not include fill in the blank or one word answers).	Y	N	

Based on your overall observations, rate the quality of the comprehension instruction you observed.

	Not Observed	Minimal/Erratic	Partially Effective	Good	Excellent
10. Comprehension	N/O	1	2	3	4

Based on your overall observations, rate the teachers' management/responsiveness to students*.

	Minimal/Poor		Fair	Good	Excellent
11. The teacher maximized the amount of time available for instruction.		1	2	3	4
12. The teacher managed student behavior effectively to avoid disruptions and provide productive learning environments.		1	2	3	4
13. The teacher redirected discussion if a student response was leading the group off topic/focus.	N/O	1	2	3	4

* Items are adapted from Teacher Competency Checklist (Foorman & Schatschneider, 2003).

Based on your overall observations, rate student engagement during the observation.

	Few engaged	Many engaged	Most engaged
14. Student engagement during the **first half** of the observation session.	1	2	3
15. Student engagement during the remainder of the observation session.	1	2	3

Intervention-Specific Classroom Observation Form: CRISS

Background Information (or label)

Observer _____ Today's Date____/____/_____
mm dd yyyy

School _____

District _____ Start time _____ a.m. p.m.

Teacher _____ End time _____ a.m. p.m.

State _____

Grade _____

Subject (circle all that apply):
Social Studies Science
Reading/LA Intervention

	Number		Number
Maximum number of students observed in classroom		**Maximum number of adults observed providing instruction or educational support in the classroom (including teacher)**	

Any special circumstances that interrupted instruction? **(please explain)**

Notes to Rater:

1. **Focus on regular classroom teacher for rating purposes. If student teacher or substitute is leading class, please do not observe.**
2. **Make sure that the teacher is teaching with expository text for your observation.**

Star each section that you observe today. Answer the questions in that section only. Do *not* answer the questions in the sections that you do not observe.

Does the teacher…		**About how many of the students…**			
Section I. Preparing for Understanding					
1. Provide instruction or lead activities to generate background knowledge about (or review) a topic or concept before students read about it?	Y N	**1. Actively participate in the activities or discussion?**	Few	Many	Most
2. Help students set goals and determine a purpose before the students begin reading?	Y N				
Section II. Engaging Students with Content and Transforming Information					
3. Have students read a written text?	Y N	**3. Actively participate in reading a written text?**	Few	Many	Most
4a. Lead students during and/or after reading in transforming information activities (e.g., graphic organizer, guided discussion)?	Y N	**4a. Actively participate in transforming information activities?**	Few	Many	Most
4b. Include informal or formal writing in the transforming activities?	Y N	**4b. Actively participate in informal or formal writing?**	Few	Many	Most
5. Use the transforming activities to teach the *content* of the lesson?	Y N				
6. Discuss or reflect on students' metacognitive processes during the transforming activities?	Y N	**6. Actively participate in metacognitive discussion or reflection?**	Few	Many	Most
Section III. Reflecting on Content and Learning Processes					
7. Lead the whole class in a reflection discussion at the end of the lesson using questions *such as*: **A) Metacognition: How did you evaluate your comprehension?**	Y N	**7. Actively participate in the reflection discussion at the end of the lesson?**	Few	Many	Most

B) Background knowledge: Did I assist you in thinking about what you already knew? **C) Purpose Setting: Did you have clear purposes?** **D) Active Involvement: How were you actively engaged?** **E) Discussion: How did discussion clarify your thinking?** **F) Writing: How did you use writing to help you learn?** **G) Transformation: What were the different ways you transformed information? How did this help you?** **H) Teacher modeling: Did I do enough modeling?**					

Please note: You may see all three sections in one sitting. Or you may see Sections I and II, or Sections II and III, or Section II alone. You should never see Sections I and III together. It is also unlikely that you will see Sections I alone or Section III alone.

Intervention-Specific Classroom Observation Form: ReadAbout

Background Information (or label)

Observer _____ Today's Date____/____/_____
mm dd yyyy

School _____

District _____ Start time _____ a.m. p.m.

Teacher _____ End time _____ a.m. p.m.

State _____

Grade _____

Subject (circle all that apply):
Social Studies Science
Reading/LA Intervention

	Number		Number
Maximum number of students observed in classroom		**Maximum number of adults observed providing instruction or educational support in the classroom (including teacher)**	

Any special circumstances that interrupted instruction? **(please explain)**

Check which of the following were observed:

- **Computer-based instruction (Number of students working on the computer:_____; Time_____ [e.g., 18 minutes])**
- **Teacher-led small group instruction (Number of students in group:_____; Time _____)**
- **Independent work by students (Number of students doing independent work: _____; Time _____)**

A. Answer these questions while observing the lesson.

Computer-based instruction		**Teacher-led small group: Did the teacher**		**Independent work: Were the students**	
1. Was the computer program implemented for the required time?	Y N	**1. Explain and model the strategy or skill?**	Y N	**1. Using the program materials?**	Y N
2. Were students engaged in the computer program's activities?	Y N	**2. Provide opportunities for guided practice?**	Y N	**2. Actively engaged in completing the assignment?**	Y N
		3. Provide students instruction on the selected 6+1 Writing Trait?	Y N	**3. What was the primary focus of the students' independent activities?** o **Author's purpose** o **Main idea/details** o **Draw conclusions** o **Fact/opinion** o **Text structure (cause/effect, compare/contrast, sequence of events, problem/solution)** o **Make inferences** o **Summarizing** o **Visualizing** o **Setting purpose** o **Monitoring (including rereading and repairing)** o **Questioning**	
		4. Use the program materials?	Y N		
5. Which components of the computer module was the student working on (circle all that apply and were observed)? **1. Initial explanation and modeling of the strategy** **2. Guided practice** **3. Writing prompt** **4. Review of key vocabulary** **5. Skill tutorial**	Y N Y N Y N Y N Y N	**5. What was the primary focus of the teacher-led instruction?** o **Author's purpose** o **Main idea/details** o **Draw conclusions** o **Fact/opinion** o **Text structure (cause/effect, compare/contrast, sequence of events, problem/solution)** o **Make inferences** o **Summarizing** o **Visualizing** o **Setting purpose** o **Monitoring (including rereading and repairing)** o **Questioning**			

B. Rate the following items based on the overall observation.

	Most of the time	Some of the Time	Rarely	Not at all	Not observed
1. Were the students actively engaged in instruction?					
2. Did the teacher monitor on-going student progress/work during the lesson?					
3. Did the teacher provide corrective feedback to students?					

Intervention-Specific Classroom Observation Form: Reading For Knowledge
Days 2 and 4 version

Background Information (or label)

Observer _____ Today's Date____/____/_____
mm dd yyyy

School _____

District _____ Start time _____ a.m. p.m.

Teacher _____ End time _____ a.m. p.m.

State _____

Grade _____

Subject (circle all that apply):
Social Studies Science
Reading/LA Intervention

	Number		Number
Maximum number of students observed in classroom		**Maximum number of adults observed providing instruction or educational support in the classroom (including teacher)**	

Any special circumstances that interrupted instruction? **(please explain)**

Please record the following:

1. Unit #____ **2. Week #____** **3. Day #____** **4. Book Title_____**

Notes to Rater:

1. **Focus on the regular classroom teacher for rating purposes. If student teacher or substitute teacher is leading class, please do not observe.**
2. **If today's class period includes testing, please do not observe and reschedule the observation visit.**

A. Answer these questions while observing the lesson.

To what extent does the teacher follow the recommended procedures suggested in the teachers' manual…				**About how many of the students…**			
I. Set the Stage					Few < 25%	Many 25%-75%	Most > 75%
a. Post the reading goal?	Little	Some	Most				
b. Present the reading goal?	Little	Some	Most	**b. Actively engage in listening to the presentation of the reading goal?**	Few	Many	Most
c. Present the cooperative learning goal?	Little	Some	Most	**c. Actively engage in listening to the presentation of the cooperative learning goal?**	Few	Many	Most
d. Provide the vocabulary instruction or practice?	Little	Some	Most	**d. Actively participate in whole group vocabulary instruction and practice?**	Few	Many	Most
II. Active Instruction					Few < 25%	Many 25%-75%	Most > 75%
a. Use a whole group or partner activity to discuss key points about the day's skill/strategy?	Little	Some	Most	**a. Actively participate in the whole group or partner activity to discuss key points about the day's skill/strategy?**	Few	Many	Most
b. Provide feedback and prompts to partner pairs during partner reading?	Little	Some	Most	**b. Actively participate in partner reading <u>and</u> discussion?**	Few	Many	Most
c. Chart individual students' progress on the setting goals and charting progress forms during partner reading?	Little	Some	Most				
d. Review routines for Team Talk discussion?	Little	Some	Most				
e. Read aloud Team Talk questions?	Little	Some	Most				
f. Circulate the classroom and monitor team discussions and	Little	Some	Most	**f. Actively participate in the group Team**	Few	Many	Most

provide prompts?				**Talk discussion?**			
g. Ask team members to share with the class their responses <u>and</u> reasoning to Team Talk questions?	Little	Some	Most	**g. Appear engaged in the whole group Team Talk discussion?**	Few	Many	Most

B. Answer these two overall questions at the end of the lesson.

The teacher followed the recommended pacing for the lesson. (Recommended pacing is 35 minutes +/- 5 minutes.)	Y N
The teacher awarded cooperation and/or improvement points at some point in the lesson.	Y N

Intervention-Specific Classroom Observation Form: Read for Real
Phase: Learn

Background Information (or label)

Observer _____ Today's Date ____/____/____
mm dd yyyy

School _____

District _____ Start time _____ a.m. p.m.

Teacher _____ End time _____ a.m. p.m.

State _____

Grade _____

Subject (circle all that apply):
Social Studies Science
Reading/LA Intervention

	Number		Number
Maximum number of students observed in classroom		**Maximum number of adults observed providing instruction or educational support in the classroom (including teacher)**	

Any special circumstances that interrupted instruction? (please explain)

Phase: Learn

Does the teacher follow the procedures suggested in the Teacher Guide?

1. Unit Introduction
The teacher:

a. Asks a student to read the title of the unit. Y N
b. Comments/asks questions to pique interest. Y N
c. Introduces the reading partner for the unit. Y N

2. Before Reading
The teacher:

a. Asks a student to read the explanation of the Before Reading focus strategy. Y N
b. Discusses the Before Reading focus strategy with the students. Y N
c. Encourages students to apply the Before Reading focus strategy. Y N
d. Calls students' attention to the "My Thinking" box. Y N
e. Asks a student to read what the reading partner says about the Before Reading focus strategy by reading the information in the "My Thinking" box. Y N

Student Engagement

About how many students are engaged during this portion of the lesson?
_____Few _____Many _____Most

3. During Reading
The teacher:

a. Asks a student to read the explanation of the During Reading focus strategy. Y N
b. Discusses the During Reading focus strategy with the students. Y N
c. Asks a student to read the information in the "My Thinking" box. Y N
d. Encourages students to share their thinking about the During Reading focus strategy. Y N
e. Reads or asks students to read the selection aloud.
_____Never _____Sometimes _____Always
f. Stops and discusses the "My Thinking" notes at each "red strategy button."
_____Never _____Sometimes _____Always
g. Asks comprehension questions.
_____Never _____Sometimes _____Always

Student Engagement

About how many students are engaged during this portion of the lesson?
_____Few _____Many _____Most

4. After Reading

The teacher:

a. Asks students to read the After Reading focus strategy. Y N
b. Discusses the After Reading focus strategy with the students. Y N
c. Asks a student to read the information in the "My Thinking" box. Y N
d. Calls on students to implement the After Reading focus strategy. Y N

Comprehension

e. Administers the comprehension test Y N
f. Corrects tests with the class. Y N
g. Discusses responses. Y N

Organizing Information

h. Asks a student to read the information from the reading partner. Y N
i. Conducts activity with graphic organizer. Y N

Writing for Comprehension

j. Asks a student to read the information from the reading partner Y N
k. Reads or asks a student to read the summary. Y N
l. Identifies how the paragraphs and sentences in the summary correspond to the information on the graphic organizer. Y N
m. Discusses the three parts of a summary:
   Introduction Y N
   Body Y N
   Conclusion Y N
n. Informs students that the author went through several steps (rough drafts, editing, and proofreading) to get to finished product. Y N

Vocabulary

o. Instructs students in the vocabulary skill. Y N
p. Asks students to complete the vocabulary activity:
_____as a whole class _____ in small groups _____independently ____in partners

Fluency

q. Asks a student to read the fluency tip. Y N
r. Asks a student to read the selection. Y N
s. Gives students time to practice the selection. Y N

Student Engagement

About how many students are engaged during this portion of the lesson?
_____Few _____Many _____Most

English Learners

Integrates activities for English Language Learners into the lesson.
_____Never _____Sometimes _____Always

Intervention-Specific Classroom Observation Form: Read for Real
Phase: Practice

Background Information (or label)

Observer _____ Today's Date ____/____/____
mm dd yyyy

School _____

District _____ Start time _____ a.m. p.m.

Teacher _____ End time _____ a.m. p.m.

State _____

Grade _____

Subject (circle all that apply):
Social Studies Science
Reading/LA Intervention

	Number		Number
Maximum number of students observed in classroom		**Maximum number of adults observed providing instruction or educational support in the classroom (including teacher)**	

Any special circumstances that interrupted instruction? (please explain)

Phase: Practice

Does the teacher follow the procedures suggested in the Teacher Guide…

1. Before Reading
The teacher:

a. Tells student there is a new reading partner for this story. Y N
b. Tells students the reading partner will give them suggestions at the beginning and end of the selection to help them practice the strategies. Y N
c. Asks a student to read the Before Reading focus strategy. Y N
d. Discusses the Before Reading focus strategy with the students. Y N
e. Asks students to implement the Before Reading focus strategy. Y N
f. Lists students' comments on board. Y N

Student Engagement

About how many students are engaged during this portion of the lesson?
_____Few _____Many _____Most

2. During Reading
The teacher:

a. Asks a student to read the During Reading focus strategy. Y N
b. Asks a student to read the note from the reading partner. Y N
c. Reminds students to write notes about the During Reading focus strategy the way the reading partner did in the previous selection. Y N
d. Reads or asks a student to read the first paragraph or two of the selection. Y N
e. Discusses selection with students. Y N
f. Remainder of the selection is read: (Check all that apply.)
_____as a whole class _____ in small groups _____ in partners ____*independently.*
g. Reminds students to stop at the red buttons and write notes on their paper. Y N
h. Asks comprehension questions.
_____Never _____Sometimes _____Always

Student Engagement

About how many students are engaged during this portion of the lesson?
_____Few _____Many _____Most

3. After Reading

The teacher:

a. Asks students to read the After Reading focus strategy. Y N
b. Discusses or asks question about the After Reading focus strategy. Y N
c. Gives written assignment highlighting the After Reading focus strategy. Y N

Comprehension

e. Administers the comprehension test. Y N
f. Corrects tests with the class. Y N
g. Discusses responses. Y N

Organizing Information

h. Asks students to complete graphic organizer.

Writing for Comprehension

i. Asks students to write a summary based on their completed graphic organizer. Y N

Vocabulary

j. Instructs students in the vocabulary skill. Y N
k. Asks students to complete the vocabulary activity:
_____as a whole class _____ in small groups _____independently ____in partners

Fluency

l. Asks a student to read the fluency tip. Y N
m. Asks a student to read the selection. Y N
n. Gives students time to practice the selection. Y N

Student Engagement

About how many students are engaged during this portion of the lesson?
_____Few _____Many _____Most

English Learners

Integrates activities for English Language Learners into the lesson.
_____Never _____Sometimes _____Always

Intervention-Specific Classroom Observation Form: SRA

Background Information (or label)

Observer _____

School _____

District _____

Teacher _____

State _____

Grade _____

–

Today's Date ____/____/____
mm dd yyyy

Start time _____ a.m. p.m.

End time _____ a.m. p.m.

Subject (circle all that apply):
Social Studies Science
Reading/LA Intervention

Maximum number of students observed in classroom **Number** _____

Maximum number of adults observed providing instruction or educational support in the classroom (including teacher) **Number** _____

Any special circumstances that interrupted instruction? (please explain)

Day 1: Social Studies/Science		
Part I. Lesson Segments		
BEFORE READING		
Does the teacher…		
1. Discuss students' prior experiences and knowledge by helping them make connections with the content of the lesson (i.e., text to self, text to world, or text to text?	Y	N
2. Ask a volunteer to read the title aloud?	Y	N
3. Ask a volunteer to identify the skill that will be practiced in the lesson?	Y	N
4. Direct students' attention to the text?	Y	N
5. Discuss the characteristics of the text with students, such as highlighted or boldfaced words, boxed text, subheads visuals, captions, etc.?	Y	N
6. Ask students to demonstrate understanding of the lesson's concepts and /or vocabulary (i.e., examples, synonyms, locate, point to)?	Y	N
DURING/AFTER READING		
Does the teacher*…*		
1. Read each section of the text aloud?	Y	N
2. Ask the students to follow along as he/she reads each section of the text?	Y	N
3. Involve students in the discussion of how the features/organization of the text is used as a means of drawing the reader into the selection (e.g., graphic organizers, maps, visuals, models, numbers, subheads, steps, etc.)?	Y	N
4. Give students practice using any one or more of the following strategies? (check all that apply) *_____Predicting/confirming predictions* *_____Summarizing* *_____Asking questions/finding answers* *_____Monitoring/clarifying (includes helping students use strategies for dealing with unfamiliar words in the text)* *_____Visualizing* *_____Monitoring for meaning (adjusting speed and/or rereading)* *_____Making connections*	Y	N
Part II. Other Components		
The teacher also… (circle one or more) *1. has volunteers look up the words in the dictionary for more detailed definitions of vocabulary* *2. models comprehension strategies when students struggle with practicing how to use comprehension strategy* *3. provides guided practice with map skills*		

Day 2: Social Studies/Science		
Part I. Lesson Segments		
BEFORE READING		
Does the teacher…		
1. Ask a volunteer to read the title aloud?	Y	N
2. Remind students that the text on this page relates to the previous page?	Y	N
3. Direct students' attention to the text?	Y	N
4. Discuss with students the characteristics of the text, such as highlighted or boldface words, boxed text, subheads, visuals, captions, etc.?	Y	N
DURING/AFTER READING:		
***Does the teacher**…*		
1. Read each section of the text aloud?	Y	N
2. Ask the students to follow along as he/she reads each section of the text?	Y	N
3. Involve students in the discussion of how the features/organization of the text is used as a means of drawing the reader into the selection (e.g., graphic organizers, maps, visuals, models, numbers, subheads, steps, etc.)?	Y	N
4. Give students practice using any one or more of the following strategies? (check all that apply) *_____Predicting/confirming predictions* *_____Summarizing* *_____Asking questions/finding answers* *_____Monitoring/clarifying (including helping students use strategies for dealing with unfamiliar words in the text)* *_____Visualizing* *_____Monitoring for meaning (adjusting speed and/or rereading)* *_____Making connections*		
Part II. Other Components		
The teacher also… (circle one or more) *1. models comprehension strategies when students struggle with practicing how to use a comprehension strategy* *2. provides guided practice with map skills, charts, graphs, etc.* *3. provides modeling for other skill work such as taking notes, drawing conclusions, etc.* *4. uses a "think aloud" technique to model a comprehension strategy* *5. invites students to "think aloud"*		

Day 3: Science		
Part I. Lesson Segments		
DURING/AFTER READING:		
***Does the teacher**…*		
1. Read each section of the text aloud?	Y	N
2. Ask the students to follow along as he/she reads each section of the text?	Y	N
3. Have students analyze the features/organization of the text in terms of how it is used as a means of drawing the reader into the selection (e.g., graphic organizers, maps, visuals, models, numbers, subheads, steps, etc.)?	Y	N
4. Give students practice using any one or more of the following strategies? (check all that apply) *_____Predicting/confirming predictions* *_____Summarizing* *_____Asking questions/finding answers* *_____Monitoring/clarifying (including helping students use strategies for dealing with unfamiliar words in the text)* *_____Visualizing* *_____Monitoring for meaning (adjusting speed and/or rereading)* *_____Making connections*	Y	N
Part II. Other Components		
The teacher also… (circle one or more) *1. models comprehension strategies when students struggle with practicing how to use a comprehension strategy* *2. provides guided practice with map skills*		

Day 3: Social Studies		
Part I. Lesson Segments		
BEFORE READING		
Does the teacher…		
1. Direct students' attention to the text?	Y	N
2. Discuss with students the characteristics of the text such as highlighted or boldfaced words, boxed text, subheads, visuals, captions, etc.?	Y	N
DURING/AFTER READING:		
***Does the teacher**…*		
1. Read each section of the text aloud?	Y	N
2. Ask the students to follow along as he/she reads each section of the text?	Y	N
3. Have students analyze the features/organization of the text in terms of how it is used as a means of drawing the reader into the selection (e.g., graphic organizers, maps, visuals, models, numbers, subheads, steps, etc.)?	Y	N
4. Give students practice using any one or more of the following strategies (check all that apply) *_____Predicting/confirming predictions* *_____Summarizing* *_____Asking questions/finding answers* *_____Monitoring/clarifying (including helping students use strategies for dealing with unfamiliar words in the text)* *_____Visualizing* *_____Monitoring for meaning (adjusting speed and/or rereading)* *_____Making connections*	Y	N
Part II. Other Components		
The teacher also… (circle one or more) *1. models comprehension strategies when students struggle with practicing how to use a comprehension strategy* *2. provides guided practice with map skills*		