Effectiveness of Selected Supplemental Reading Comprehension Interventions: Impacts on a First Cohort of Fifth-Grade Students

Executive Summary

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE

Institute of Education Sciences

Effectiveness of Selected Supplemental Reading Comprehension Interventions: Impacts on a First Cohort of Fifth-Grade Students

Executive Summary

May 2009

Susanne James-Burdumy
Wendy Mansfield
John Deke
Nancy Carey
Julieta Lugo-Gil
Alan Hershey
Aaron Douglas
Mathematica Policy Research, Inc.

Russell Gersten
Rebecca Newman-Gonchar
Joseph Dimino
RG Research Group

Bonnie Faddis
RMC Research Corporation

Audrey Pendleton
Project Officer
Institute of Education Sciences

IES NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
Sue Betka
Acting Director

National Center for Education Evaluation and Regional Assistance
Phoebe Cottingham
Commissioner

May 2009

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the reports.

To order copies of this report,

- Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 1398, Jessup, MD 20794-1398.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327 (800-USA-LEARN). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 800-437-0833.
- Fax your request to 301-470-1244.
- Order online at www.edpubs.org.

This report also is available on the IES website at http://ies.ed.gov/ncee.

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The research team for this evaluation consists of a prime contractor, Mathematica Policy Research, and two major subcontractors: RG Research Group and RMC Research Corporation. None of these organizations or their key staff members have financial interests that could be affected by findings from the study. None of the members of the Technical Working Group, convened by the research team to provide advice and guidance, have financial interests that could be affected by findings from the study.

EXECUTIVE SUMMARY

EFFECTIVENESS OF SELECTED SUPPLEMENTAL READING COMPREHENSION INTERVENTIONS: IMPACTS ON A FIRST COHORT OF FIFTH-GRADE STUDENTS

There are increasing cognitive demands on student knowledge in middle elementary grades where students become primarily engaged in reading to learn, rather than learning to read (Chall 1983). Children from disadvantaged backgrounds often lack general vocabulary, as well as vocabulary related to academic concepts that enable them to comprehend what they are reading and acquire content knowledge (Hart and Risley 1995). They also often do not know how to use strategies to organize and acquire knowledge from informational text in content areas such as science and social studies (Snow and Biancarosa 2003). Instructional approaches for improving comprehension are not as well developed as those for decoding and fluency (Snow 2002). Although multiple techniques for direct instruction of comprehension in narrative text have been well demonstrated in small studies, there is not as much evidence on teaching reading comprehension within content areas (National Institute of Child Health and Human Development 2000).

Improving the ability of disadvantaged students to read and comprehend text is an important element in federal education policy aimed at closing the achievement gap. Title I of the No Child Left Behind Act (NCLB) calls on educators to close the gap between low- and high-achieving students using approaches that scientifically based research has shown to be effective. Such rigorous research is relatively scarce, however, so it is difficult for educators to determine how best to use Title I funds to improve student outcomes. Identifying interventions that improve reading comprehension is part of this challenge.

The Institute of Education Sciences (IES) of the Department of Education (ED) has undertaken a rigorous evaluation of curricula designed to improve reading comprehension as one step toward meeting that research challenge. In 2004, ED contracted with Mathematica Policy Research, Inc. (MPR) and its subcontractors to conduct the study.[1] The study team worked with ED to refine the study design and select the curricula to be tested, and then recruited districts and schools, collected data on implementation and outcomes, and analyzed the data. The study was conducted based on a rigorous experimental design for assessing the effects of four reading comprehension curricula on reading comprehension among fifth-grade students in selected districts across the country, where schools were randomly assigned to use one of the four treatment curricula or to a control group.

---

[1]These subcontractors were RMC Research Corporation, RG Research Group, the Vaughn Gross Center for Reading and Language Arts at the University of Texas at Austin, the University of Utah, and Evaluation Research Services.

The experimental design ensures a strong basis for answering the study's key research questions:

1. What is the impact of the reading comprehension curricula as a whole on reading comprehension, and how do the impacts of the individual curricula compare to one another?

2. How are student, teacher, and school characteristics related to impacts of the curricula?

3. Which instructional practices are related to impacts of the curricula?

This report focuses on findings based on the first year of data collected for the study. It presents findings about the impacts of the reading comprehension interventions over one school year (2006-2007) for a first cohort of fifth graders.[2] The main finding from the first year of the study regarding the basic question of intervention effectiveness is:

- **Reading comprehension test scores in schools randomly assigned to use one of the four reading comprehension curricula were not statistically significantly higher than scores in control schools.** In addition, there was evidence that test scores were lower in treatment schools than in control schools (4 of the 20 impacts[3] comparing treatment and control group test scores were negative and statistically significant, effect sizes: -0.14 and -0.21 for Reading for Knowledge on the composite test score and science comprehension test score, respectively, and -0.08 for the combined treatment group on both the composite test score and Group Reading Assessment and Diagnostic Evaluation (GRADE) test score).[4]

---

[2]Impacts are reported as "effect sizes" to facilitate comparisons of impacts on different outcomes. The effect size is the impact divided by the standard deviation of the outcome for students in the control group. For example, an impact of 4 units on an outcome with a standard deviation of 20 would be reported as effect size of 0.20. When control group means are shown in tables in the report, they are the *actual* control group means (they are not regression-adjusted means). Unadjusted means for treatment groups are presented in Table K.1 in Appendix K.

[3]The 20 impacts arise from having 4 reading comprehension assessments (Group Reading Assessment and Diagnostic Evaluation (GRADE) (Williams 2001), ETS science comprehension (Educational Testing Service 2007a), ETS social studies comprehension (Educational Testing Service 2007b), and a composite test score that is an average of the three tests listed here) and 5 intervention groups for whom impacts were estimated (4 individual intervention groups and the combined treatment group, which groups the 4 interventions together).

[4]To put this in perspective, for a student at the 50th percentile, an effect size of 0.10 represents about 4 percentile points, an effect size of 0.15 represents about 6 percentile points, and an effect size of 0.20 represents about 8 percentile points. To provide additional perspective, a meta-analysis by Rosenshine and Meister (1994) found an average effect size of 0.32 across nine studies examining the impact of multiple reading comprehension strategy instruction on standardized test scores (this meta-analysis focused on reciprocal teaching, which involves the use of guided practice and dialogue between students and teachers to teach students about four comprehension strategies including question generation, summarization, prediction, and clarification). Another meta-analysis by Rosenshine, Meister, and Chapman (1996) found an average effect size of 0.36 across 13 studies examining the impact of question generation on standardized test scores.

The main finding from the first year of the study regarding questions about for whom and under what conditions the interventions may be effective is:

- **Reading comprehension test scores in schools using the selected reading comprehension curricula were statistically significantly lower than scores in control schools for some subgroups defined by student, teacher, and school characteristics.** These subgroups include students with:

    - above-average baseline fluency levels (effect size: -0.23 for the combined treatment group on the social studies comprehension test score),

    - students with baseline comprehension levels in the lowest third of the sample (effect sizes: -0.08 and -0.09 for the combined treatment group on GRADE and composite test scores, respectively),[5]

    - students in schools with below-average School Professional Culture scale scores[6] (effect size: -0.14 for the combined treatment group on the composite test score),

    - students in schools with an above-average concentration of students eligible for free or reduced-price lunch (effect size: -0.11 for the combined treatment group on the composite test score),

    - students in schools with a below-average concentration of English language learners (effect sizes: -0.15 for the combined treatment group on the composite test score and -0.19 for the difference in impacts [on the composite test score and for the combined treatment group] between students in schools with below-average and above-average concentrations of English language learners),

    - students whose teachers had more than five years of experience (effect size: -0.09 for the combined treatment group on the composite test score), and

    - students whose teachers had more than 10 years of experience (effect size: -0.36 for Reading for Knowledge on the science comprehension test score).

---

[5]These effect sizes are from impact models that included the *middle* and bottom third of students to permit an assessment of whether there was a difference in impacts between these two groups. We found effect sizes of -0.14 and -0.15 on the GRADE and composite test scores for students in the lowest third of the sample when the impact models included the *top* and bottom third of students.

[6]The School Professional Culture scale is based on 35 items from the study's Teacher Survey and reflects teachers' perceptions of the culture in their school, including relationships with colleagues, access to professional development, experiences with changes being implemented in their school, and leadership support in their school. See Chapter I and Appendix F for details.

Study Design

Drawing on input from the Title I Independent Review Panel (IRP) and the study's technical working group (TWG), IES decided on an evaluation plan (Glazerman and Myers 2004).[7] The study focused on upper elementary students—fifth graders—so that it complemented other IES initiatives to understand the effectiveness of Reading First for younger students, and to reflect the concern that disadvantaged students in upper elementary grades may still struggle with reading. The focus of the study was on testing curricula designed to improve comprehension of expository text. Outcomes were defined as the ability to comprehend such text generally and in two specific content areas, science and social studies.

SUMMARY OF FIRST-YEAR EVALUATION DESIGN

Intervention: Four reading comprehension curricula (Project CRISS, ReadAbout, Read for Real, and Reading for Knowledge) were selected as interventions for the study based on public submissions and ratings by an expert review panel.

Participants: 10 districts, 89 schools, 268 teachers, and 6,350 fifth-grade students. Districts were recruited from among those with at least 12 Title I schools, and schools were recruited only if they did not already use any of the four selected curricula. Students in those schools were eligible to participate if they were enrolled in fifth-grade classes when the baseline tests were administered in fall 2006 or if they enrolled after the baseline administration but before January 1, 2007. Students in combined fourth-/fifth- or fifth-/sixth-grade classes were excluded, as were those in special education classes, although special education students mainstreamed in regular fifth-grade classes were eligible to participate.

Research Design: Within each district, schools were randomly assigned to an intervention group that would use one of the four curricula or a control group that did not have access to any of the four curricula being tested. For example, in a district with 10 schools, 2 schools were assigned to each treatment group and 2 schools were assigned to the control group. Control group teachers could, however, use other supplemental reading programs. The study administered tests to students in intervention and control schools near the beginning and end of the 2006-2007 school year. It also observed classrooms during the school year and collected data from teacher questionnaires, student and school records, and from the intervention developers.

Outcomes: Impact estimates focused on student reading comprehension test scores.

Schools in districts that agreed to participate were randomly assigned to one of the five study arms (four intervention groups and one control group). Teachers in schools assigned to an intervention group developed their own strategies for incorporating the assigned reading comprehension curriculum into their daily schedules and their core reading instruction. (As described in more detail in the next section, the curricula being evaluated in this study were designed to supplement—not replace—the core curriculum being used by each teacher.) Teachers in control group schools continued to teach reading using whatever methods they had

---

[7]The Title I Independent Review Panel (IRP) was set up by Congress to provide ED with policy recommendations on Title I research. The MPR study design team worked with the IRP, TWG, and IES on defining the key elements of this study's design (as laid out in the text that follows) (Glazerman and Myers 2004).

been using in the absence of the study. Due to the experimental design, differences in outcomes of students in the treatment and control groups are attributable to the curricula being tested.[8]

This study provides educators with a sense of the effectiveness of these curricula when used for the first time by teachers in "real-world" conditions. Although the study team worked to facilitate study activities such as the collection of data in study schools, the developers provided teacher training and follow-up support to teachers throughout the year, and teachers and schools could discontinue use of the curricula if they believed they were ineffective or too challenging to use. Therefore, the study conditions may be comparable to those many districts might face if they implemented these curricula in their schools.

Selecting Curricula for the Study

The goal of the reading comprehension evaluation is to test "high quality" supplemental curricula that would be available to schools searching for ways to improve students' comprehension skills. An open, competitive process was used to solicit proposals from curriculum developers and to select study curricula. The plan, based upon the evaluation design and available resources, was to select four curricula for the study.

Proposals were formally solicited by the study team. The Request for Proposals (RFP) described the type of interventions to be included in the study. The reading comprehension interventions needed to supplement—not displace—the core reading, science, and/or social studies instruction in fifth-grade classrooms. They also needed to take an average of 30 to 45 minutes per day to implement and to encompass an entire school year.

In response to the RFP, a total of 13 proposals were submitted to the study team. Those that met a set of predetermined, minimum requirements were forwarded to the panel of reading experts for review.[9] The expert panel then assessed the extent to which the proposals met substantive criteria for inclusion in a pilot implementation stage. These criteria related to the theoretical and empirical underpinnings of the curriculum, evidence of the intervention's effectiveness, the support developers proposed to provide for teachers, the developers' institutional capability, and the appropriateness of the curriculum for the study's target population.

Five programs were selected to participate in a pilot implementation for the 2005-2006 academic year.[10] After the pilot year, four of the five curricula that were included in the pilot year were selected for the full implementation of the study. Based on the study team's recommendations, IES selected the following curricula:

---

[8]The study design just discussed is also described in James-Burdumy et al. (2006). Early study design proposals are laid out in Glazerman and Myers (2004).

[9]To meet the minimum requirements, a proposal needed to include a technical discussion of the intervention, teacher training materials, classroom materials, and a budget.

[10]During the pilot year, each developer recruited three Title I schools, trained an average of three teachers per school, and provided support to teachers during the year. The study team observed training and instruction, reviewed training and instructional materials, and provided formative feedback to the developers.

- **Project CRISS (**developed by CRISS) (Santa et al. 2004): Project CRISS focuses on five keys to learning—background knowledge, purpose setting, author's craft (which involves using text structure to improve comprehension), active learning, and metacognition. The program is designed to be used each day during language arts, science, or social studies periods.

- **ReadAbout** (developed by Scholastic) (Scholastic 2005): Students are taught reading comprehension skills such as author's purpose, main idea, cause and effect, compare and contrast, summarizing, and inferences, primarily through a computer program. Students apply what they have learned to a selection of science and social studies trade books.

- **Read for Real** (developed by Chapman University and Zaner-Bloser) (Crawford et al. 2005): In Read for Real, teachers work with a six-volume set of books to teach reading strategies students can use before, during, and after reading (such as previewing, activating prior knowledge, setting a purpose, main idea, graphic organizers, and text structures). Each of these units includes vocabulary, fluency, and writing activities.

- **Reading for Knowledge** (developed by the Success for All Foundation) (Madden and Crenson 2006): Reading for Knowledge makes extensive use of cooperative learning strategies and a process called SQRRRL (Survey, Question, Read, Restate, Review, Learn).

Recruiting Districts and Schools for the Study

The study team recruited school districts for the study beginning in January 2006. The team focused on districts that served low-income students and had enough schools to support the random assignment of schools in each participating district to the five arms of the study.

Interested districts worked with the study team to identify schools that served low-income students and did not already use any of the four curricula identified for the study (or other similar comprehension curricula). By August 2006, participating districts and schools had been identified and participation agreements with districts obtained. A total of 10 districts and 89 schools agreed to participate. As expected—given the types of districts and schools being recruited—the participating districts and schools were statistically significantly different from schools and districts nationwide in several respects. They had higher poverty levels (63 percent of students in study districts were eligible for free or reduced-price lunch, compared to 40 percent of students in districts nationally), were larger (38,026 students per study district, compared to 3,153 students per district nationally), and were more urban than districts and schools nationally (70 percent of study districts were in urban areas, compared to 11 percent of districts nationally).

Collecting Data

Addressing the study questions required information about the curricula and how they were implemented, study participants, and students' performance outcomes. Information about teaching and implementation of the curricula was collected to support an examination of the fidelity of implementation to each curriculum design, the ways the curricula affected more general (non-curriculum-specific) teaching practices related to comprehension and vocabulary instruction, and the resources required to implement the curricula. Data on all three "levels" of study participants—schools, teachers, and students—were collected as a basis for describing their characteristics as they entered the study. Student outcomes were measured through assessments administered towards the end of the 2006-2007 school year. More information on the study's key data sources is provided below (see box for a summary).

Data Source	Time Collected	Description of Data
Classroom Observations (Developed by study team. See Appendix J for a copy of the instrument.)	January-April 2007	Observers documented the number of times they observed instructional practices related to vocabulary and comprehension instruction. In treatment classrooms, observers also documented whether the teachers adhered to the curriculum content and procedures prescribed by the developers.
Teacher Survey (Developed by study team. See Appendix J for a copy of the instrument.)	August-November 2006	This survey gathered data on teacher characteristics, experience, educational credentials, impressions about the culture in their school, and attitudes about student engagement, instructional strategies, and classroom management.
School Information Form (Developed by study team. See Appendix J for a copy of the instrument.)	April-June 2007	This form collected data on school characteristics such as enrollment, the percentage of students classified as English Language Learners, and the percentage of students eligible for free or reduced-price lunch.
Student Records (Developed by study team. See Appendix J for a copy of the instrument.)	May-October 2007	This form gathered data on student characteristics such as gender, date of birth, race, ethnicity, and eligibility for free or reduced-price lunch.
Test of Silent Contextual Reading Fluency (TOSCRF) (Hammill 2006)	August-October 2006	This assessment measured students' skills in word identification, word meaning, and sentence structure.
Passage Comprehension Subtest of the Group Reading Assessment and Diagnostic Evaluation (GRADE) (Williams 2001)	August-October 2006 (baseline), April-June 2007 (follow up)	This test assessed students' general reading comprehension skills.
Science and Social Studies Reading Comprehension Assessments (Educational Testing Service 2007a and 2007b)	April-June 2007 (follow up)	These assessments focused on students' reading comprehension of science and social studies text.

Information About Teaching and Implementation of the Curricula. Three data collection activities focused on teachers, teaching, and implementation of the four reading comprehension curricula. Two of these involved classroom observations, conducted in spring 2007 for two purposes. To support interpretation of the impact estimates, intervention-specific "fidelity" observations of classes taught by treatment group teachers were conducted to determine the extent to which the teachers adhered to the curriculum content and procedures prescribed by each developer. To describe more general teacher practices related to comprehension and vocabulary instruction (as opposed to practices linked to a specific intervention) and determine whether these practices were correlated with intervention impacts, "quality of instruction" observations were carried out in both treatment and control group classrooms to record the frequency with which teachers engaged in behaviors that research suggests are effective comprehension and vocabulary teaching practices. The third data collection activity that addressed the implementation of the curricula was a survey of developers on the cost of their curriculum to school districts.

To help summarize the large amount of "quality of instruction" observation data collected on general (non-intervention-specific) teaching practices related to comprehension and vocabulary instruction, the following three summary scales were created (for details on these scales, see Chapter II and Appendix F):[11]

- *Traditional Interaction.* This scale captures interactive teaching practices, primarily focused on vocabulary instruction and drawing inferences from text, that have been in use for many decades in American schools (Durkin 1978-1979; Brophy and Evertson 1976).

- *Reading Strategy Guidance.* This scale captures teachers' use of aspects of strategy instruction (such as using text structure and generating summaries to improve comprehension) to build students' comprehension ability.

- *Classroom Management and Student Engagement.* This scale captures teaching practices related to the management of student behavior and students' engagement.

Data on Teacher Characteristics. The Teacher Survey, conducted in early fall 2006, served three main purposes. First, it allowed the study team to describe the teachers participating in the study. Second, it was used to assess the similarity of treatment and control group teacher characteristics. Third, it made it possible to examine the relationship between teacher characteristics and impacts, including examining the relationship between impacts and school culture and teachers' ability to benefit from the professional development provided to treatment group teachers as part of the study.

The Teacher Survey data were used to create two scales for this third purpose (see Appendix F for details):

---

[11]These scales were used in two ways: (1) to describe teacher practices in the treatment and control groups and (2) to examine the nonexperimental relationship between impacts on student reading comprehension outcomes and these scale scores.

- ***School Professional Culture.*** The School Professional Culture scale is intended to capture conditions in schools that affect the quality of instruction (Consortium on Chicago School Research 1999; Carlisle 2003). The scale's 35 items were included in the Teacher Survey developed for this study. They reflect teachers' perceptions of the culture in their school, including relationships with colleagues, access to professional development, experiences with changes being implemented in their school, and leadership support in their school.

- ***Teacher Efficacy.*** The Teacher Efficacy scale is intended to capture teachers' ability to benefit from professional development (Sparks 1988; permission to use scale provided by Hoy and Woolfolk 1993). The scale's 12 items, included in the Teacher Survey developed for this study, ask about teachers' attitudes concerning student engagement, instructional strategies, and classroom management.

Data on School and Student Characteristics. The School Information Forms, collected at the end of the 2006-2007 school year, captured data on school characteristics, which were used to describe the study context, contribute school-level variables to the impact analysis, and examine the relationship between impacts and conditions in schools. At the end of the 2006-2007 school year, the study team also asked schools to provide records data on each student, including several stable items that could be used to describe students' baseline characteristics (such as gender, race, and ethnicity).

Data on Students' Baseline Achievement Levels. Two student assessments administered at the start of the 2006-2007 school year allowed the study team to characterize the achievement level of study students at baseline:

- ***Passage Comprehension subtest of the Group Reading Assessment and Diagnostic Evaluation (GRADE).*** This assessment, published by Pearson Learning Group, measures a student's ability to comprehend text passages (Williams 2001).

- ***Test of Silent Contextual Reading Fluency (TOSCRF).*** This assessment yields a score that reflects skills such as word identification, word meaning, and sentence structure, all of which are important skills for reading comprehension (Hammill et al. 2006).

Data on Student Outcomes. Data on student outcomes were collected from two sources at the end of the fifth-grade year (spring 2007). First, students were retested using the GRADE (Williams 2001). In addition, students were tested for comprehension of social studies and science informational text, using assessments specially developed by the Educational Testing Service (ETS) for the study (Educational Testing Service 2007a and 2007b). To reduce burden, half the students were randomly assigned to take the science test and half to take the social studies test.

Summary of Study Findings

The study's key findings focus on curriculum implementation and impacts on student achievement. The implementation analyses document treatment teachers' training and feelings of preparedness to implement the curricula, adherence to their assigned curriculum, and teaching practices observed among teachers in the treatment and control group classrooms. The impact analyses examine how student outcomes were affected by the curricula and how the impacts relate to conditions and practices in study schools and classrooms.

Implementation Findings. Five key findings emerged from the implementation analyses:

1. **During summer and early fall 2006, over 90 percent (91-100 percent) of treatment teachers were trained to use the curricula.** Ninety-one percent of Read for Real teachers, 96 percent of Reading for Knowledge teachers, and 100 percent of Project CRISS and ReadAbout teachers were trained in the use of the curricula.

2. **More than half of the teachers (56 to 80 percent) reported feeling very well prepared by the training to implement the curricula.** Fifty-six percent of Reading for Knowledge teachers, 69 percent of Project CRISS teachers, 72 percent of ReadAbout teachers, and 80 percent of Read for Real teachers reported that they felt very well prepared to implement their assigned curricula.

3. **At the time of the classroom observations in the spring, over 80 percent (81 to 91 percent) of treatment teachers reported using their assigned curriculum.** Eighty-one percent of Read for Real teachers, 83 percent of Reading for Knowledge teachers, 87 percent of ReadAbout teachers, and 91 percent of Project CRISS teachers reported using their assigned curriculum.

4. **Classroom observation data showed that teachers implemented 55 to 78 percent of the behaviors deemed important by the developers for implementing each curriculum.** ReadAbout and Project CRISS teachers implemented, on average, 71 and 78 percent of such behaviors, respectively. Reading for Knowledge teachers implemented 58 and 65 percent of the behaviors deemed important for the two types of instructional days that are part of the curriculum. Finally, Read for Real teachers implemented 55 and 71 percent of the behaviors deemed important for the two types of instructional days that are part of that curriculum.

5. **Two of three teacher practice scales were not statistically significantly different between the treatment and control groups.** For the purposes of describing teacher practices, the study team constructed scales summarizing teacher practices in three areas. There were no statistically significant differences in the Reading Strategy Guidance and Classroom Management and Student Engagement scales. Scores on the third scale, Traditional Interaction, were statistically significantly lower for the treatment group than the control group (effect size: -0.52).

Impact Findings. The effectiveness of the study curricula was gauged by experimental comparisons of reading comprehension test scores between students in treatment and control schools. Effects on test scores were estimated using a statistical model that accounts for clustering of students within schools, adjusts tests of statistical significance for the multiple

comparisons being made in the study, and includes covariates to increase statistical precision. The study's key impact findings, described below, were robust to a variety of sensitivity tests including variations in model specification, method of estimation, and method of adjusting for multiple comparisons.

In this report, two types of impacts are presented. First, impacts are presented for each intervention (for example, outcomes of students in ReadAbout schools are compared to outcomes of students in the control group). These impacts provide information on the effectiveness of each intervention, which may be helpful to readers considering implementing one of the interventions included in the study. Second, impacts are presented for the combined treatment group. In this analysis, the outcomes of students in all four intervention groups *combined* are compared to outcomes of students in the control group. These impacts provide information on the effectiveness of reading comprehension interventions more broadly (not the specific impacts of any one intervention). Impacts for the combined treatment group are presented for two main reasons. First, although the details of each intervention differ, the four interventions share common strategies for improving reading comprehension, so examining the interventions as a group is a reasonable approach to address the question of whether the use of these types of interventions, in general, improves comprehension. Second, examining the combined treatment group gives the study more power than looking at an individual treatment group.

The analysis of impacts was designed to answer two types of questions: (1) confirmatory (primary) questions about whether the reading comprehension interventions "work" and (2) exploratory (secondary) questions about for whom and under what conditions they might work. Answers to the confirmatory questions, all of which are supported by the experimental design and have a causal interpretation, indicate whether or not the interventions have the intended effect of improving reading comprehension. Answers to the second set of questions can help interpret the answers to confirmatory impact questions and guide future research on reading comprehension interventions. Answers to these exploratory questions do not always allow causal conclusions to be drawn about the impacts of the interventions for subgroups. A subgroup analysis that maintains the properties of random assignment allows causal conclusions about the impacts of the intervention for that subgroup to be drawn because it ensures that there are no systematic differences between subgroup members in the treatment and control groups. In this report, such subgroup analyses are those in which the subgroups are based on teacher, student, or school characteristics that could not have been influenced by the intervention, including teacher experience, students' prior test scores and English language learner status, and the schools' concentration of English language learners and students eligible for free or reduced-price lunch. A subgroup analysis that does *not* maintain the properties of random assignment does *not* allow causal conclusions about the impact of the intervention for that subgroup to be drawn because subgroup members in the treatment and control groups might differ systematically. In this report, such subgroup analyses are those in which the subgroups are based on teacher characteristics that could have been influenced by the intervention, including teachers' reported professional development participation, teaching efficacy, and professional culture in the school (all of which could be affected by the product-specific training teachers in the treatment group received during the summer before the intervention year).

Answers to Confirmatory Questions on Intervention Effectiveness. Figures 1 through 4 show observed score differences on the GRADE, ETS science comprehension assessment, ETS social studies comprehension assessment, and a composite score based on an average of the GRADE and ETS test scores. All differences are shown in effect size units, which (as noted above) allows for a comparison of results for tests scored in different units.

Figure 1: Effects of Reading Comprehension Curricula on GRADE Score

*Statistically different from the control group at the .05 level.

Figure 2: Effects of Reading Comprehension Curricula on Social Studies Reading Comprehension Assessment Score

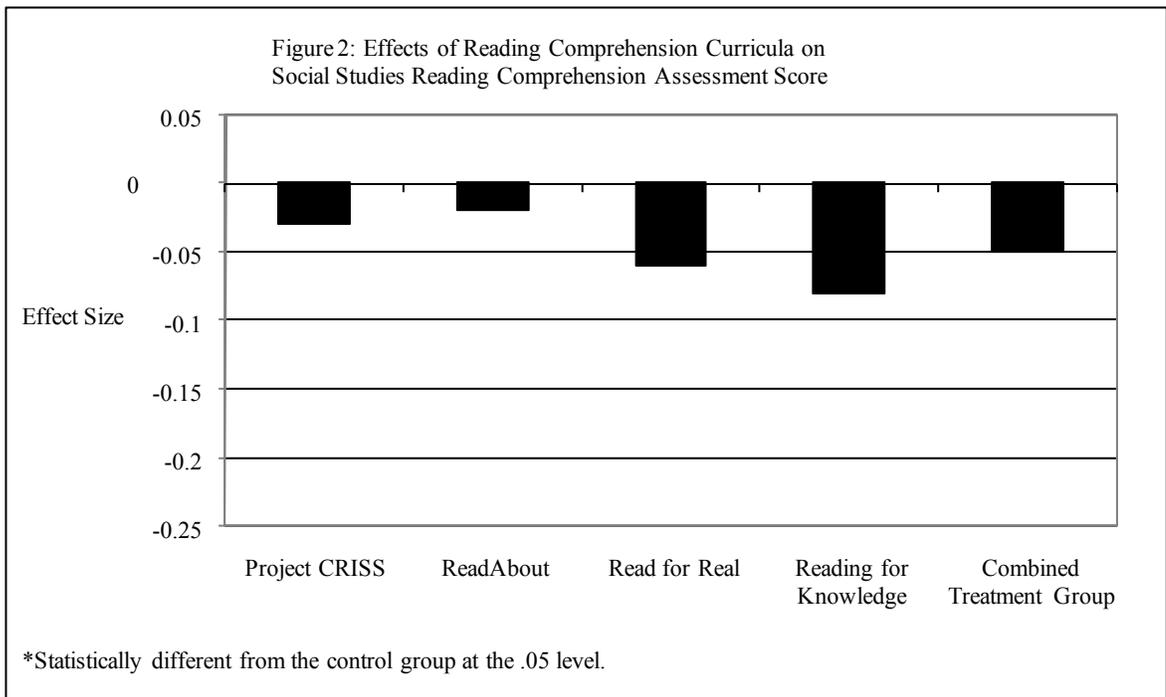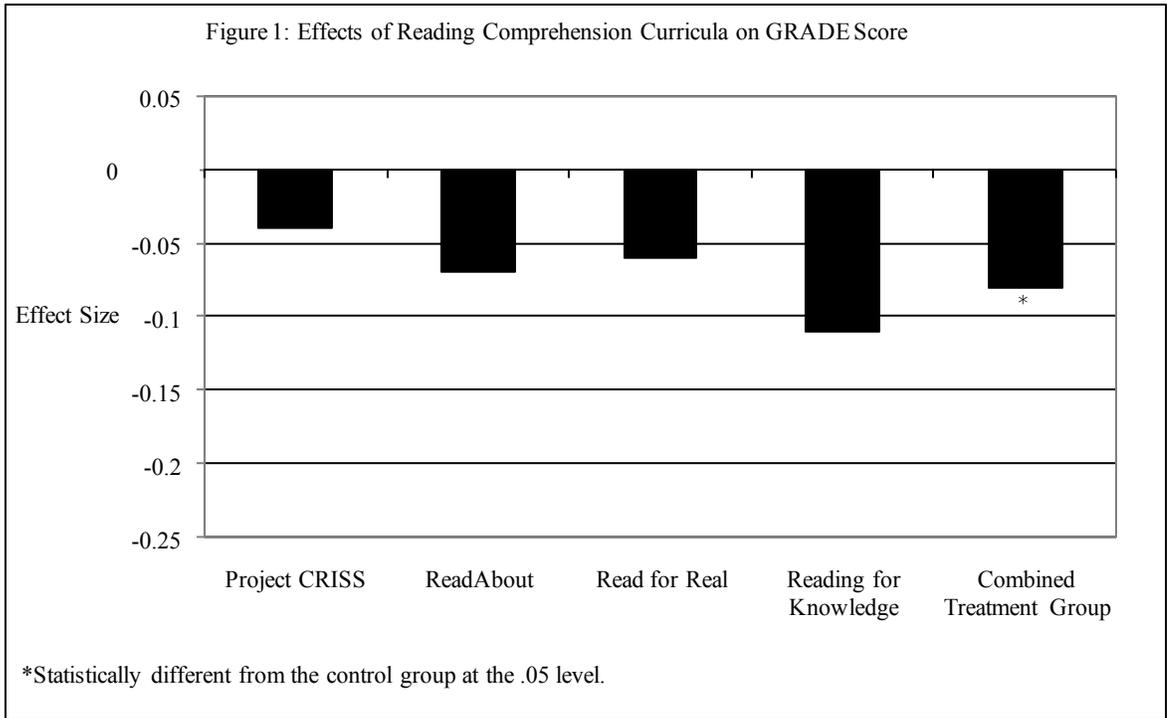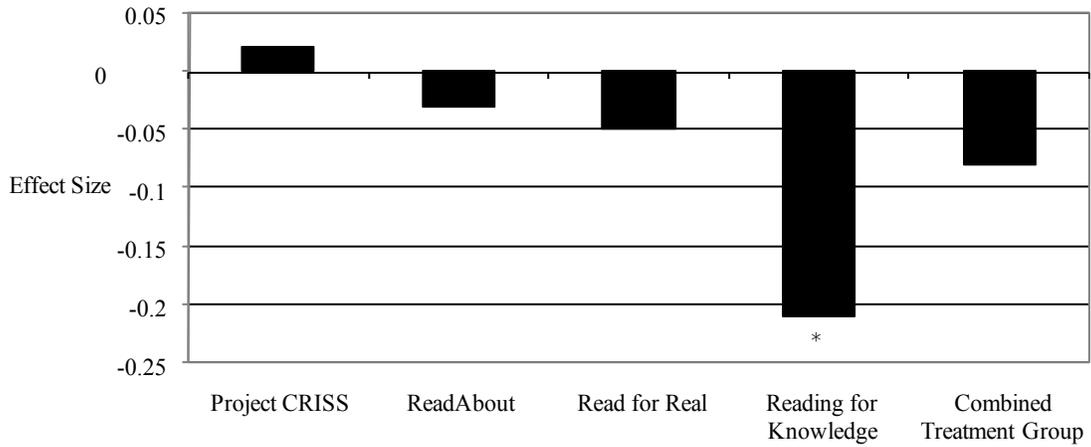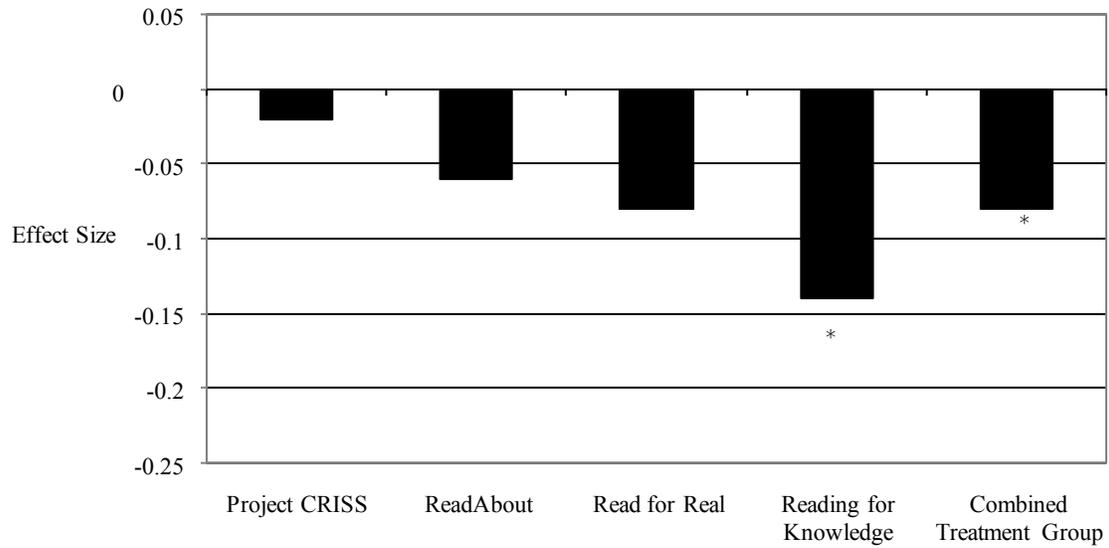*Statistically different from the control group at the .05 level.

Figure 3: Effects of Reading Comprehension Curricula on Science Reading Comprehension Assessment Score

*Statistically different from the control group at the .05 level.

Figure 4: Effects of Reading Comprehension Curricula on Composite Test Scores

*Statistically different from the control group at the .05 level.

Reading comprehension test scores were not statistically significantly higher in schools using the selected reading comprehension curricula than in control schools. In fact, students' reading comprehension test scores were statistically significantly lower in treatment schools than in control schools. The treatment group as a whole scored lower than the control group on the GRADE assessment (Figure 1, effect size: -0.08), and the Reading for Knowledge group scored lower than the control group on the ETS science comprehension assessment (Figure 3, effect size: -0.21). On the composite test score, the treatment group as a whole scored lower than the control group and the Reading for Knowledge group scored lower than the control group (Figure 4, effect sizes: -0.08 and -0.14, respectively).

Answers to Exploratory Questions on the Effectiveness of the Interventions for Subgroups of Students. The student subgroups examined were defined based on variables that can be observed by teachers, and thus could be used as a basis for targeting the interventions to specific students (for example, students with below-average fluency levels might respond better to a particular intervention). Similarly, the teacher and school subgroups examined were defined using characteristics that might be used by teachers and principals to target interventions to specific settings (for example, certain interventions might be more effective in schools with above-average concentrations of English language learners or they might be more effective for teachers with below-average years of experience).

Although reading comprehension test scores in treatment schools were statistically significantly lower than scores in control schools for subgroups of students defined by certain baseline characteristics of students, teachers, and schools, no clear pattern to these findings emerged. For the combined treatment group, negative impacts (treatment students scoring lower than control students) were observed for the following subgroups, all of which have a causal interpretation because the subgroups are defined in terms of characteristics that were measured at the beginning of the study's implementation year, and thus could not have been influenced by the intervention:

- students with above-average baseline fluency levels (effect sizes: -0.23 for the social studies comprehension test score, where above-average is defined as above the average of 100 for the national norm sample, and -0.14 for the social studies comprehension test score, where above-average is defined as above the sample median of 89),

- students with baseline comprehension levels in the bottom third of the sample (effect sizes: -0.08 and -0.09 for the GRADE and composite test scores, respectively),

- students of teachers with more than five years teaching experience (effect size: -0.09 for the composite test score),

- students in schools with an above-average concentration of students eligible for free or reduced-price lunch (effect size: -0.11 for the composite test score), and

- students in schools with a below-average concentration of English language learners (effect sizes: -0.15 for the composite test score and -0.19 for the difference in impacts [on the composite test score] between students in schools with below-average and above-average concentrations of English language learners).

For the subgroups that did not maintain the properties of random assignment because teachers in the treatment group might have been affected by the product-specific training they received in the summer before the intervention year, the study found:

- for the combined treatment group, negative impacts for students in schools with a below-average School Professional Culture scale score (effect size: -0.14 for the composite test score), and

- no statistically significant impacts for the subgroups based on teachers' past professional development or teaching efficacy.

For Reading for Knowledge, statistically significant negative impacts were observed for students whose teachers had 10 or more years of teaching experience (effect size: -0.36 for the science comprehension test score). Other characteristics examined were not statistically significantly related to impacts. These include students' English language learner status, teachers' Teacher Efficacy scale scores, and teachers' past reading professional development. Impacts for subgroups defined by the Teacher Efficacy scale, School Professional Culture scale, and teachers' professional development cannot be interpreted causally, because treatment group teachers received additional professional development prior to the administration of the Teacher Survey (which could have affected the teachers' responses to questions on the survey about their professional development, teaching efficacy, and professional culture in their schools).

Answers to Exploratory Questions on the Relationship Between Intervention Effects and Teacher Practices. The study team also examined the relationship between intervention effects and classroom practices. These relationships must be interpreted cautiously because the interventions may have affected the extent to which teachers engage in specific practices or the types of teachers who choose to engage in those practices. More specifically, because the research design did *not* randomly assign interventions to teachers with different levels of teacher practices, factors that led teachers to have a certain level of teacher practices could explain the observed correlations. As a result, treatment and control teachers who engage in teaching practices to the same degree may differ in unmeasurable ways.[12] Therefore, it is important to note that these estimates of the relationship between intervention effects and teacher practices do *not* allow causal conclusions to be drawn.

Keeping these caveats in mind, several statistically significant relationships between teacher practices and intervention effects were observed. Students in Reading for Knowledge classrooms whose teachers had below-average scores on the Reading Strategy Guidance scale had statistically significantly lower composite test scores than students in control group classrooms in which teachers had below-average Reading Strategy Guidance scale scores (effect size: -0.23). Students in Read for Real classrooms of teachers with Classroom Management scale scores below the sample median had statistically significantly lower scores than students in control group classrooms taught by teachers with Classroom Management scale scores below the sample

---

[12]If the intervention affected teacher practices, then that impact on teacher practices might explain the overall impact on student test scores. However, it is not possible to make causal statements about that relationship (causal statements would require a different study design than the one we used on this study, such as one in which teachers or schools were randomly assigned to implement the interventions to different degrees or amounts).

median (effect sizes: -0.23 for the composite test score and -0.35 for the social studies reading comprehension test score).[13] In both cases, these differences raise questions for further research, but—as noted above—the estimates do *not* provide experimental or causal evidence.

A second study report will use a second year of data to examine two further questions: (1) whether the curricula are more effective after teachers and schools have had more experience using them, and (2) whether the curricula have any lasting impacts on student outcomes. To address the first question, we enrolled a second cohort of fifth-grade students in study schools and will determine whether impacts over one school year for those students are more positive than the impacts reported in the first report for the first cohort. To address the second question, students from the first fifth-grade cohort are being tested again at the end of the 2007-2008 school year to assess whether the impact results observed in the first year persist or change after a second year.

---

[13]See Appendix Figures F.1A through F.3 for information on how the frequency of specific teacher practices corresponds to different scale scores.

REFERENCES

Brophy, J., and C. Evertson. *Learning from Teaching: A Developmental Perspective*. Boston, MA: Allyn and Bacon, 1976.

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

Chall, J. *Stages of Reading Development.* Fort Worth, TX: Harcourt-Brace, 1983.

Consortium on Chicago School Research. "Improving Chicago's Schools: The Teachers' Turn, 1999; Elementary School Teacher Survey, 1999." Chicago: CCSR, 1999. http://www.consortium-chicago.org.

Crawford, L.W., C.E. Martin, and M.M. Philbin. *Read for Real: Nonfiction Strategies for Reading Results*. Columbus, OH: Zaner-Bloser, 2005.

Durkin, D. "What Classroom Observations Reveal About Reading Comprehension Instruction." *Reading Research Quarterly,* vol. 14, no. 4, 1978-1979, pp. 481–533.

Educational Testing Service. *Science Reading Comprehension Assessment* (unpublished). Princeton, NJ: ETS, 2007a.

Educational Testing Service. *Social Studies Reading Comprehension Assessment* (unpublished). Princeton, NJ: ETS, 2007b.

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

Hammill, D., J. Wiederholt, and E. Allen. *Test of Silent Contextual Reading Fluency (TOSCRF), Examiner's Manual*. Austin, TX: PRO-ED, Inc., 2006.

Hart, B., and T.R. Risley. *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Brooks, 1995.

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

James-Burdumy, S., D. Myers, J. Deke, W. Mansfield, R. Gersten, J. Dimino, J. Dole, L. Liang, S. Vaughn, and M. Edmonds. "The National Evaluation of Reading Comprehension Interventions: Design Report." Final report submitted to the U.S. Department of Education. Princeton, NJ: Mathematica Policy Research, Inc., May 2006.

Madden, N.A., and V. Crenson. *Reading for Knowledge.* Baltimore, MD: Success for All Foundation, 2006.

National Institute of Child Health and Human Development. *Report of the National Reading Panel, Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction.* (NIH publication no. 00-4769.) Washington, DC: U.S. Government Printing Office, 2000.

Rosenshine, B., and C. Meister. "Reciprocal Teaching: A Review of the Research." *Review of Educational Research,* vol. 64, no. 4, 1994, pp. 479–530.

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.

Santa, C.M., L.T. Havens, and B.J. Valdes. *Project CRISS: Creating Independence through Student-Owned Strategies* (3rd ed.). Dubuque, IA: Kendall/Hunt Publishing, 2004.

Scholastic. *ReadAbout: The Personal Reading Coach for Every Student.* New York, NY: Scholastic, 2005.

Snow, C.E. *Reading for Understanding: Toward a Research and Development Program in Reading Comprehension.* Santa Monica, CA: RAND, 2002.

Snow, C.E., and G. Biancarosa. *Adolescent Literacy and the Achievement Gap: What Do We Know and Where Do We Go From Here?* New York: Carnegie Corporation of New York, 2003.

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

Williams, K.T. *Group Reading Assessment and Diagnostic Evaluation (GRADE) Technical Manual.* Circle Pines, MN: American Guidance Service, Inc., 2001.