NCEE 2011–4007 U.S. DEPARTMENT OF EDUCATION

An Experimental Study of the Project CRISS Reading Program on Grade 9 Reading Achievement in Rural High Schools

Final Report

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

REL NORTHWEST
Regional Educational Laboratory

An Experimental Study of the Project CRISS Reading Program on Grade 9 Reading Achievement in Rural High Schools

April 2011

Authors

Jim Kushman, Principal Investigator
Education Northwest

Makoto Hanita
Education Northwest

Jacqueline Raphael
Education Northwest

Project Officer

Ok-Choon Park
Institute of Education Sciences

NCEE 2011-4007
U.S. Department of Education

U.S. Department of Education

Arne Duncan

Secretary

Institute of Education Sciences

John Q. Easton

Director

National Center for Education Evaluation and Regional Assistance

Rebecca A. Maynard

Commissioner

April 2011

This report was prepared for the National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, under contract ED-06C0-0016 with Regional Educational Laboratory Northwest administered by Education Northwest.

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 report.

This report is available on the Institute of Education Sciences website at http://ncee.ed.gov and the Regional Educational Laboratory Program website at http://edlabs.ed.gov.

Alternate Formats 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 conflict of interest

None of the authors or staff members from Education Northwest involved in this study, and none of the members of Chesapeake Research involved in providing technical advice and selecting the random sample for this study, have financial interests that could be affected by the content of this report.*

*Contractors carrying out research and evaluation projects for IES frequently need to obtain expert advice and technical assistance from individuals and entities whose professional work may not be entirely independent of or separable from the tasks they are carrying out for the IES contractor. Contractors endeavor not to put such individuals or entities in positions in which they could bias the analysis and reporting of results, and their potential conflicts of interest are disclosed.

Contents

Figures

Tables

Summary

Students entering high school face many new academic challenges. One of the most important is their ability to read and understand more complex text in literature, mathematics, science, and social studies courses as they navigate through a rigorous high school curriculum. The Regional Educational Laboratory (REL) Northwest conducted a study to examine the effectiveness of a teacher professional development program called Project CRISS, which stands for Creating Independence through Student-owned Strategies. Through Project CRISS, high school teachers learn how to apply research-based learning principles and reading/writing strategies in all major subject or content areas using materials, training, and follow-up support provided by the developer. The ultimate goal of Project CRISS is to help students learn new ways to read and comprehend, practice reading and writing strategies in different classes, and eventually internalize and use successful reading and writing strategies independently, leading to improved reading comprehension.

Improving adolescent literacy is a concern both nationally and in the Northwest Region states (Alaska, Idaho, Montana, Oregon, and Washington). National statistics such as grade 8 results from the National Assessment of Educational Progress show that 32 percent of students entering high school are not proficient readers (U.S. Department of Education 2009). Since 2004, REL Northwest needs assessments have identified secondary school reading as a high-priority area for improvement.

There are two prior randomized studies of Project CRISS, which met What Works Clearinghouse evidence standards, addressing impact on reading comprehension. One experimental study conducted by the developer (Horsefall and Santa 1994 as cited in U.S. Department of Education 2010) found positive effects for grade 4–6 students while an independent experimental study (James-Burdumy et al. 2009 as cited in U.S. Department of Education 2010) found no effects on grade 5 students. The current experimental study of Project CRISS was conducted in high schools that typify the Northwest Region in order to provide local educators and policymakers with rigorous evidence of the program's effectiveness. This study also contributes to the research on Project CRISS and adolescent reading programs by examining their effectiveness in a new context.

The working theory behind Project CRISS is that, if the program is well implemented, high school students are exposed to a set of consistent and effective learning and reading comprehension strategies through English language arts, mathematics, science, and social studies core classes. Over a two-year period, all core content teachers in a high school are trained and encouraged to use a consistent set of reading comprehension strategies across the curriculum. As students are exposed to these strategies and supported in their learning, effective learning strategies become "student owned" or internalized by students. The Project CRISS treatment tested in this study consisted of a series of training and technical assistance visits provided by a certified national trainer plus expectations for follow-up activities by schools.

This study used random assignment at the school level to test the impact of Project CRISS on grade 9 student reading comprehension. The primary impact question addressed was:

- *What impact does Project CRISS have on the reading comprehension of grade 9 students in high schools in rural and town locales in Northwest Region states?*

One implementation question and one exploratory question were also addressed:

- *To what extent was Project CRISS implemented with fidelity in the treatment schools?*
- *Does the impact of Project CRISS on student reading comprehension differ for boys and girls?*

The study used a cluster randomized trial with a baseline measure to assess the effects of Project CRISS. Schools recruited for the study were high schools in rural or town locales across six western states that enrolled 250–1,000 students. The intent was to test Project CRISS on the smaller rural and town schools that typify the Northwest Region. Schools selected for the study pool were required to have no Project CRISS training for the prior five years in order to avoid control group schools that adopted this program. Across two waves or recruitment cohorts, a total of 52 schools were randomly assigned to a treatment condition (Project CRISS) or a control condition (business as usual). A blocking design was used in which random assignment occurred within blocks or strata defined by cohort (Cohort 1 or 2), state (six states participated), and poverty index (at or below the median poverty level of all schools in the regional population or above the median poverty level of all schools in the regional population). This helped ensure that groups would be equalized along these variables. Baseline equivalence measures showed no significant differences between treatment and control conditions with regard to selected student characteristics, pretest scores, or selected teacher characteristics.

Over two school years, the treatment consisted of 24 hours of formal teacher training plus an additional four to five days of on-site consultation and assistance by a certified Project CRISS trainer. Additionally, a school teacher was selected to serve as a local facilitator, who received additional training and coaching in order to become a district certified trainer. Local facilitators were generally highly experienced teachers with literacy backgrounds. They were expected to provide frequent, job-embedded professional development to help teachers implement the Project CRISS learning principles and instructional strategies. Actual implementation was compared with the developer description of full implementation to assess implementation fidelity.

Within each school, treatment impact was assessed using the Stanford Diagnostic Reading Test, Fourth Edition, Comprehension Subtest, which was administered to groups of grade 9 students by research team staff and a school test coordinator at each treatment and control school. Pre/post impact assessment occurred over a single school year (fall to spring) in the second year of program implementation, by which time teachers would conceivably be more fully using Project CRISS in their classrooms. The impact analysis used hierarchical linear modeling procedures. Due to sample attrition, the final analysis sample consisted of 49 schools and 4,959 students: 23 schools and 2,460 students in the treatment condition and 26 schools and 2,499 students in the control condition.

Addressing the implementation question, schools fully received and used the Project CRISS services provided by the parent company, Lifelong Learning, Inc. These materials and services were consistent with full implementation. Schools engaged in some follow-up activities prescribed by full implementation (such as regular local facilitator–sponsored Project CRISS meetings, individual consultations that included instructional observation and coaching, and principal walkthroughs). However, the self-reported level of implementing or attending these follow-up activities was less than prescribed in the full implementation model. In essence, full professional development services were provided, but implementation fell short of an ideal or full implementation defined by developer standards.

Regarding the impact question, the difference in student reading comprehension test scores between treatment and control conditions was not statistically significant. A sensitivity analysis was performed to determine whether explicit modeling of the effect size variability across the blocks used for random assignment would alter the substantive result, but the difference between treatment and control conditions remained nonsignificant. The exploratory analysis found no moderating effect of gender: the impact of Project CRISS on reading comprehension did not differ between male and female students.

1. Study background

Students entering high school face many new academic challenges. One of the most important is their ability to read and understand more complex text in literature, mathematics, science, and social studies courses as they navigate through a rigorous high school curriculum. Given the importance of helping high school students become proficient readers, the Regional Educational Laboratory (REL) Northwest conducted a study to examine the effectiveness of a teacher professional development program called Project CRISS, which stands for Creating Independence through Student-owned Strategies. Through Project CRISS, high school teachers learn how to apply research-based learning principles and reading/writing strategies in all major subject or content areas, using materials, training, and follow-up support provided by the developer. The ultimate goal of Project CRISS is to help students learn new ways to read and comprehend, practice reading and writing strategies in different classes, and eventually internalize and use successful reading and writing strategies independently, leading to improved reading comprehension. This report presents the findings of a randomized study of Project CRISS.

Need for the study

Students graduating from high school need strong literacy skills to prepare for higher education, family-wage careers, and life success in a faster, more technologically driven information age. Yet, results from the National Assessment of Educational Progress (NAEP) continue to show that a majority of students enter high school lacking the proficiency to read with full comprehension and understanding (U.S. Department of Education 2009). Between 2005—when this study was conceived—and the most recent 2009 results, performance in grade 8 reading improved only slightly on the NAEP. In 2009, the percentage of grade 8 students nationally at the proficient level was 32 percent, compared with 31 percent in both 2007 and 2005, a statistically significant difference (U.S. Department of Education 2009). A quarter (25 percent) of grade 8 students in 2009 was reading below the basic level, compared with 26 percent in 2007 and 27 percent in 2005 (U.S. Department of Education 2009). Again, this represents a small but statistically significant improvement. While it is encouraging that small improvements in adolescent reading skills have occurred since 2005, still only a third of students nationally are reading at a proficient level, and a quarter do not possess a basic level of reading as they enter high school.

As shown above, national statistics support the need for early high school programs that can improve reading proficiency. Likewise, regional needs surveys and focus groups conducted by the REL Northwest also identify secondary school reading as a high priority for improvement. A 2004 regional survey asked superintendents in the Northwest Region states (Alaska, Idaho, Montana, Oregon, and Washington) to rate priority education issues in their districts and schools. The results indicated that 80 percent of superintendents rated improving junior and senior high reading comprehension as needing more or much more effort (Barnett and Greenough 2004). Further, 79 percent of principals in high-poverty secondary schools rated improving junior and senior high school reading comprehension as needing more or much more improvement, compared with 62 percent of principals in high-poverty elementary schools

(Barnett and Greenough 2005). Similar findings were derived from needs-sensing state forums with broad stakeholder groups, including representatives from state education agencies, legislatures and governors' offices, local education agencies, and professional educator associations. A thematic analysis of these focus groups conducted by Gilmore Research Group in 2006 identified low test scores in secondary reading and writing as an ongoing concern, along with coordinating reading and writing instruction across the curriculum. One subtheme in the discussions was the difficulty of identifying what works to improve literacy at the secondary level and a perception of "conflicting sources of information to evaluate the merits of best practices" for improving adolescent literacy (Gilmore Research Group 2006, p. 58).

Given both the national data on reading proficiency and REL Northwest needs assessment results, Project CRISS was selected for study as a program with the potential to help Northwest Region students improve reading comprehension. Project CRISS includes specific reading and writing strategies for improving student comprehension of secondary school–level text. It is a reading program that was developed over two decades from the growing body of evidence on reading comprehension strategies that successful readers use. Carol M. Santa, Ph.D., is the founder and lead developer of Project CRISS. She developed the program with support from the U.S. Department of Education and the National Diffusion Network, beginning in the late 1970s, and has continued to refine the program based on emerging theory and research concerning reading, writing, and learning (Santa 2004). Santa served as the president of the International Reading Association from 1999 to 2000.

Another factor in selecting Project CRISS was its widespread use, despite the lack of solid evidence of effectiveness. The developer estimated that more than 1,000 elementary and secondary schools have received Project CRISS training throughout the United States and abroad, including large adoptions in Florida and Illinois when this study began (C.M. Santa, Project CRISS founder, personal communication, October 2006). The program had not undergone independent rigorous evaluation at the high school level at the initiation of this study. (The research on Project CRISS effectiveness is presented later in this chapter.) The purpose of this study was to evaluate the program's effectiveness in helping grade 9 students become more proficient readers, thereby increasing the likelihood of student success across core courses in high school that require significant reading comprehension.

It was beyond the scope of this study to test a logic model of Project CRISS's long-term effectiveness by tracking students longitudinally. The goal of this study was more modest in testing whether Project CRISS is an effective program for improving grade 9 reading comprehension as measured by a standardized test. Reflecting the Northwest Region's largely rural geography, the study focused on relatively small high schools in rural areas and towns. To the extent that there are many rural areas and small towns throughout the United States, this study can also have relevance beyond the Northwest Region states.

Project CRISS research base and conceptual framework

Project CRISS attempts to help teachers apply principles drawn from reading research, cognitive psychology, and social learning theory in order to help students become better readers. It also uses writing to improve reading comprehension. A brief summary of the research base and theory of action guiding Project CRISS is presented below.

Project CRISS began in 1979 as a single-district demonstration project in Kalispell, Montana, and was then disseminated more widely between 1985 and 1995 through the U.S. Department of Education's National Diffusion Network. The early Project CRISS research base was an integration of research-based practices from cognitive science (Bloom and Broder 1950; Bruner 1977; Simon 1979); reading comprehension research (Duke and Pearson 2002; Palinscar and Brown 1984); social learning research (Bandura 1977; Vygotsky 1978); and "transactional strategy instruction" methods in which teachers model learning strategies followed by student guided practice and application (Brown, Pressley, Van Meter, and Schuder 1996; Duffy et al. 1987; Duke and Pearson 2002). Concepts derived from these strands of research form the foundation of the CRISS philosophy and learning principles. In everyday terms, the main idea behind Project CRISS is that proficient readers are goal directed, monitor their own understanding, use specific strategies to solve reading problems, and are intentional in their approach to reading text. Research on readers who struggle with text supports the idea that explicit teaching of comprehension strategies increases knowledge retention (Brown, El-Dinary, Pressley, and Coy-Ogan 1995; Duke and Pearson 2002; Palinscar and Brown 1984; Paris, Lipson, and Wixson 1994; Pearson and Fielding 1991; Pressley 2002). The CRISS approach attempts to support students while effective reading comprehension strategies are learned, practiced, and then used independently.

More recent syntheses of research on adolescent reading provide support for the teaching philosophy and instructional methods embedded within Project CRISS. In *Reading Next: A Vision for Action and Research in Middle and High School Literacy* (Biancarosa and Snow 2004), a national panel of reading instruction scholars concluded that high school students need explicit instruction in the strategies that proficient readers use to achieve a full understanding of increasingly more challenging text. Project CRISS applies the idea from *Reading Next* that struggling adolescent readers need to learn these strategies through explicit classroom instruction, thus becoming "strategic readers." Project CRISS also embodies the *Reading Next* conclusion that effective instructional principles should be embedded in content, such as language arts teachers teaching from content-area texts and content teachers providing instruction and practice in reading and writing skills in their subject areas (Biancarosa and Snow 2004). Content-area teachers are taught how to use the Project CRISS strategies in English language arts, mathematics, science, and social studies classes.

A more recent analysis of the research on secondary school reading comprehension appears in an Institute of Education Sciences practice guide, *Improving Adolescent Reading: Effective Classroom Intervention Practices* (Kamil et al. 2008). This guide includes recommendations for improving adolescent reading and writing that mirror the findings of *Reading Next* and the practices of Project CRISS. The recommendations include the need to provide explicit instruction in vocabulary and comprehension (supported by strong research evidence) and opportunities for extended classroom discussion of the meaning and interpretation of text (supported by moderate evidence).

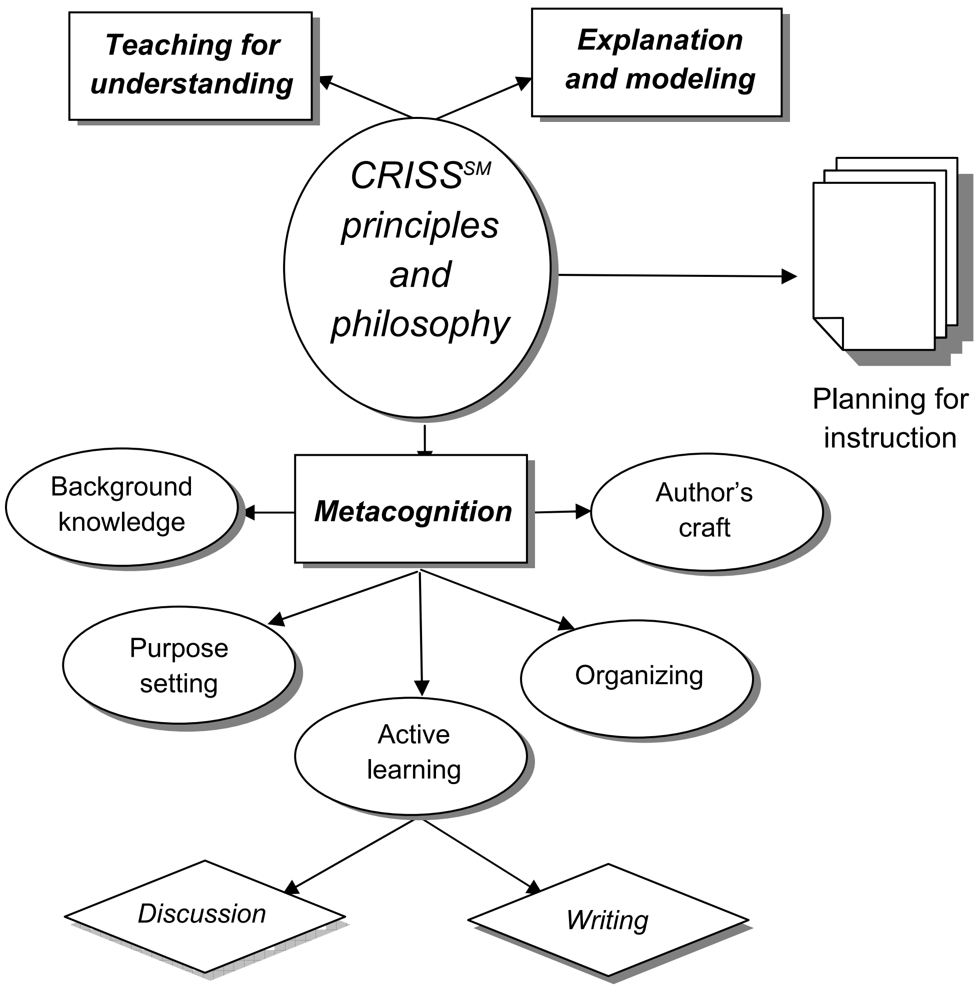
In summary, the working theory behind the Project CRISS treatment is that, if the program is well implemented, high school students are exposed to a set of consistent and effective learning and reading comprehension strategies through English language arts, mathematics, science, and social studies core classes. Over a two-year period, all core content teachers are trained and encouraged to use a consistent set of reading comprehension strategies across the curriculum. As students are exposed to these strategies and supported in their learning, effective learning strategies become student owned, or internalized by students.

Figure 1 presents the conceptual framework for Project CRISS, showing its three basic learning principles and multiple teaching and student learning practices. As figure 1 indicates, Project CRISS involves a fairly complex set of expectations for teachers and students that can be described as a teaching philosophy or approach involving many instructional strategies. As such, the successful implementation of Project CRISS requires time, practice, and opportunities for teachers to try out and discuss these strategies with other teachers and with a Project CRISS in-school facilitator or teacher who ideally has literacy experience.

There are three core learning principles at the heart of Project CRISS: teaching for understanding, explanation and modeling, and metacognition. Teaching for understanding involves the idea that understanding written text goes beyond knowing and being able to recall the information conveyed. Project CRISS helps teachers expand understanding by having students use the information in thought-demanding activities such as explaining, finding examples, producing evidence, generalizing, and representing the topic in new ways.

Explanation and modeling require that teachers at first take charge of the lesson by showing, explaining, and demonstrating content and techniques such as active reading that lead to improved comprehension of the text. As students begin to comprehend the material and understand their own processes for improving comprehension, the teacher acts more as a guide to help students learn and internalize effective reading strategies. The Project CRISS manual that forms the basis of training contains numerous examples in different subject areas to help teachers model learning strategies and provide students opportunities to reflect on the effectiveness of these strategies for their own understanding (Santa, Havens, and Valdes 2004).

Figure 1. Project CRISS conceptual framework



Source: Santa, Havens, and Valdes 2004.

Metacognition is viewed as the most central of the three principles. In everyday terms, this idea from cognitive psychology means being aware of one's own thinking process. The premise behind Project CRISS is that expert readers are strategic readers who know what kinds of behaviors help them comprehend what they are reading. Such proficient readers self-regulate their behaviors as they read to ensure more complete and deeper understanding of what they are reading. As figure 1 indicates, Project CRISS helps teachers use a range of instructional strategies to guide students in becoming more effective readers and reflecting on their own reading processes. These strategies include having students determine their own background or prior knowledge of the topic, setting a clear purpose so students determine what information they should pay attention to in the text, using active learning that involves both discussing and writing about what the text means, incorporating tools for organizing text such as graphic organizers to increase understanding, and helping students understand author craft or technique so that they can anticipate certain predictable elements in different kinds of text. In all, these strategies are intended to help students become strategic readers who monitor their own understanding and use effective techniques to increase their understanding of many different kinds of text across subjects (Santa, Havens, and Valdes 2004).

As figure 1 shows, planning for instruction is an important element of applying these principles and strategies. Project CRISS does not provide a ready-made teacher curriculum. It asks teachers to use their schools' existing curriculum, adding supplementary high-interest reading materials if needed, around which lesson plans are developed that apply the learning principles and associated strategies depicted in figure 1. In essence, Project CRISS requires a considerable amount of teacher application. Teachers are strongly encouraged to work in disciplinary or cross-disciplinary teams to develop ways of teaching that mirror the CRISS philosophy and approach resulting in a "CRISS Strategic Learning Plan" to guide instruction. For this reason, the model includes two years of teacher professional development plus training of an in-school teacher facilitator to help a school faculty fully integrate effective learning strategies into all core subject classes.

Prior research on Project CRISS

At the time this experimental study was conceived, there was prior evidence of Project CRISS effectiveness based on developer studies. The developer conducted several unpublished quasi-experimental evaluation studies that are reported in summary form on the Project CRISS website (http://www.projectcriss.com/criss_research.php). These early study summaries lack sufficient methodological detail and are not reviewed here.

More recently, the What Works Clearinghouse (WWC) at the Institute of Education Sciences conducted a thorough search of studies involving the Project CRISS professional development program. The WWC published the results of two randomized studies that fell within the review protocol for adolescent literacy studies and that met WWC evidence standards (U.S. Department of Education 2010). The first was a developer random-assignment study of elementary and secondary school students across rural, urban working class, and suburban settings in three different states (Horsefall and Santa 1994, as reported by U.S. Department of Education 2010). The WWC based its effectiveness ratings for this study on 120 grade 4 and 6 students attending six Project CRISS classrooms, compared with 111 students attending six control classrooms. (Effectiveness for the grade 8 and 11 samples could not be determined because these studies did not meet WWC standards.) As in the Project CRISS quasi-experimental studies cited earlier, the outcome measure was a pre/post, staff-developed, free-recall comprehension test on a subject-appropriate reading segment. The posttest was taken one semester or approximately 18 weeks after program participation. The study found a statistically significant positive effect for students in treatment classrooms.

The second evaluation was an independent experimental study of Project CRISS and three other reading comprehension programs of grade 5 students throughout the United States, including a control group (James-Burdumy et al. 2009, as reported by U.S. Department of Education 2010). The WWC based its effectiveness ratings on 1,155 grade 5 students attending 17 Project CRISS schools, compared with 1,183 students attending 21 control schools. The outcome measure was a passage comprehension subtest of the Group Reading Assessment and

Diagnostic Evaluation of either a science or social studies text, measured after nine months of student participation. This study did not find any statistically significant effects of Project CRISS.

Based on these two randomized studies, the WWC concluded that Project CRISS had potentially positive effects on the outcome domain of reading comprehension.

Need for an experimental study

The REL Northwest conducted this randomized control trial on Project CRISS in Northwest Region high schools to meet a strong regional need for programs that improve adolescent reading. Built on reading and cognitive sciences research, Project CRISS has the potential to help grade 9 students become more proficient readers as they enter high school and are expected to master more challenging text. Prior evidence of Project CRISS's effectiveness has been mixed. The model was classified as having potentially positive effects by the WWC based on two randomized studies, although there were different results in different settings and different outcome measures across the two studies. The findings of the current REL Northwest study will help Northwest Region educators make more informed decisions about the program's effectiveness, particularly in contexts similar to their own. The results will also contribute to a body of research on Project CRISS and adolescent reading programs by adding another context and outcome measure on which to judge effectiveness.

Organization of the report

Chapter 2 presents the study methods, beginning with the research questions and then describing the experimental design, measures, and analysis procedures used. Chapter 3 covers Project CRISS implementation in the schools randomly assigned to the treatment condition, allowing the reader to judge how faithfully Project CRISS was carried out in comparison to its design. Chapter 4 presents the main impact analysis. Chapter 5 analyzes the exploratory research question. And chapter 6 presents a summary of findings and study limitations. Appendixes follow with important details and technical information not fully described in the text.

2. Study design and methods

This study used random assignment at the school level to test the impact of the Project CRISS treatment on grade 9 student reading comprehension. The study addressed one primary impact question, as well as one implementation question and one exploratory question. The questions are presented below and followed by the overall study design, sampling and recruitment methods, data collection procedures, and data analysis methods.

Research questions

The primary impact question addressed by this study was:

- *What impact does Project CRISS have on the reading comprehension of grade 9 students in high schools in rural and town locales in Northwest Region states?*

One implementation question and one exploratory question were also addressed:

- *To what extent was Project CRISS implemented with fidelity in the treatment schools?*
- *Does the impact of Project CRISS on student reading comprehension differ for boys and girls?*

Study design

This study used a cluster randomized trial with a baseline measure to assess the effects of Project CRISS. The trial consisted of two levels of clustering. Students were nested within a school (student at level 1, school at level 2) and not within a teacher or a classroom. This is because Project CRISS is a school-level program in which each high school student was exposed to the CRISS treatment from multiple teachers across core subject classes. Outcome measures were taken on two occasions: at baseline and posttreatment. The unit of assignment was the school, whereas the primary unit of analysis was student. The outcome measure was a standardized test of reading comprehension, the Stanford Diagnostic Reading Test, Fourth Edition, Comprehension Subtest.

The overall design for student impact is represented as follows:

Treatment	*R (School)*	O_b X O
Control	*R (School)*	O_b O

where *R (School)* stands for randomization by the school cluster; O stands for the observation, with O_b meaning baseline observation; and X stands for the treatment to the student participating in the study, which occurs during the second year of teacher training.

In the treatment schools, teachers received intensive Project CRISS training during the first year and follow-up training and coaching during the second year. It was expected that in the second treatment year teachers fully understood the principles and mechanics of Project CRISS strategies and were able to use and adapt them in their classrooms. During the second treatment year, grade 9 students were tested pre/post from fall to spring on reading comprehension (as represented by O_b and O). Two waves or cohorts of grade 9 students were studied; half of the treatment and control schools began the project in the 2007/08 school year and the other half in 2008/09 when full recruitment was completed.

Descriptive data on implementation of Project CRISS were also collected in the treatment schools. Several additional descriptive measures of teacher, principal, and school characteristics were collected in treatment and control schools to assess baseline equivalence of treatment and control conditions. Control schools were offered the Project CRISS treatment on a two-year delayed schedule as an upfront inducement for schools to participate in a randomized study.

Sample size and recruitment methods

At the outset of the study, a power analysis was conducted to determine adequate sample size for the study. This was followed by a recruitment process to obtain schools from the Northwest Region states and later from contiguous states, in order to obtain a sample of schools for study from rural and town locales.

Power analysis

An initial power analysis was performed for detecting a main effect of treatment on the student outcome. The goal of the power analysis was to estimate the necessary number of schools to sample in order to maintain the power of 0.80 for a minimum detectable effect size in the range of $\delta = 0.10$–0.25. The initial plan was to recruit up to 60 schools if possible (30 treatment and 30 control) to achieve an estimated minimum detectable effect size of $\delta = 0.19$. A total of 52 schools were recruited for the study after two rounds of recruitment that exhausted the population of schools that fit the sampling criteria. At the start of the study, the minimum detectable effect size for the recruitment sample of 52 schools that signed up for the study (26 treatment and 26 control) was recalculated; the resulting estimated minimum detectable effect size was $\delta = 0.21$.

A retrospective power analysis was performed on the final intention-to-treat analysis sample of 49 schools (23 treatment and 26 control) and 4,959 students (2,460 treatment and 2,499 control). The resulting minimum detectable effect size was $\delta = 0.25$, within the range originally set. Appendix A contains details of the power analysis.

Target population and recruitment

The population of interest was students in rural and town high schools in the Northwest Region states plus other contiguous states in order to expand the recruitment pool. Eligible schools were those with 250–1,000 students, located in rural areas or towns. This lower limit of school size is based on several considerations. Schools with fewer than 250 students can have too few teachers within a discipline to collaborate in the implementation of CRISS strategies (a program design feature) and too few grade 9 students to reliably estimate a school mean. While there are a few schools in large towns with more than 1,000 students, very large high schools tax the limits of the program in that the developer sets a training cap of 30 teachers at a time in order to cover the material and use interactive training with small group practice. More important, the size restrictions fit the intent of the study to examine Project CRISS in typical settings in the region predominated by rural areas and towns with smaller high schools.

Four of the five states in the Northwest Region served as the population of schools at the initiation of the study: Idaho, Montana, Oregon, and Washington. Alaska was not included because most nonurban high schools in the state fall below the 250 student minimum. Rural or town schools in Alaska are predominantly very small isolated schools in Alaska Native villages accessible only by airplane. Village school populations are primarily 100 percent Alaska Native students who are schooled in very small grade K–12 configurations with mixed grades and perhaps one or two teachers at the high school level. The remoteness, travel costs, and distinctive education context of high school education in Alaska factored into the decision to exclude the state.

Recruitment occurred in two rounds because a sufficient number of schools were not achieved after the first round. In the first round (Cohort 1), National Center for Education Statistics locale and sublocale codes were used to identify schools in small towns, rural areas in metropolitan counties, and rural nonmetropolitan areas in the four Northwest Region states (table 1). In the second round (Cohort 2), two additional locale codes—large town and urban fringe—were added (see table 1). This increased the sampling pool in the original regional states. Additionally, all five locale codes were applied in the contiguous states of Nevada, Utah, and Wyoming and in northern areas of California (which is geographically similar to southern Oregon). The sampling pool had to be increased to achieve a desired sample of at least 50 schools. The two cohorts combined included 353 schools in the population, of which 52 agreed to all conditions of the study—or about one school in seven.

Table 1. Distribution of eligible schools by cohort

Cohort and state	Number of schools
Cohort 1: small towns, rural areas in metropolitan counties, and rural nonmetropolitan areas in the Northwest Region states	161
Idaho	31
Montana	30
Oregon	47
Washington	53
Cohort 2: large town and urban fringe areas in Northwest Region states and all five locale codes for contiguous states	192
Northwest Region states (new locale codes only)	37
Wyoming	56
Nevada	32
Utah	26
Northern California	41

Note: Eligible schools included comprehensive high schools serving grades 9–12 with 250–1,000 students.

Source: Authors' analysis of recruitment data.

Recruitment procedures

Research team members contacted all eligible schools in the population of rural and town schools of a certain size (as described above) in an attempt to recruit 50–60 schools for the study. Participation in the study was voluntary. Schools in the final sample are self-selected treatment subjects who agreed to participate in Project CRISS and collect data for the study.

In the first recruitment round, research staff members were used to recruit schools. In the second round, three former school principals were used in the hope that they would have more credibility in selling Project CRISS and the study. These were highly experienced, respected, and articulate former principals with many contacts throughout the region. They were trained by research staff to conduct the recruiting. More onsite visits were conducted in the second year because these recruiters had more flexible schedules. Recruiters and research staff interacted frequently during the recruitment process via regular meetings and email to ensure a smooth process. The recruitment in the two rounds resulted in a total sample of 52 schools (26 treatment and 26 control schools) in six states: California (northern portions), Idaho, Montana, Oregon, Washington, and Wyoming. No schools from Nevada or Utah chose to participate. The first recruitment process yielded 23 schools (Cohort 1) and the second 29 schools (Cohort 2). Cohort 1 schools entered the study in fall 2007, and Cohort 2 schools entered one year later in fall 2008.

Several motivational factors helped the recruitment effort. First, Education Northwest's experience and reputation in the Northwest Region as a service organization enabled easy contact with schools. The major incentive was that both treatment and control schools were offered a substantial teacher professional development program through their participation in the study. Schools had to agree to random assignment, with control schools receiving the same services on a two-year delay. Schools also received some additional resources from the study to help offset teacher time required for data collection. In the recruitment, Project CRISS was presented as a value-added program that could fit into a larger comprehensive reform or reading program and that could be implemented without purchasing additional curriculum materials.

Random assignment and sample attrition

Schools agreeing to the study were blocked by state and poverty level within each cohort before random assignment. Although randomization is the best method for equalizing treatment and control groups on preselection variables, there is still a small probability of a "bad draw," in which the two groups of school clusters are not balanced. Therefore, a matrix of 24 blocking cells was created, defined by the combination of two cohorts (Cohort 1 and Cohort 2), six states (California, Idaho, Montana, Oregon, Washington, and Wyoming), and two levels of district poverty (at or below the median poverty level of all schools in the regional population—15 percent—and above the median poverty level). This set of 2 x 6 x 2 = 24 blocking cells had seven empty cells because not all six states were represented in each of the two cohorts. Chesapeake Research Associates, the methodological consultant on the study, conducted the random assignment. Random assignment (within each block or strata) occurred after all schools had agreed to participate in the study. Chesapeake had no prior relationship with any of the schools under study.

The main strategy for retaining the sample was a memorandum of understanding with the school principal and a formal contract with the district ensuring that schools understood their obligations to collect data in exchange for services with monetary value. The memorandum of understanding and contract specified all study conditions. Prior to the memorandum of understanding, school principals were provided explicit timelines, checklists, and reminders about data collection procedures, training schedules, and expectations. Before schools were brought into the study, research staff talked personally with school principals to ensure they understood all the conditions of the study.

Principals were asked to fully consult with teachers, either through a vote or consensus decision process, to achieve the necessary buy-in so that sample attrition would be reduced once the study was under way. The services were offered to all teachers of core content classes (English/language arts, mathematics, science, and social studies) as a strong incentive to participate, even though data collection occurred only on students in grade 9. All schools agreed to have their full core faculty trained, creating a sense of school purpose around the program and decreasing the likelihood of school dropouts during the study.

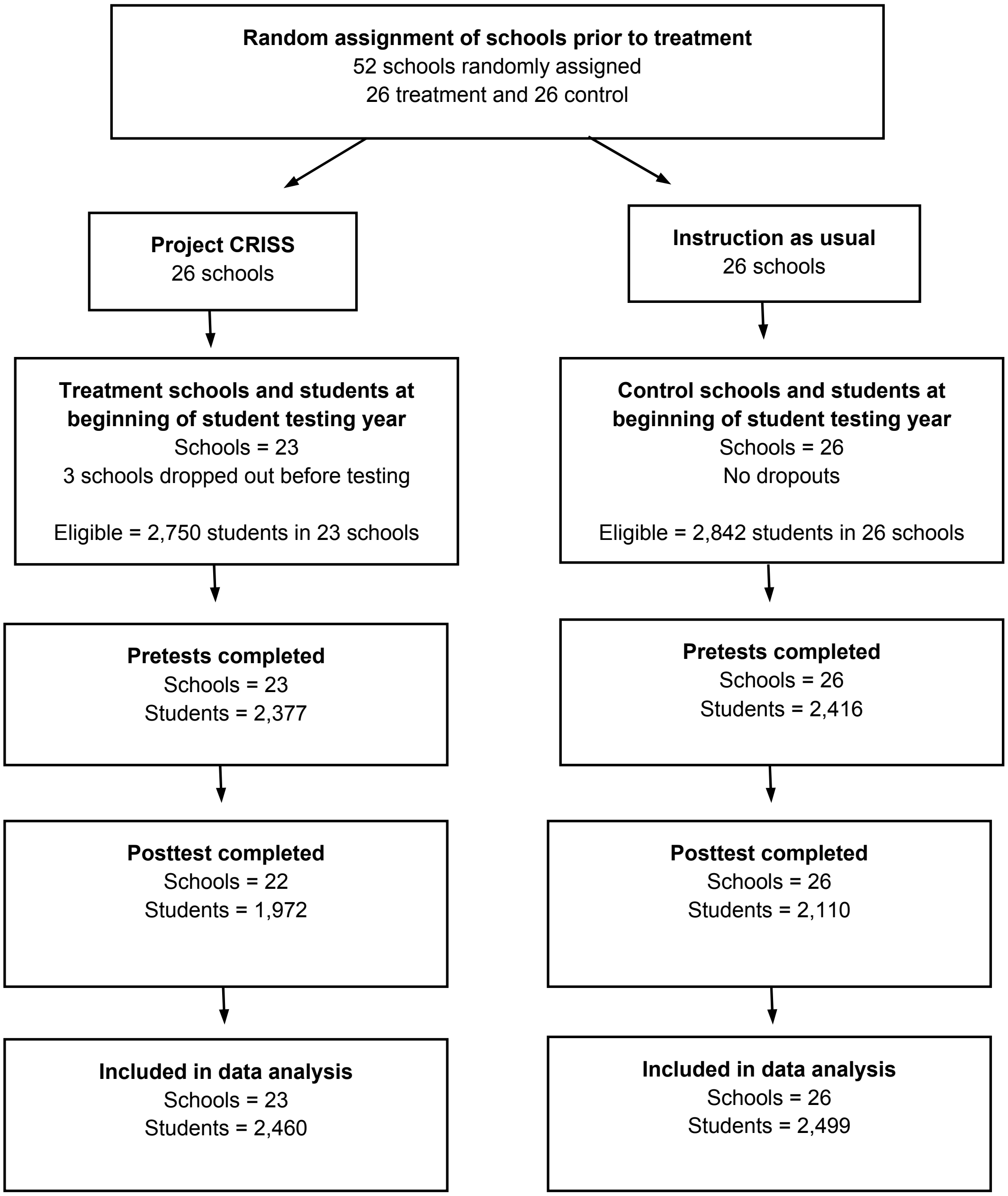
Schools were contacted regularly throughout the course of the study (announcements, emails, phone calls) to ensure that principals and teachers knew what to expect next and could ask questions or express concerns. This was viewed as relationship building with important clients who gave the privilege of conducting the study in their school. Principals were notified early on, and again during the summer before student data collection, that parents could opt out of the study. Passive consent forms were provided, and procedures were suggested for schools to use with these forms for their incoming freshmen. Passive consent was approved by the Institutional Review Board for the study because procedures were established such that no personally identifiable student information would leave the school.

Despite these efforts to retain the full sample, 3 of the 52 schools dropped out of the data collection and the program after randomization; all three were in the treatment condition. Repeated efforts failed to convince these schools to participate in the student data collection. (As with the other schools, they were offered cash compensation of $500 for their staff time to help collect data for the study.)

The reasons for dropping out that the three schools reported to the research team included a change in the lead literacy teacher and/or principal after the MOU had been signed in spring but before implementation began in fall, and a desire to focus on other more pressing professional development needs around school climate rather than reading. The three schools that dropped out of the program and data collection were found to be statistically similar to the remaining 49 schools regarding available data on school size (student enrollment) and district poverty level.[1] However, it is important to note that all three dropout schools were from the treatment group and dropped out after random assignment, indicating the possibility of response-to-assignment attrition.

The final intent-to-treat sample consisted of 49 schools, a loss of 3 schools or about 6 percent. Figure 2 presents a chart indicating sample attrition at various phases of the study for both schools and students.

Figure 2. Sample size at various phases of the study



Note: Grade 9 students were tested during the second year of treatment implementation. At the time of random assignment, the number of eligible students was not known. Because the three treatment schools that dropped out did not cooperate in data collection, the number of eligible students in the three schools is not known. Testing occurred on a single day on which some students were absent. Imputation of missing data was used for students who took a pretest but missed the posttest and for students who were known to be absent for the pretest and took a posttest, resulting in the final numbers of students included in the analysis.

Source: Authors' analysis of student data files.

Baseline equivalence of treatment and control groups

Through random assignment, treatment and control groups are assumed to be statistically equivalent on important variables that can impact the outcome measure, differing only on whether they were exposed to the treatment. However, it is still possible for treatment and control groups to differ in important ways that can affect the outcome measure and bias results. To reduce the likelihood of this happening, the blocking procedure described earlier was used prior to random assignment. Nevertheless, data were collected to assess the statistical equivalence of the final intent-to-treat sample of 23 treatment schools and 26 control schools, along key available variables that could bias the results.

Table 2 presents baseline equivalence based on grade 9 class size (an index of school size), measured as students eligible for the pretest; percentage of White students (who make up the majority population in the schools) as an index of race; percentage of female students as a gender index; pretest score equivalence; and various grade 9 teacher characteristics from a teacher questionnaire. No statistically significant differences ($p < .05$) were found.

Table 2. Baseline characteristics and equivalence of treatment and control schools

Characteristic	Treatment schools/teachers	Control schools/teachers	Difference	Test statistic
Number students eligible for pretest	Mean = 123.1 Standard deviation = 48.4	Mean = 106.2 Standard deviation = 44.8	16.9	$t = 1.27$ $p = 0.21$
Percentage of White students	Mean = 79.4 Standard deviation = 20.6	Mean = 78.4 Standard deviation = 21.3	1.0	$t = 0.17$ $p = 0.87$
Percentage of female students	Mean = 49.8 Standard deviation = 4.1	Mean = 48.4 Standard deviation = 5.2	1.4	$t = 1.01$ $p = 0.32$
Percentage of families below poverty level	Mean = 14.3 Standard deviation = 4.9	Mean = 15.4 Standard deviation = 6.3	–1.1	$t = -0.67$ $p = 0.50$
Mean pretest scale score, reading comprehension test used in study	Mean = 687.9 Standard deviation = 11.3	Mean = 681.3 Standard deviation = 19.2	6.6	$t = 1.42$ $p = 0.16$
Mean years teaching middle or high school	Mean = 14.5 Standard deviation = 10.4	Mean = 13.1 Standard deviation = 10.0	1.4	$t = 1.21$ $p = 0.23$
Mean years in present school	Mean = 9.4 Standard deviation = 8.8	Mean = 8.1 Standard deviation = 8.3	1.3	$t = 1.31$ $p = 0.19$
Percent of teachers with advanced degree	Percent = 56	Percent = 50	6.0	$X^2 = 1.16$ $p = 0.28$

Note: The first five characteristics—number of students, White students, female students, poverty, and pretest scores—are mean school values (n = 23 treatment schools, n = 26 control schools). The school poverty level is actually represented by U.S. Census percentage of families at or below the poverty level in the district census area. The teacher characteristics—years teaching, years in present school, advanced degree—are analyzed at the teacher level with n = 163 teachers in the treatment schools and n = 147 in the control schools. Student and teacher data are for grade 9 only. Listwise deletion of missing data was used, resulting in a loss of three or fewer cases for some analyses.

Source: Authors' analysis based on data described in text, National Center for Education Statistics Common Core of Data.

Data collection instruments and procedures

A variety of data collection instruments were used throughout the course of the study to address the main impact research question and the implementation question, as described below.

Implementation assessment

Table 3 shows the key features of Project CRISS implementation and data sources used to assess project implementation in the treatment schools.

Table 3. Data sources for assessing Project CRISS implementation

Prescribed element of CRISS treatment	Data source
Training and technical assistance provided to teachers	Documentation of training/technical assistance visits to schools; teacher attendance figures for Level I training
Selection of a qualified local facilitator	Documentation of local facilitator assignment, local facilitator questionnaire on background characteristics
Local facilitator Level II training (qualification for district certification)	Documentation of local facilitator attendance at a Level II summer institute
Monthly follow-up activities by local facilitator	Local facilitator monthly logs during the two years
Teacher participation in follow-up activities	Teacher questionnaire

Source: Authors' compilation of data sources used.

Local facilitator activity log. Because the CRISS-designated local facilitator is crucial to the follow-up support after the initial Level I training, a local facilitator monthly activity log was developed, and local facilitators were asked to complete the online questionnaire once a month. This questionnaire took no more than five minutes to complete. The log asked about local facilitator activities specified in the model that are intended to support teachers in the implementation of Project CRISS. The first log also asked local facilitators to provide some basic background information. Follow-up phone contact was made to local facilitators who were not consistently completing the log. Overall, local facilitators completed their logs 92 percent of the months required in year 1 of implementation and 87 percent of the months required in year 2.

Teacher questionnaire. A teacher questionnaire was administered to all grade 9 teachers in the treatment schools and included background information and several questions about participation in key Project CRISS follow-up activities prescribed by the model. To reduce the data collection burden, only grade 9 core subject teachers were administered the questionnaire because the study assessed impact on grade 9 students only. A similar questionnaire was used in the control schools to gather the teacher background used to establish baseline equivalence. Questionnaires were administered to teachers in the spring of implementation years 1 and 2. The teacher questionnaire used in treatment schools is presented in appendix B.

The data sources described above were used to address the implementation research question: to what extent was Project CRISS implemented with fidelity in the treatment schools? Descriptive results from the 23 treatment schools in the intent-to-treat analysis sample are presented in chapter 3, where comparisons are made between what occurred in the treatment schools and full CRISS implementation. The full implementation model and how it was derived is discussed in chapter 3 along with the results.

Implementation data are limited to the reported activities of Project CRISS national trainers, who provided training and technical assistance, and the self-reported activities of local facilitators and teachers with regard to their participation in prescribed follow-up activities. The extent to which teaching methods changed in the classroom to reflect Project CRISS literacy activities was not assessed in either the treatment or control schools.[2]

Impact assessment

The outcome variable was student reading comprehension. Because this study included schools from multiple states, a common reading comprehension instrument was used across all treatment and control schools: the Stanford Diagnostic Reading Test, Fourth Edition, a validated, group-administered, norm- and criterion-referenced reading test designed to pinpoint student strengths and weakness in several aspects of reading. To reduce burden on the schools, only the comprehension subtest was administered, which has a test time of 50 minutes plus student orientation. The test is machine scored by the publisher, Pearson Testing. The study used reading comprehension scale scores for the impact analysis. (See Karlsen and Gardner 1996, for the technical manual and student norms for the Stanford Diagnostic Reading Test, Fourth Edition.)

Because the Stanford Diagnostic Reading Test was not a district- or state-required test, it was administered by a small cadre of research team test administrators who were trained in test administration and other data collection required at the school. This was done to ensure complete testing and quality control over the test administration. A test administrator worked in concert with each high school's regular test coordinator to plan and administer the test.

A student roster was prepared for each school that included a space for student name, unique research identification number (sequential number plus school code), gender, race/ethnicity, and absenteeism. The roster with student identification numbers was sent to the school two weeks before the testing date, and the school test coordinator was asked to complete the names and background information for all grade 9 students. A research team test administrator arrived one day before the scheduled pretest to ensure the roster was complete and to help the school coordinator prepare for the testing. The original roster with student names was kept by the test coordinator (plus one copy for safekeeping in the principal's office). The copy returned to the Regional Educational Laboratory (REL) Northwest identified students only by number and included student characteristics.

At the posttest, the school list with names was retrieved so that pretest student numbers could be matched to the same posttest numbers when the research team test coordinator arrived for the posttest. Student names were placed on removable notes that were placed on answer sheets with a prebubbled student number so that tests would go to the correct students and the names could be removed from the answer sheets. This procedure allowed for pretest and postmatching of individual student scores without student-identifiable information leaving the school.

Because no identifiable student information left the school, passive parent consent was used. Passive consent forms were sent to each school during the summer prior to the pretest so that they could be distributed to all parents or guardians of incoming grade 9 students. Students whose parents completed a nonconsent form were not tested (a total of 82 students across all treatment and control schools). In addition, other students not tested from the eligible list of grade 9 students were students with Individualized Education Programs that required special accommodations for testing (such as reading items) and English language learner students with a state English language learner test score indicating that they were nonproficient in English. Otherwise, all grade 9 students present on the specified day of testing were administered the test in groups that were proctored by the research team test administrator and school test coordinator. The test administrator arrived one day early to help with the completion of roster forms and test preparation and remained until all tests were packaged and labeled for shipment back to the REL Northwest. Pretests occurred at the beginning of the school year between September and early October. Posttests occurred in the spring between mid-March and early May. Given the number of schools and wide geographic locations, not all schools could be tested at the same time.

Impact analysis methods

Preliminary analyses to clean data and to assess missing data were conducted, including examining out-of-range values, plausible means, standard deviations, and univariate outliers; removing cases with missing or inaccurate identification numbers that could not be tagged to a school (these were rare and attributed to data entry errors on the bubbled answer sheets); and determining the amount, pattern, and distribution of missing data. Missing data procedures are described later in this chapter.

The reading comprehension scale scores came from students attending 49 schools in the final analysis sample, which were randomly assigned to the treatment or the control condition within each of 17 blocks defined by the combination of cohorts, states, and district poverty level as discussed earlier. The impact analysis for the primary research question—what impact does Project CRISS have on the reading comprehension of grade 9 students in high schools in rural and town locales in Northwest Region states?—was performed using the following hierarchical linear modeling procedures.

Two-level hierarchical linear modeling with individual pretest score as the covariate

Level 1 model (that is, student-level model)

$$y_{ij} = \beta_{0j} + \beta_{1j}\mathrm{PRE}_{ij} + e_{ij}$$

Level 2 model (that is, school-level model)

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\mathrm{TRT}_j + \gamma_{02}\mathrm{Block2}_j + \gamma_{03}\mathrm{Block3}_j + \ldots + \gamma_{017}\mathrm{Block17}_j + r_j$$
$$\beta_{1j} = \gamma_{10}$$

where y_{ij} *is* the posttest score of student *i* at school; β_{0j} is the pretest-score-adjusted mean for school *j*; β_{1j} is the slope for the pretest score at school *j*; PRE_{ij} is the pretest score of student *i* at school *j* grand mean–centered; e_{ij} is the residual associated with each student (assumed to be normally distributed with mean 0 and variance $\sigma_{|PRE}^2$); γ_{00} is the pretest-adjusted block-unweighted mean for the control group; γ_{01} is the treatment effect (estimated for the impact analysis; the model implicitly weights the treatment effect according to the size of the block); TRT_j is the indicator variable (treatment school is indicated by 1, control school by 0); γ_{02}–γ_{017} are the difference between the adjusted mean for the control group schools in each block and the adjusted unweighted estimated mean for the control group; $Block2_j$–$Block17_j$ are indicator variables (effect coding is used with block 1 as the referent category); r_j is the residual associated with each school (assumed to be normally distributed with mean of 0 and variance $\tau_{|PRE}$); and γ_{10} is the slope for the pretest score.

At level 1, the posttest score of a student is defined as a function of the school mean, the student's own pretest score, plus the random error associated with each student. The school mean, however, is adjusted for the differences in the pretest score of its student body, as the pretest score will be grand mean–centered.

At level 2, the pretest-adjusted school mean is defined as a function of the pretest-adjusted mean for the control group, the fixed effect of the treatment, the fixed effect of the block, plus the random effect associated with the school. The block was coded using the effect coding method, with block 1 as the referent category. Consequently, the intercept, γ_{00}, represents the pretest-adjusted block-unweighted mean for the control group (that is, the mean of the block means). The treatment effect for each block is assumed to be constant across blocks. This constraint was imposed for the impact analysis in order to estimate the overall treatment effect that is implicitly weighted by the size of the block. The slope for the pretest score is assumed to be invariant across schools for the impact analysis.

The random assignment to conditions took place at the school level. Consequently, the treatment effect will show up at the school level. The grand mean–centered pretest was used as a covariate to adjust the school mean, which led to a more accurate estimate of the treatment effect as well as a gain in statistical power. Other student- or school-level covariates were not included, as the inclusion of these additional covariates was not expected to improve the precision of the model substantially. Additional covariates can make the model unnecessarily complex, which may cause convergence issues and instability in parameter estimates. This problem could be dealt with by constraining the model—that is, making simplifying assumptions about the behavior of those covariates in the model. However, without some ideas about the behavior of those covariates, it is easy to impose incorrect assumptions, which would lead to biases in the parameter estimates.

The hierarchical linear model above reflects the random assignment procedure that used 17 blocks defined by a combination of cohorts, states, and district poverty rates, as described earlier. The block was modeled as a fixed-effect variable because the study did not select the two cohorts randomly out of a universe of cohorts, the study did not select the states randomly out of a universe of all 50 states, and the two categories for the district poverty rate form a finite, exhaustive set. Blocks were defined as 17 distinct "sites," rather than as cross-classified (that is, cohort x state x poverty rate) categories. This was because there was no interest in the effects of block-level background variables, but rather blocks were used solely for the purpose of improving the precision of the impact estimates.

Substituting the higher level model equations with the variables in the level 1 model yields a combined linear mixed model for the above hierarchical linear model:

$$y_{ij} = \gamma_{00} + \gamma_{01}\mathrm{TRT}_j + \gamma_{10}\mathrm{PRE}_{ij} + [\gamma_{02}\mathrm{Block2}_j + \ldots + \gamma_{017}\mathrm{Block17}_j] + r_j + e_{ij}.$$

The combined model makes it clear that the proposed hierarchical linear model will yield the estimates of three fixed effects: the intercept, the treatment effect, and the slope for the covariate, while accounting for the fixed effects of blocks, the random effect of schools, and the random effect of students.

Sensitivity analysis

Sensitivity analysis was conducted to fit the model in which the possible cross-block variability in the treatment effect was modeled. This was done by including the following additional terms to represent the block by treatment interaction to the level 2 equation for the school intercept:

$$\gamma_{018}\mathrm{Block2{*}TRT}_j + \gamma_{019}\mathrm{Block3{*}TRT}_j + \ldots + \gamma_{033}\mathrm{Block17{*}TRT}_j.$$

With the effect coding of the blocks, such a model produced the estimate of the treatment effect that is unweighted by the size of the block (that is, the simple mean of the block-specific treatment effects), as well as the deviation of the block-specific treatment effect from it. These estimates were compared with the treatment effect estimated in the impact analysis to see if the result of the impact analysis needed to be qualified. The slope for the pretest score is assumed to be fixed across blocks.

Missing student data

Figure 2, presented earlier, indicates that 5,592 grade 9 students (2,750 treatment and 2,842 control) were eligible for the study based on September enrollments in the study schools. A total of 4,793 students (2,377 treatment and 2,416 control) completed the baseline pretest. Student rosters indicated that 166 enrolled students (83 treatment and 83 control) failed to complete the baseline measure due to absence. When these absent students are included, there were a total of 4,959 students (2,460 treatment and 2,499 control) in the analysis sample at the time of baseline measurement. Some enrolled students were not tested because of their English language learner status, special education status, or parental nonconsent, as described earlier.

Students who came into the study and were added to the student rosters at the time of the posttest ("arrivers") were not included in the study because there were no data to indicate how long they were exposed to the Project CRISS treatment.

A total of 488 students of the 2,460 treatment group baseline tests (19.8 percent) were missing the posttest score. This number included 120 students who were attending a treatment school that dropped out prior to the posttest. Not counting these 120 students from the dropout school, there were 368 students (15.0 percent) across other treatment schools that were absent for their posttest. These students were classified as "leavers." In comparison, 389 students of the 2,499 control group students who took the baseline test (15.6 percent) were missing the posttest score across all of the control schools. These students were also classified as leavers. Pretest to posttest attrition between treatment school leavers (15.0 percent) and control group leavers (15.6 percent) differed by less than one percentage point when the treatment school that dropped out was excluded from the comparison. However, total attrition between pretest and posttest for the treatment schools was 19.8 percent, compared with 15.6 percent for control schools, a difference of 4.2 percentage points. The overall attrition rate was 877 students (488 treatment and 389 control) out of 4,959, or 17.7 percent.

The leavers plus the students at the treatment schools that dropped out of the study before the posttest were included in the analysis to arrive at an intent-to-treat estimate of the treatment effect ($N = 4{,}959$). Their posttest scores were imputed. For the 166 enrolled students who failed to complete the baseline measure due to absence, their pretest scores were imputed.

Students who had both their pretest and posttest scores were classified as "stayers." Analysis revealed that stayers and leavers were statistically alike with regard to their treatment status and gender, as shown in tables 4 and 5. However, the mean pretest scale score for stayers was 685.60 (95 percent confidence interval: 684.31–686.90), compared with 669.24 (95 percent confidence interval: 666.03–672.46) for leavers. Stayers had significantly higher mean scores than leavers, as evidenced by the lack of overlap between the 95 percent confidence intervals for their estimated means.

Table 4. Comparison of missing status across treatment conditions

	Treatment status		
Missing status	**Control**	**Treatment**	**Total**
Both pre- and posttest scores (stayers)	2,027	1,889	3,916
Pretest scores only (leavers)	389	368	757
Pretest scores only (school dropout)	0	120	120
Posttest scores only (enrolled at pretest)	83	83	166
Total	2,499	2,460	4,959

Note: Pearson chi-square comparing the frequencies of stayers versus leavers across the treatment conditions = 0.036, *p*-value = 0.85.

Source: Authors' analysis of data described in text.

Table 5. Comparison of missing status across gender

	Gender			
Missing status	**Boys**	**Girls**	**No information**	**Total**
Both pretest and posttest scores (stayers)	1,931	1,845	140	3,916
Pretest scores only (leavers)	357	342	58	757
Pretest scores only (school dropout) or Posttest scores only (enrolled at pretest)	85	80	121	286
Total	2,373	2,267	319	4,959

Note: Pearson chi-square comparing the frequencies of stayers versus leavers across gender = 0.001, *p*-value = 0.97.

Source: Authors' analysis of data described in text.

Table 5 shows that a total of 319 students (6.4 percent) were missing gender information. The gender of these students was imputed for the exploratory analysis. Due to concerns over the reliability of the imputed gender, however, a sensitivity analysis was performed in which those 319 students without gender information were dropped from the analysis.

The findings from the missing data analysis, especially the stayers versus leavers, made it clear that missing-completely-at-random could not be assumed as the mechanism behind missing outcome data. Furthermore, the rate of missing outcome data also exceeded the preset cutoff of 5 percent, below which listwise deletion was to be used. As a result, multiple imputation was used to handle missing data. Appendix C presents the data imputation procedures.

3. Implementation of Project CRISS

This chapter describes the implementation of Project CRISS and addresses the following research question: to what extent was Project CRISS implemented with fidelity in the treatment schools? Presented first is the professional development approach used by Project CRISS, which influenced how implementation measures were selected. Second, the expectations for full implementation are discussed, including how they were arrived at in consultation with the developer. Descriptive statistics are then presented for key measures, leading to an assessment of the degree of implementation in Project CRISS treatment schools compared with full or ideal implementation. There were no systematic treatment contrast measures collected in control schools, which were viewed as operating in a "business as usual" mode without the benefit of the highly specified Project CRISS professional development program.[3]

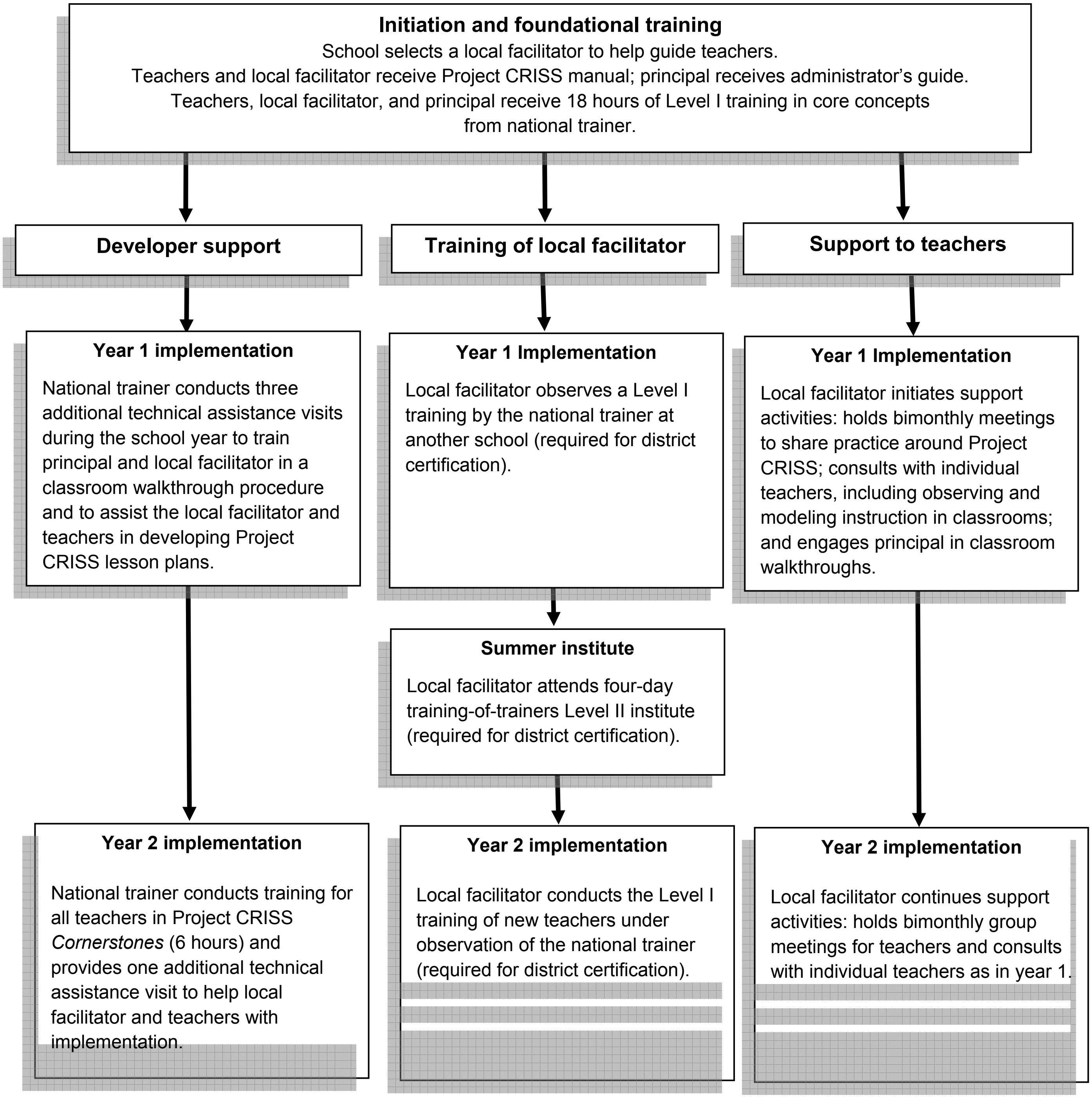
Project CRISS professional development

The Project CRISS treatment tested in this study was two years in duration and consisted of a series of training and technical assistance visits provided by a certified national trainer, plus expectations for follow-up activities by schools. National trainers are typically active or retired teachers throughout the United States who have used Project CRISS extensively in their own classrooms and have undergone a rigorous training-of-trainers program to become certified. Only a certified national trainer can lead a school staff through the Project CRISS professional development program.

National trainers work under the supervision of a training director at Lifelong Learning, Inc., the home office of Project CRISS, located in Kalispell, Montana. Lifelong Learning provides the copyrighted training materials, training agendas, and prescribed schedule of services. Trainers have some latitude to improvise by including their own supplemental materials (such as sample lesson plans) that have worked well in their classrooms. Beyond the first Level I formal training, trainers are also allowed to tailor Project CRISS professional development to the needs and circumstances of specific schools and to teach from their own experience as a Project CRISS teacher. The approach is a teacher-to-teacher professional development network, in which highly experienced CRISS-certified expert teachers work with new districts and schools that contract for a Project CRISS adoption.

Figure 3 shows the Project CRISS professional development process and main activities. A typical implementation begins with selection of a school local facilitator during the spring or summer prior to the Level I training at the start of the school year. The Level I training occurs in groups of up to 30 teachers in order to engage participants in small-group, interactive learning. Teachers are taught how to apply the three core principles and multiple learning strategies discussed in chapter 1. Numerous applications and sample lessons are provided in the 300-page manual that trainers and teachers use throughout the two-year program. Level I training typically occurs during two professional development days in August prior to students returning to school and a third day in early fall, for a total of 18 hours of training.

Figure 3. Project CRISS professional development process



Source: Authors' compilation from Project CRISS materials.

After the Level I training is complete, developer support, training the local facilitator, and support to teachers continue for two years, as shown in figure 3. Continued developer support includes further training and technical assistance visits to help teachers develop lesson plans around the Project CRISS literacy activities, assist the school principal in conducting classroom walkthroughs (that is, teacher observations and feedback), and provide help as needed with implementation challenges. Coupled with this direct support, the local facilitator is expected to undergo further training to become a certified district trainer. This is accomplished by observation of Level I training in another school, attending a four-day summer institute (Level II training), and training new teachers in the school during year 2, under the observation of the national trainer. This enables the local facilitator to provide ongoing support as national trainer activities come to an end after two years. Local facilitators are expected to provide direct support to teachers, including bimonthly meetings with the whole faculty and departmental groups, classroom visits and modeling of instructional strategies, and individual consultations with teachers who need further assistance. Through these three main activities—developer support for two years, training of the local facilitator, and support of teachers by both the national trainer and local facilitator—ongoing and job-embedded professional development is put in place after the initial Level I training. Implementation was measured as the extent to which this prescribed professional development model was carried out in treatment schools.

Project CRISS is not a prescriptive program in which a teacher-ready curriculum and set of lesson plans are provided. As noted in chapter 1, it is a philosophy of teaching backed up with an array of strategies that teachers are expected to learn, experiment with, and apply to their own students. The developers of the program—Dr. Carol Santa and her associates Ms. Lynn Havens and Ms. Bonnie Valdes—express a vision for successful implementation in the preface to their training manual, where they write:

> Successful implementation occurs when teachers and administrators work together to share, extend ideas, and problem solve. The initial workshop must be supported by follow-up sessions in which participants have opportunities to talk about how they are implementing the project. Our strongest adoptions occur in schools and districts where several participants become Certified District Trainers responsible for continuing to disseminate and support the project within their own districts (Santa, Havens, and Valdes 2004, p. viii).

This statement of successful implementation formed the basis for determining fidelity of implementation. To measure full implementation, it was important not only to document the provision of services and materials by the developer, but also to assess local facilitator follow-up support activities and teacher participation in them. The training design encourages the kind of successful implementation described in the quote. During the 18 hours of Level I training, teachers are encouraged to sit together and work in disciplinary teams so that they learn together and begin a regular process of collaboration throughout the school year to apply what they have learned. The local facilitator role is to continue this process by facilitating activities that help teachers experiment with, refine, and ultimately use the instructional practices on a daily basis. By the second year of implementation, the local facilitator becomes a certified district trainer by going through a number of prescribed steps. The national trainer starts the second year with a one-day training (6 hours) in *Cornerstones*, which includes exercises to help teachers collaborate toward fully integrating Project CRISS into a schoolwide approach for instructional improvement. Both the training and the follow-up that schools engage in define a successful implementation.

Expectations for fidelity of implementation

The research team conferred with Project CRISS's Carol Santa and Lynn Havens during a pilot phase prior to the study to set more precise expectations for implementation fidelity, using their idea of successful implementation cited above as a conceptual starting point. A set of criteria for a "five-star school" was found in the training materials as a representation of full implementation. However, the criteria left several key activities not fully specified. One area was the level of support that should be provided by the pivotal local facilitator role, the lead teacher/facilitator who provides the follow-up action and momentum to keep the program going. During planning meetings with the research team, the developers estimated that a facilitator should spend five to six hours a week working with teachers to achieve a high-quality implementation. A full implementation model was thus developed representing what the developers believe is necessary for Project CRISS learning strategies to become embedded in daily teacher practice and internalized by students. Fidelity of implementation is viewed as the extent to which schools achieved full implementation as defined by the developers. Full implementation includes specified training and technical assistance provided by Project CRISS and follow-up actions on the part of the local facilitator and teachers to use, refine, and implement new learning strategies in the classroom. Specific elements of full implementation are presented in appendix D.

The research team developed a memorandum of understanding for participating school principals and a corresponding formal contract with the school district for provision of services. These agreements used the elements of the full implementation model to set an expectation for how the program would be implemented. However, it was the responsibility of the Lifelong Learning staff and national trainers to ensure that each treatment school received materials, training, and support throughout the two years of implementation. The Regional Educational Laboratory (REL) Northwest provided funds for these services through a subcontract arrangement but did not attempt to influence the provision of actual services nor influence the extent to which local facilitators and teachers participated in follow-up activities. This was an intent-to-treat study testing the effectiveness of Project CRISS under real-world conditions in which districts and schools implement the services in collaboration with the provider.

Extent of implementation of Project CRISS in treatment schools

A total of 26 schools were randomly assigned to the treatment condition following recruitment, but, as described earlier, three dropped out of the program and data collection. The implementation results that follow are based on averages across the intent-to-treat analysis sample of 23 treatment schools.

A set of implementation measures was drawn from the full implementation model described earlier. These measures were selected to represent services provided by Project CRISS staff and local facilitator/teacher follow-up activities designed to ensure full implementation. They provide a parsimonious set of observable activities and behaviors describing how faithfully the training and technical assistance model was implemented. The three broad implementation areas and their associated indicators for full implementation are described below.

1. **Selection and retention of the local facilitator**, who:
   - Is an experienced teacher or administrator in the school and who has five to six hours per week release time from regular duties to support teachers.
   - Provides teacher support for two years and participates in Level II training to become a district certified trainer.
2. **Provision of Project CRISS materials, training, and technical assistance**, as follows:
   - Each teacher receives a Project CRISS manual and *Cornerstones* workbook; principals receive an administrator's guide.
   - Level I training is provided to all teachers, with an expected attendance rate of at least 75 percent, plus attendance by the principal.
   - The trainer visits the school at least eight times during the two years to provide Level I training and additional technical assistance to support full implementation.
3. **Follow-up activities to support implementation**—the local facilitator is expected to spend about five to six hours per week facilitating Project CRISS through:
   - One monthly meeting for all teachers and one small-group monthly meeting (for example, by academic department) to provide a forum for sharing practices and discussing implementation problems.
   - Individual consultations with teachers, including observing and modeling instruction in the classroom.
   - Working with the school principal to help him/her use a walkthrough process that involves observing instruction and providing feedback on Project CRISS instructional strategies.

The data sources included documentation from the Project CRISS national center, a monthly activity log completed online by the local facilitator, and a teacher questionnaire administered during the spring of implementation years 1 and 2.

Selection and retention of the local facilitator

Project CRISS national trainers helped school principals select their local facilitators during the spring or summer prior to the first Level I training. A list of local facilitators was maintained by the Project CRISS office and shared with the REL Northwest. The research team also collected data on basic local facilitator background characteristics through a questionnaire.

Local facilitators were selected before initiation of the introductory Level I training. Four schools had local facilitator turnover during the course of the treatment, but in each case a new local facilitator was selected as a replacement. A basic requirement was that local facilitators be in-school teachers or assistant administrators so that they were readily available to help teachers. All of the assigned local facilitators met this requirement. Table 6 provides basic background information on the local facilitators, indicating that these individuals were primarily experienced secondary school teachers (mean years of secondary experience = 15.0) who taught for many years in their current schools (mean years in school = 10.1). Most had a master's degree (72.7 percent) and taught English/language arts as their core class (68.2 percent) and were therefore already experienced in reading and writing learning strategies.

Table 6. Project CRISS local facilitator background characteristics

Teacher characteristic	Statistic	Value
Length of time secondary teacher	Mean (years)	15.0
	Range (years)	1–32
Length of time at present school	Mean (years)	10.1
	Range (years)	1–32
Highest academic degree	Bachelor's degree (percent)	27.3
	Master's degree (percent)	72.7
Core subject taught	English/language arts (percent)	68.2
	Other core or no core class (percent)	31.8

Note: This table represents the initial set of local facilitators assigned across the 23 treatment schools, four of whom left and were replaced.

Source: Authors' analysis of background questionnaire completed by local facilitators.

Project CRISS held summer regional institutes available to all study schools in order to provide the four-day Level II training-of-trainers for project local facilitators. In addition to the annual gathering in Kalispell, Montana, a second regional institute was held in Portland, Oregon, at the REL Northwest headquarters to offer an alternative location for project schools. This was done to make travel easier and less expensive, particularly for the northern California, Oregon, and Washington project schools. Based on information from the Lifelong Learning training coordinator, it was found that all 22 local facilitators who were selected attended a four-day Level II institute to strengthen their Project CRISS knowledge and training skills, thereby meeting a requirement for district certification as required in the full implementation model.

Provision of Project CRISS materials, training, and technical assistance

The Project CRISS home office supplied documentation indicating that all treatment schools participating in the program received required materials, including teacher training manuals and administrator guides specified in the full implementation model. Project CRISS calls for a total of eight professional development days across two years (five for formal training, three for additional assistance), exclusive of optional days needed for new-teacher Level I make-up training. Documentation from Project CRISS and from the local facilitator log of activities indicated that the national trainers provided 20 of the 23 treatment schools (87 percent) with between 8 and 12 total training/consulting days over two years, while three schools had fewer than the prescribed eight days. The observed differences across schools were due to the needs of individual schools, including the new-teacher training in year 2, and are consistent with the idea that the national trainer uses some professional discretion in how much follow-up schools require.

The foundational training for Project CRISS is the Level I training that occurs at the beginning of the school year, a two- or three-day kickoff event before students return to school for the fall. (Schools can either have all three training days at the beginning of the year or have two training days before school starts and one day later in the early fall, depending on professional development days available.) Attendance was tracked for all teachers eligible to attend the training (that is, those teaching core subjects) on this important event as a sign of early commitment to Project CRISS by the treatment schools. The average participation rate across the 23 treatment schools was 78 percent. Fifteen schools (65.2 percent) had at least 75 percent attendance as specified in full implementation; three of the eight schools that did not had attendance below 50 percent. Among the schools participating in the Level 1 training, all principals attended this training along with their faculties, which is also part of full implementation.

In summary, Lifelong Learning provided the requisite materials and core training to all treatment schools. Treatment schools showed an overall high rate of participation in the core Level I training, with a school average of 78 percent attendance by eligible teachers. The average number of professional development visits made to each school during the two-year implementation period was nine. These results are consistent with full implementation of Project CRISS services.

Follow-up activities to support implementation

The local facilitators in treatment schools used a web-based log created for the study to document the project activities they undertook each month. They were asked about key support activities that are listed in the full implementation description, such as holding monthly meetings and individual teacher consultations to help the faculty implement Project CRISS. The logs were maintained during the two years of the implementation, beginning when Level I training started and ending with the school year.[4]

The local facilitators were asked to estimate the number of hours they spent each month facilitating implementation of Project CRISS by working with teachers or the principal. On average across all 23 treatment schools (including the school that did not implement Project CRISS in the intent-to-treat sample), just over three-quarters of the local facilitator monthly reports (76.4 percent for year 1 and 77.4 percent for year 2) indicated five or fewer hours per month facilitating Project CRISS. This translates to about one hour or less per week. These reported activity levels were lower than the prescribed full implementation by a wide margin, which recommends five to six hours per week.

Items were also included about specific professional development follow-up activities to provide a clearer picture of the ways in which teachers were supported or sought support to implement Project CRISS practices. The results below show the frequency of key activities presented from the points of view of the local facilitator providing the support (table 7) and the teachers who were to receive it (table 8), noting in each table the prescribed full implementation level.

Table 7 presents results for formal group meetings and individual teacher consultations to support Project CRISS. Results are presented separately for each of the two years of implementation. (The impact analysis occurred during the second year of implementation, by which time teachers would conceivably be more fully using Project CRISS in their classrooms.)

Table 7. Local facilitator frequency of providing assistance to teachers with Project CRISS

Activity	Frequency Year 1	Frequency Year 2
Hold formal meetings with teachers participating in Project CRISS (full implementation: two per month)		
Percentage of months at least one formal group meeting held	51.5	51.3
Mean number of formal meetings per school year	8.7	8.1
Standard deviation (meetings per year)	6.4	7.4
Hold individual consultations with teachers participating in Project CRISS (full implementation: as needed by teachers)		
Percentage of months at least one individual teacher consultation	55.1	48.7
Mean number of consultations per school year	16.2	13.8
Standard deviation (consultations per year)	16.8	16.7

Note: n = 23 schools

Source: Authors' analysis of monthly log completed by local facilitators.

Table 7 indicates that, on average across all schools and months of the treatment, local facilitators held at least one formal group meeting for only about half of the months they worked as a local facilitator (51.5 percent in year 1 and 51.3 percent in year 2). Local facilitators held at least one individual teacher consultation for about half the months, with some variation between year 1 (55.1 percent) and year 2 (48.7 percent). The implication is that for about half the months, there were no group meetings or individual teacher consultations centered on Project CRISS. The full implementation prescribes about two group meetings per month (for example, one with full faculty and one with an academic department), while individual consultations are recommended on an as-needed basis.

The number of events for both group meetings and individual consultations show wide variation across schools, as evidenced in the standard deviations (see table 7). As a further illustration of the range across different schools, during year 1 there were three schools that held 20 or more formal group meetings during the school year, compared with nine schools that held 5 or fewer. This variation could be due to varying school size (number of core teachers), the degree to which the local facilitator had sufficient free time to facilitate and guide teachers, and the interest level and time of teachers.

The teacher questionnaire included items asking teachers to estimate their own participation in local facilitator–sponsored Project CRISS activities. Teachers were also asked about the frequency of principal or administrator classroom observations or walkthroughs related to Project CRISS (see table 8).

Table 8. Teacher self-reported participation in Project CRISS activities

Activity	Response	Percent Year 1	Percent Year 2
Attended meetings facilitated by Project CRISS local facilitator (full implementation: two per month)	Never	15.3	11.3
	Several times a year or less	70.6	60.4
	Once a month	10.4	16.4
	More than once a month	3.7	11.9
Have had local facilitator observe or assist in classroom with Project CRISS principles and strategies (full implementation: as needed by teachers)	Never	32.3	32.9
	Several times a year or less	57.1	58.9
	Once a month	8.1	6.3
	More than once a month	2.5	1.9
Have had principal or other administrator observe in classroom and provide feedback about Project CRISS principles and strategies (walkthroughs) (full implementation: no specific recommendation)	Never	31.9	32.7
	Several times a year or less	60.1	57.9
	Once a month	6.1	6.9
	More than once a month	1.8	2.5

Note: n varies from 158 to 163 for specific items because of listwise deletion of missing data.

Source: Authors' analysis of teacher questionnaires administered at the end of year 1 and year 2.

During year 1, 70.6 percent of the teachers reported attending a formal Project CRISS meeting several times a year or less, with 15.3 percent reporting never attending a meeting, and 10.4 percent reporting attending once a month. There was an increase in meeting attendance by year 2: 60.4 percent reported attending several times a year or less, 11.3 percent reported never attending, 16.4 percent reported attending once a month, and 11.9 percent reported attending more than once a month. This increase could be due to the design of the treatment, in which the *Cornerstones* guide is introduced at the beginning of year 2 to provide teachers with group exercises to deepen their understanding and use of Project CRISS. As noted above, full implementation specifies attending two group meetings a month.

Individual assistance in the form of the local facilitator assisting or observing in the classroom is also reported in table 8. About a third of teachers (32.3 percent in year 1 and 32.9 percent in year 2) reported never having the local facilitator in their classroom for observation or assistance. A majority of teachers (57.1 percent in year 1 and 58.9 percent in year 2) reported receiving local facilitator classroom assistance several times a year or less. Full implementation prescribes individual consultations on an as-needed basis.

Finally, as an index of principal involvement, teachers were asked to estimate the frequency of principal or administrator classroom observations (called walkthroughs in the material) that occurred for the explicit purpose of providing feedback on Project CRISS learning principles and strategies. In the full implementation model, these principal walkthroughs are encouraged to signal the importance of Project CRISS and provide feedback from the school's instructional leader. As with local facilitator consultations, this type of assistance did happen, but not on a regular basis. Just under a third of teachers (31.9 percent in year 1 and 32.7 percent in year 2) reported that the principal or another administrator never conducted walkthroughs to provide feedback on Project CRISS adoption. A majority of respondents (60.1 percent in year 1 and 57.9 percent in year 2) said this type of feedback occurred several times a year or less.

In summary, the local facilitator self-reports and teacher self-reports should be taken as separate views of implementation that are not directly comparable. The local facilitator self-reported activity data were recorded monthly, while the teacher self-report data were collected through end-of-year surveys using a more general response set for judging frequency. Nevertheless, the results of both viewpoints, while not directly comparable, do indicate a similar trend: there was less follow-up activity reported in the schools than prescribed by full implementation, in which teachers and local facilitators are encouraged to meet frequently and regularly to share and improve practice.

Summary of implementation

A national trainer provided regular assistance to help treatment school teachers learn and adopt a set of learning principles and strategies embedded in the Project CRISS conceptual framework. Schools participating in the service selected experienced teachers for the local facilitator position; two-thirds of the local facilitators taught English/language arts as their core subject. The facilitators underwent additional training to become district-certified trainers. These results were close to full implementation as prescribed by the developer.

Regarding local facilitator and teacher follow-up activities, the results did not reach a level of high implementation compared with the prescribed program. Placed against full implementation, local facilitator and teacher follow-up activities to strengthen and deepen practice were less than expected. Key follow-up activities were inconsistent and less frequent than full implementation as prescribed by the developer.

The current study measured implementation in reference to a full professional development model prescribed by Project CRISS, as indicated by the activities of national trainers, local facilitators, and teachers in project schools. The extent to which teacher practices might have changed or not changed in treatment schools compared with control schools is not addressed in this kind of analysis.

4. Results: did Project CRISS improve student reading comprehension?

This chapter presents the results of analyses performed to answer the primary research question: what impact does Project CRISS have on the reading comprehension of grade 9 students in high schools in rural and town locales in Northwest Region states? The first section presents results of the impact analysis used to estimate the treatment effect; the second section presents results of two sensitivity analyses to determine whether explicit modeling of the effect size variability across blocks would alter the substantive result.

Impact analysis

The estimated effect of the treatment is illustrated as the difference between the covariate-adjusted posttest scores of the treatment group students and the control group students. The standard error of the estimate is adjusted for the nesting of students within schools, as are the significance test and the confidence interval for the difference. The estimated treatment effect was 2.15 on the Stanford Diagnostic Reading Test, Fourth Edition, comprehension scale score (table 9). Although the treatment group mean score was higher than that of the control group, this difference was not statistically significant.

Table 9. Estimated impact of Project CRISS adjusted for pretest score and blocks (n = 4,959)

Outcome measure	Treatment group[a]	Control group[a]	Difference: estimated treatment effect	Standard error of estimate	Test statistic	95 percent confidence interval	Effect size[b]
Posttest score adjusted for pretest score and blocks	694.04	691.89	2.15	2.80	$t = 0.77$ $p = 0.44$	–3.34 to 7.64	0.05

a. The covariate-adjusted posttest score for the treatment group represents the predicted posttest score of a treatment group student who had a pretest score at the grand mean, controlling for the fixed effect of blocks. Likewise, the covariate-adjusted posttest score for the control group represents the predicted posttest score of a control group student who had a pretest score at the grand mean, controlling for the fixed effect of blocks.

b. Hedges's *g* (standardized difference using the pooled within-condition standard deviation of the posttest scores).

Source: Authors' analysis based on data described in text.

The Stanford Diagnostic Reading Test reading comprehension scale score ranges from 490 to 814, with a scale score of 693 corresponding to the 50th percentile for students tested in the fall of grade 9. The test developer's scale score as the metric is not comparable to other tests or studies. Consequently, for the purposes of illustrating effect size, this value was standardized using the pooled within-condition standard deviation of the posttest score (Hedges's g). Hedges's g was calculated using the formula suggested by the What Works Clearinghouse for effect size calculations for multilevel analysis (U.S. Department of Education 2008). This formula derives the pooled within-condition standard deviation from sample statistics, as a weighted average of within-condition standard deviations. The effect size in standard deviation units was 0.05. Appendix E presents the complete multilevel model results for the impact research question.

Sensitivity analyses

Two sensitivity analyses were performed to investigate the effects of the blocking procedure and the effects of not collecting posttest scores for some students.

Sensitivity analysis one

Because random selection was conducted within a blocking design defined by cohorts, states, and poverty status, a sensitivity analysis was performed to determine whether explicit modeling of the effect size variability across blocks would alter the substantive result.

The estimated treatment effect was in the opposite direction from the original impact analysis model (table 10). However, the substantive result remained the same. With or without the constraint of equal treatment effect, the estimated treatment effect was not statistically significant.

Table 10. Estimated impact of Project CRISS adjusted for pretest score and blocks, and with explicit modeling of effect size variability across blocks (n = 4,959)

Outcome measure	Treatment group[a]	Control group[a]	Difference: estimated treatment effect	Standard error of estimate	Test statistic	95 percent confidence interval	Effect size[b]
Posttest score adjusted for pretest score and blocks	693.54	696.20	–2.66	10.81	$t = -0.25$ $p = 0.81$	–23.85 to 18.53	–0.06

a. The covariate-adjusted posttest score for the treatment group represents the predicted posttest score of a treatment group student who had a pretest score at the grand mean, controlling for the fixed effect of blocks. Likewise, the covariate-adjusted posttest score for the control group represents the predicted posttest score of a control group student who had a pretest score at the grand mean, controlling for the fixed effect of blocks.

b. Hedges's *g* (standardized difference using the pooled within-condition standard deviation of the posttest scores).

Source: Authors' analysis based on data described in text.

Sensitivity analysis two

The impact analysis data included a group of students who did not have posttest scores. Consequently, the impact analysis data contained imputed posttest scores for these 120 students. The imputing of the posttest scores for these students was deemed appropriate, since the imputation model contained the pretest scores. Still, it was considered prudent to run a sensitivity analysis using the data without those 120 students.

The estimated treatment effect was in the opposite direction from the original impact analysis model (table 11). However, the substantive result remained the same: in both cases the test statistic was not statistically significant.

Table 11. Estimated impact of Project CRISS adjusted for pretest score and blocks, after removing students attending a dropout school (n = 4,839)

Outcome measure	Treatment group[a]	Control group[a]	Difference: estimated treatment effect	Standard error of estimate	Test statistic	95 percent confidence interval	Effect size[b]
Posttest score adjusted for pretest score and blocks	692.52	692.60	–0.08	2.25	*t* = –0.04 *p* = 0.97	–4.49 to 4.33	–0.002

a. The covariate-adjusted posttest score for the treatment group represents the predicted posttest score of a treatment group student who had a pretest score at the grand mean, controlling for the fixed effect of blocks. Likewise, the covariate-adjusted posttest score for the control group represents the predicted posttest score of a control group student who had a pretest score at the grand mean, controlling for the fixed effect of blocks.

b. Hedges's *g* (standardized difference using the pooled within-condition standard deviation of the posttest scores).

Source: Authors' analysis based on data described in text.

5. Results of additional exploratory analysis

An exploratory analysis was performed to answer the following question: does the impact of Project CRISS on student reading comprehension differ for boys and girls? The rationale for this analysis is that there tends to be a gender gap between boys and girls in reading proficiency; for example, National Assessment of Educational Progress (NAEP) reading proficiency scores for grade 8 students show a long-term trend of female students consistently scoring higher than male students (U.S. Department of Education 2009). Given the plausibility that gender differences in reading proficiency might exist as students enter high school, the authors chose to explore whether there might be a gender by treatment interaction effect in order to more fully understand Project CRISS effects.

It was hypothesized that the program might have a stronger effect on boys who can benefit more from the reading strategies because they enter high school with generally lower reading proficiency. This kind of subgroup analysis provides more fine-grained information to school superintendents and principals considering whether to adopt Project CRISS. Exploratory analyses were also considered for ethnic subgroups, which show differential grade 8 reading proficiency on NAEP. However, these analyses were not performed due to restricted subgroup sample sizes.

Formally, the question concerns the moderator effect of gender. Gender is a student-level variable, and the moderating effect of gender was estimated by adding a cross-level interaction term "treatment by gender" to the impact analysis model. Gender main effect was also added. The resulting model for the exploratory analysis was:

$$y_{ij} = \gamma_{00} + \gamma_{01}\text{TRT}_j + \gamma_{10}\text{PRE}_{ij} + \gamma_{20}\text{Boy}_{ij} + \gamma_{21}\text{TRT}_j{*}\text{Boy}_{ij} + [\gamma_{02}\text{Block2}_j + \ldots + \gamma_{017}\text{Block17}_j] + r_j + e_{ij}.$$

The model adjusted the treatment effect for the individual pre-score, which was grand mean–centered. Also included in the model were the fixed effects of blocks, which were effect-coded with block 1 as the referent. Gender was dummy-coded, with girl as the referent category. As such, the intercept (γ_{00}) represented the estimate of the postscore for a girl in the control group whose pre-score was at the grand mean. The coefficient for the treatment main effect (γ_{01}) represented the treatment effect for girls. The coefficient for the treatment by gender interaction (γ_{21}) was the moderating effect of gender, representing the gender difference in the treatment impact. The treatment effect for boys, therefore, could be calculated as $\gamma_{01} + \gamma_{21}$. Gender main effect was represented by the coefficient for Boy (γ_{20}). Two random effects, the school random effect and the individual student residual, were included in the model. The test based on student's *t* distribution was performed to test the significance of the moderator effect of gender (γ_{21}), to answer the research question does the impact of Project CRISS on student reading comprehension differ for boys and girls? In addition, the main effect of gender (γ_{20}) was also tested with student's *t*.

There were 2,373 identified boys and 2,267 identified girls in the impact analysis sample. The remaining 319 students in the analysis sample lacked gender information. The gender information for these 319 students was imputed as a part of multiple imputations for the impact analysis. The dataset for the impact analysis was also used for the exploratory analysis.

The imputation model contained available variables from the pretesting and the posttesting, such as scale scores, raw scores, and number of items attempted. It also contained the student race/ethnicity and the dummy-coded school identification number (design variable). Because none of those variables was a good predictor for gender, the imputed gender was not likely to be reliable. To address this concern, a sensitivity analysis was performed using the same imputed dataset with all cases for the 319 students whose gender information was originally missing deleted. A check was also performed to see whether the missing gender information was clustered in some ways, because this information came from each school separately.

Exploratory analysis results

The estimated treatment effect for boys was larger than that for girls by 0.593 on the scale score, although the difference between the two was not statistically significant (table 12).

Table 12. Estimated moderator effect of gender on the Stanford Diagnostic Reading Test scale score, adjusted for pretest (n = 4,959)

Outcome measure	Treatment effect for girls	Treatment effect for boys	Difference: moderator effect	Standard error of estimate	Test statistic	95 percent confidence interval
Stanford Diagnostic Reading Test scale score adjusted for pretest	1.793 *n* = 2,267	2.386 *n* = 2,373	0.59	2.21	$t(30.9) = 0.27$ $p = 0.79$	–3.02 to 5.65

Note: The gender code was missing for 319 students, whose gender was subsequently imputed for the purpose of analysis. Imputed data on average contained 51 percent boys and 49 percent girls. Total *n* = 4,959.

Source: Authors' analysis based on data described in text.

Across the genders, the treatment students had higher estimated mean scores than the control students, though the difference was not statistically significant (table 13). Irrespective of the treatment status, girls had higher estimated mean scores than boys. This difference was significant ($t = 4.35$, $p < 0.001$).

Table 13. Estimated Stanford Diagnostic Reading Test scale-score means for girls and boys in the treatment and control groups, adjusted for pretest (n = 4,959)

Experimental condition	Girls	Boys
Treatment	696.88	691.34
Control	695.08	688.95

Note: Each mean score is covariate-adjusted, representing the predicted score of a student who had a pretest score at the grand mean.

Source: Authors' analysis based on data described in text.

Sensitivity analysis results

Prior to the sensitivity analysis, the data were checked to see if the missing gender information was clustered in some ways. Of 319 cases with missing gender information, 302 (95 percent) came from three schools that failed to provide student demographic data on the student rosters. The remaining 17 cases with missing gender data (5 percent) came from schools other than these three.

In the sensitivity analysis, the estimated treatment effect for boys was larger than that for girls by 0.243 on the scale score; this difference was not statistically significant (table 14).

Table 14. Estimated moderator effect of gender on the Stanford Diagnostic Reading Test scale score, adjusted for pretest (n = 4,640)

Outcome measure	Treatment effect for girls	Treatment effect for boys	Difference: moderator effect	Standard error of estimate	Test statistic	95 percent confidence interval
Stanford Diagnostic Reading Test scale score adjusted for pretest	–0.213 n = 2,267	0.030 n = 2,373	0.24	2.20	$t(33.5) = 0.11$ $p = 0.91$	–3.77 to 4.84

Note: The gender code was missing for 319 students, and those students were excluded from the analysis. Total n = 4,640.

Source: Authors' analysis based on data described in text.

The treatment effect was negative for girls and positive for the boys, although this difference was not significant (see table 14). Irrespective of the treatment status, girls had higher estimated mean scores than boys (table 15). This difference was significant ($t = 4.22$, $p < 0.001$). Taken together, the pattern of results differs between the exploratory analysis and its sensitivity analysis. However, this difference was small and within the realm of random error. The substantive conclusion remained the same across the two analyses: No moderating effect of gender was found.

Table 15. Estimated Stanford Diagnostic Reading Test scale-score means for girls and boys in the treatment and control groups, adjusted for pretest (n = 4,640)

Experimental condition	Girls	Boys
Treatment	695.46	689.65
Control	695.67	689.62

Note: Each mean score is covariate-adjusted, representing the predicted score of a student who had a pretest score at the grand mean.

Source: Authors' analysis based on data described in text.

6. Summary of findings and study limitations

This chapter summarizes the findings regarding the main impact question on the effects of Project CRISS on grade 9 student reading comprehension and discusses the study's limitations.

Effect of Project CRISS on student reading comprehension

The primary purpose of this study was to estimate the impact of the Project CRISS teacher professional development program on grade 9 student reading comprehension, as measured by the Stanford Diagnostic Reading Test, Fourth Edition, Comprehension Subtest. The pretest and posttest were administered in the fall and spring, respectively, of the second implementation year of the treatment. It was hypothesized that by year 2 of the treatment, teachers would have had familiarity with how to use the Project CRISS instructional methods, given the training and guided coaching they received during the first year of implementation. During the second year, when student impact was assessed, the model specified that teachers would continue to receive some training from the national trainer and technical assistance from the district-certified, in-school local facilitator.

There was no statistically significant difference between the treatment group and control group on the mean reading comprehension scores. Because random selection was conducted within a blocking design defined by cohorts, poverty status, and state, a sensitivity analysis was performed to determine whether explicit modeling of the effect size variability across blocks would alter the substantive result. The substantive result remained the same: there was no statistically significant treatment effect.

In addition to the main impact analysis, one exploratory research question was addressed: does the impact of Project CRISS on student reading comprehension differ for boys and girls? Concerning the moderating effect of gender, no statistically significant impact difference was found between boys and girls

Study limitations

Other randomized and nonrandomized studies of Project CRISS have been conducted (see chapter 1), and no single study provides a definitive answer to the question of program effectiveness. This experimental study contributes to the growing body of empirical evidence on both Project CRISS and strategies to improve adolescent reading comprehension more generally. However, this study and the findings reported here are limited by several design and contextual factors. The study limitations are presented below.

- The findings of this study apply to grade 9 students in smaller rural and town high schools (ranging in size from 250 to 1,000 students) in six western states. The study does not address whether the intervention might be more or less effective with subgroups of students except for the one exploratory analysis on gender differences.
- This was a voluntary participation study in which schools and teachers agreed to participate in the Project CRISS treatment, while receiving services at no cost to the school or district. The extent to which the results would apply to other situations, such as targeted or mandated implementations for particular types of schools or students, is unknown and cannot be inferred from this study.

- Implementation was measured as adherence to the Project CRISS professional development model, including training activities provided by the developer and self-reported follow-up activities by local facilitators and teachers. The extent to which teachers learned the content provided to them in training and put into practice what they learned was not assessed. The study does not answer the implementation questions of whether teachers fully understood the Project CRISS concepts and used them in daily classroom practice.
- The student achievement data were collected during the second year of teacher professional development under Project CRISS and covered a learning period of about seven months from pretest to posttest. The study does not answer what effects the treatment may have had during a longer duration of classroom implementation or student assessment. While there may have been some movement of teachers by the second year, the likelihood that teachers in treatment schools moved to control schools (as a threat to internal validity) was minimized by the fact that treatment and control schools were in separate districts and towns.

Appendix A. Statistical power analysis

This appendix contains a summary of a prospective power analysis during the planning phase of the study, which was performed to determine the sample size and the expected minimum detectable effect size (MDES). It also contains a summary of a retrospective power analysis based on the actual data, performed for assessing the actual MDES for the study.

Prospective power analysis for impact study

To determine the level of statistical power attainable from various sample sizes, a set of analyses was performed to estimate the MDES under several scenarios. The MDES was defined as the necessary size of effect in order to maintain statistical power of 0.80. The power analysis was performed for the detection of the main effect of the treatment on the student outcome.

The design of this study was a multisite cluster-randomized trial (multisite CRT), as reflected in the model for the impact analysis (see the original impact analysis model in chapter 2). The site (that is, block) effects were modeled as fixed effects; however, this original model could be simplified into that of a student-nested-within-school CRT. In this way, the fixed effect of the student pretest score and the fixed effects of blocks became a set of covariates to improve the precision of estimating the treatment effect (see the combined model in chapter 2).

Optimal Design Software (Raudenbush et al. 2005) was used for the power analysis. The following describes the input values used for the analysis, and the rationale for those values.

The target MDES was in the range of $\delta = 0.10–0.25$. In the prospective power analysis, the number of students (grade 9) per school (n) was set to 50 as a conservative estimate, based on actual population data of class size in the smallest schools under study. The standard default value of 0.05 for a two-tailed test of significance was used for the alpha level. Publisher data on the test used (Stanford Diagnostic Reading Test) was not available to estimate the intraclass correlation coefficient (ICC). Instead the ICC was estimated from other available test data in the region: the grade 4 data (n = 39,057) of the Oregon Writing Assessment (2005/06) and the ICC value calculated from the grade 7 and 8 data (n = 74,210) of the Washington Mathematics Assessment (2005/06). Even though these represent different subjects and test batteries on different student grade levels than the current study, there should be some similarity of general school effects on student achievement in math, writing, and reading.

An identical procedure was followed to calculate the ICC for each dataset. First, the unconditional ICC was calculated by fitting a two-level CRT model without any covariate. Then, the conditional ICC was calculated by fitting a two-level CRT model with the previous year's building mean as the school-level (level 2) covariate. When a two-level CRT model without covariate was fit to the Oregon Writing Assessment data, school-level variance (τ) was 2.917, and student-level variance (σ^2) was 18.221. Unconditional ICC was therefore calculated as $\tau / (\tau + \sigma^2) = 0.138$ for the Oregon Writing Assessment data. When the model was fit to the Washington Mathematics Assessment data, school-level variance (τ) was 288.66, and student-level variance (σ^2) was 1,465.18. Unconditional ICC was therefore calculated as $\tau / (\tau + \sigma^2) = 0.165$ for the Washington Mathematics Assessment data. Based on these results, it was decided to use the unconditional ICC value of 0.15 for the prospective power analysis.

When a two-level CRT model with covariate was fit to the Oregon Writing Assessment data, school-level conditional variance ($\tau_{|x}$) was 1.574, and student-level variance (σ^2) was 18.225. Conditional ICC was therefore calculated as $\tau_{|x} / (\tau_{|x} + \sigma^2) = 0.079$ for the Oregon Writing Assessment data. When a two-level CRT model with covariate was fit to the Washington Mathematics Assessment data, school-level conditional variance ($\tau_{|x}$) was 15.11, and student-level variance (σ^2) was 1,465.82. Conditional ICC was therefore calculated as $\tau_{|x} / (\tau_{|x} + \sigma^2) = 0.010$ for the Washington Mathematics Assessment data. Since the conditional ICC here was unusually small for the latter, a conservative approach was followed, with the conditional ICC value of 0.05 used for the power analysis.

The effect of covariate, R^2_{L2}, was calculated from the initial school-level variance (τ) and the subsequent school-level variance conditional to the use of covariate ($\tau_{|x}$):

$$R^2_{L2} = 1 - (\tau_{|x} / \tau) = 1 - (0.05 / 0.15) = 0.667.$$

Based on the information above, the unconditional ICC of 0.15 was used. Then, the effect of covariate (R^2_{L2}) of 0.67 was entered, yielding the power estimates presented in table A1. The MDES values arrived through this method would be somewhat conservative (higher), in that a set of covariates at the school level (block fixed effects) was not accounted for in the power analysis.

Table A1. Statistical power analysis

Unconditional intraclass correlation coefficient	R^2_{L2}	Number of schools	Minimum detectable effect size (at power of 0.80)
		214	.100
		96	.150
		60	.190
		56	.200
0.15	0.67	36	.250

Source: Authors' analysis based on data described in text.

At the outset of the study, the desired sample size was set to be large enough to detect a plausible effect size given the nature of the program while also meeting cost constraints of the study. Project CRISS is a teacher professional development program requiring a two-year faculty commitment to participate in formal training plus follow-up activities (e.g., monthly meetings), special training for a local facilitator, and expenditure of school funds for training and support services. A recent evidence review of 1,300 studies addressing the effects of teacher professional development on student achievement provides some guidance on determining a plausible effect size. The review, conducted on studies of grade K–5 teacher professional development programs and students, identified nine studies that met What Works Clearinghouse evidence standards; these nine studies included 20 individual effect sizes. Across the 12 effect sizes focused on reading and language arts programs, the average effect size was 0.53 standard deviation units, with a range of -0.53 to 2.39 (Yoon et al. 2007). The average professional development hours across the nine studies was 49 teacher hours, comparable to the

48 professional development hours designed into Project CRISS; five formal training days and three additional assistance days (at six hours each) over a two-year period.

In addition to this estimate of average effect size found across studies of teacher professional development programs in reading, other considerations were taken into account. Konstantopoulos and Hedges (2008) point out that selecting the appropriate minimum detectable effect size is a balance between selecting a value that is large enough to be educationally important but not so small as to be disappointing given the effort put into the reform. These authors make the case, based on an examination of NAEP trend data, that school reform effect sizes of one-quarter standard deviation or less should not be summarily ignored. Because most achievement variation is within—rather than between—schools, effect sizes of this magnitude might still be educationally significant when compared to other educational reforms.

Given prior research on teacher professional development programs, and in order to ensure that an educationally significant effect size would not be missed, the decision was made to select enough schools and students to maintain the power of 0.80 for a minimum detectable effect size of $\delta = 0.10–0.25$. Given the costs of testing all grade 9 students among schools geographically dispersed across a six-state rural area, an initial sample size of 60 schools was sought with an estimated MDES of 0.19 (see table A1), which was within the desired range and which could be accomplished with study resources.

The goal of 60 schools was not achieved after two rounds of recruitment and exhausting the population of schools that fit the sampling criteria. The final sample was 52 schools, divided evenly between treatment and control. Recalculation of power for the new sample size was conducted. Using the same assumptions above and using $n = 52$ (26 treatment, 26 control) as the obtained sample, the updated MDES was estimated at 0.211.

Retrospective power analysis for impact study

A retrospective power analysis was performed after the impact analysis was completed. The power analysis was based on the actual data and therefore yielded the actual MDES of the current study.

The Stata output for the impact analysis (see appendix E) includes the school-level ICC, conditional to all the covariates in the model including the pretest score and the blocks. This value (0.07) was used in place of the unconditional ICC, while the effect of covariate, R^2_{L2}, was set to 0 to offset the substitution of the conditional ICC value for the unconditional ICC value in power calculation. The Stata output also includes the number of schools ($J = 49$) as well as the average number of students per school ($n = 48$). These values were used for the power calculation. Under the Optimal Decision Software Version 2.0 (2009), the power calculation for a two-level CRT with an ICC of 0.07, R^2_{L2} of 0, 49 clusters, and cluster size of 48 was selected. The resulting MDES was 0.25.

Based on the study means and standard deviations for the treatment and control groups, an MDES of 0.25 corresponds to approximately 11 scale score units on the Stanford Reading Diagnostic comprehension test. Publisher fall grade 9 norms (see Karlsen and Gardner, 1996, table 17) revealed that an 11-point scale score difference corresponds to approximately 10–12 percentile points improvement in test performance (specifically, from the 30th to 40th percentile; or from the 40th to the 51st percentile; or from the 50th to the 62nd percentile). In conclusion, the actual power of the impact analysis was within the range originally set (MDES of 0.10–0.25) and a reasonable expectation for an extensive teacher professional development program.

Retrospective power analysis for exploratory research question

For the exploratory analyses investigating the moderating effect of student gender, the retrospective power analysis result (MDES) for the impact study was adjusted using the formula in Bloom (2005, p. 141). The value for the proportion of the referent subgroup (girl) was estimated at 46 percent through a descriptive analysis. The ICC value was assumed to be the same as the one for the impact analysis (0.07). The size of n was assumed to be the same as the one for the impact analysis (48), since this was the moderator analysis comparing boys and girls. The resulting MDES of 0.218 was obtained.

Appendix B. Teacher questionnaire for treatment schools

56980

4-10

Experimental Study of Project CRISS
TEACHER QUESTIONNAIRE

This questionnaire is part of an experimental study being conducted by Education Northwest (formery Northwest Regional Educational Laboratory) under contract with the U.S. Department of Education. Your answers are critically important and will be used to better understand teacher characteristics in the schools participating in the study. Please be candid in your answers. Responses to this data collection will be used only for statistical purposes. The reports prepared for this study will summarize findings across the sample and will not associate responses with a specific district, school, or individual. We will not provide information that identifies you or your district or school to anyone outside the study team, except as required by law.

Please use a black pen or No. 2 pencil, fill in the bubbles completely, and do not fold, since your answers will be read by a scanner.

For questions 1-3, answer with respect to how often this activity occurred for you during the current school year.

Activity	Never	Several times a year or less	Once a month	More than once a month
1. I attended meetings facilitated by the Project CRISS local facilitator to share and discuss applications of Project CRISS learning principles and strategies.	◯	◯	◯	◯
2. Project CRISS local facilitator came into my classroom to observe and/or assist me on Project CRISS learning principles and strategies.	◯	◯	◯	◯
3. Principal or other administrator came into my classroom and observed, then provided feedback about my application of Project CRISS principles and strategies.	◯	◯	◯	◯

The remaining questions ask about your background and experience.

4. How many years (including this year) of middle or high school teaching experience do you have? **(bubble in number)**

☐☐ years

1	◯	◯
2	◯	◯
3	◯	◯
4	◯	◯
5	◯	◯
6	◯	◯
7	◯	◯
8	◯	◯
9	◯	◯
0	◯	◯

5. How many years (including this year) have you been teaching at *this* school? **(bubble in number)**

☐☐ years

1	◯	◯
2	◯	◯
3	◯	◯
4	◯	◯
5	◯	◯
6	◯	◯
7	◯	◯
8	◯	◯
9	◯	◯
0	◯	◯

Next →

education northwest

56980

56980

6. Which of the following core content subjects do you currently teach?
   (check all that apply, and write in other subjects under Other, if applicable)

   - ◯ Language Arts
   - ◯ Science
   - ◯ Mathematics
   - ◯ Social Studies
   - ◯ Other:

7. What is your highest educational degree completed?

   - ◯ Bachelor (B.S. or B.A.)
   - ◯ Masters (M.S., M.A., M.Ed.)
   - ◯ Doctorate (Ph.D, Ed.D, J.D)
   - ◯ Other: ____________________

Thank you.

Please return your completed questionnaire to the Education Northwest in the postage-paid envelope provided for your school.

education northwest

Below For Official Use Only:

1 2 3 4 5 6 7 8 9 0

56980

Appendix C. Missing data imputation procedures

Stata's MI IMPUTE command was used to impute missing data on the posttest score (877 students) and gender (319 students). For the 166 students who were enrolled in school at the time of the baseline measure but were absent from the pretesting, the pretest score was imputed. The pattern of missing data was not sequential (that is, nested). Consequently, imputation was done simultaneously on multiple variables, using the assumption of multivariate normal distribution.

Variables included in the imputation model were the raw pretest score, the number of attempted items for the pretest, the scale score for the pretest (pretest score to be used as a covariate in the analysis model), the raw posttest score, the number of attempted items for the posttest, the scale score for the posttest (posttest score to be used as the outcome in the analysis model), gender (dummy-coded), ethnicity (dummy-coded), and the design variable school identification (dummy-coded). Five sets of complete data were imputed separately for the treatment group and for the control group. Those 10 imputed datasets were then merged for the impact analysis. The imputation was done using the expectation maximization method to calculate the initial values, followed by an iterative procedure based on the Markov Chain Monte Carlo method. Proper convergence behavior was verified using a trace plot and an autocorrelation plot for the worst linear function.

Appendix D. Project CRISS full implementation model description

A "five-star school" description was used to measure full program implementation, using original developer materials and adding specificity to some areas. This description of full implementation was agreed upon by the developer and research team prior to the implementation of Project CRISS in the treatment schools.

The main components of full implementation from the *program delivery perspective* are presented below.

1. **Schools select a project local facilitator prior to starting Project CRISS.**
    - Schools are asked to select a local facilitator who is an experienced teacher or administrator and has time to help teachers apply Project CRISS learning principles and reading/writing strategies (such as a literacy coach, school/district curriculum specialist, reading resource teacher, assistant principal, or experienced content teacher).
    - The local facilitator is expected to have sufficient free time to spend approximately 5–6 hours per week to help and coach teachers in the implementation of learning principles and reading/writing strategies.
    - The local facilitator is expected to participate in three Project CRISS training-of-trainer activities—observe a Level I training session in another school, attend a Level II institute, and train new teachers under observation—in order to become a district-certified trainer.
2. **Schools receive the following materials in year 1:**
    - Each teacher receives a 300-page manual (updated in 2004) used for Level I training and as resource book. The manual includes extensive examples of Project CRISS applications in different content areas and grade levels.
    - Four copies of the administrator guide are provided for the principal and assistant administrators to help them monitor and provide support for Project CRISS implementation.
3. **Schools receive a total of six days of training/technical assistance in year 1.**
    - At the beginning of the school year, a two-day initial Level I workshop (12 contact hours) is provided by the national trainer for all core-subject teachers, local facilitator, and principal. Core subjects are English/language arts, mathematics, science, and social studies. Project CRISS was offered to core subject teachers in all grades.
    - In the fall, the same national trainer (who stays with the school throughout the course of the program) visits for one day (6 contact hours) in order to complete the Level I training.
    - In the spring, a one-day session (approximately 6 contact hours) is held for all teachers for teacher support and lesson plan development around Project CRISS learning principles and strategies.

- During the course of the school year, trainers spend two additional days on site for technical assistance (dates optional but typically occurring in winter and spring). This includes working with the local facilitator and principal to model a classroom walkthrough process, assessing progress, and pinpointing challenge areas to improve practice.

4. **Schools receive the following materials, training, and support in year 2:**
    - Each teacher receives a copy of *Cornerstones,* a more advanced manual with 15 team exercises that help teachers collaborate and deepen their practice around Project CRISS learning principles and reading/writing strategies.
    - One day of training (6 contact hours) is conducted for all teachers at the beginning of the school year to introduce *Cornerstones* and its use and to review key concepts from Level I training.
    - If there are new teachers in the school, two or three days of training are provided for all new teachers on Level I content. (If the group is small, the trainer attempts to cover all the Level I content for new teachers in two days.)
    - One day of additional technical assistance is provided; this is flexible technical assistance time with no training agenda. It is intended to provide further teacher support in areas such as developing lesson plans or helping with challenging implementation areas.

In addition to these planned activities directed by a national trainer working cooperatively with the school principal and faculty, full implementation includes expectations for local facilitator activity to support teachers and teacher participation in these activities. These activities are summarized below.

1. **Local facilitator activities during year 1.**
    - Conduct approximately one whole-group meeting with all content teachers, plus one small-group meeting (for example, by academic department) per month to maintain regular support for teachers.
    - Visit and observe teachers in classrooms as needed, including peer observation or coaching on the learning principles and reading/writing strategies.
    - Conduct some walkthrough protocols with the principal.
    - Be available for scheduled telephone or email contacts with the national trainer or initiate contact as needed to troubleshoot implementation challenges.
    - Observe a Level I training and attend a summer Project CRISS Level II regional institute; these are required to become district-certified trainer.
2. **Local facilitator activities during year 2.**
    - Conduct the Level I training for new teachers at the beginning of the school year under observation of the national trainer. This is the third and final step to become a district-certified Project CRISS trainer.
    - Use *Cornerstones* as a structure for regular teacher study groups in order to deepen understanding and use of Project CRISS, including holding meetings and visiting classrooms as needed.
    - Contact the national trainer by phone and email as needed for advice or assistance.

Appendix E. Complete multilevel model results for impact analysis

The model parameter estimates are based on the pooling of five sets of parameter estimates, each arrived at by the application of a linear mixed model to one of the five multiply-imputed datasets. Linear mixed models are based on maximum likelihood estimation.

Table E1 presents the results of pooled parameter estimates for the impact analysis. The confirmatory question was: what impact does Project CRISS have on the reading comprehension of grade 9 students in high schools in rural and town locales in Northwest Region states?

Table E1. Pooled linear mixed model estimates for the impact assessment of Project CRISS on student reading comprehension

Variable	Coefficient	Standard error	*t*	*p*	Degree of freedom
Fixed effects					
γ_{000}: adjusted grand mean for the control condition	691.8916	2.078767	332.84	0.000	681.1
γ_{010}: adjusted average treatment effect for Project CRISS	2.154223	2.804438	0.77	0.443	494.7
γ_{100}: average effect of pretest score on student outcome	0.6576527	0.0116619	56.39	0.000	167.7
γ_{001}: adjusted effect of block 2	4.16225	5.999848	0.69	0.488	12,736.8
γ_{002}: adjusted effect of block 3	0.3446712	8.79577	0.04	0.969	2,224.2
γ_{003}: adjusted effect of block 4	3.900254	6.441493	0.61	0.545	749.0
γ_{004}: adjusted effect of block 5	7.305791	4.393063	1.66	0.107	28.7
γ_{005}: adjusted effect of block 6	6.842049	9.002859	0.76	0.447	3,174.6
γ_{006}: adjusted effect of block 7	0.0493389	8.671661	0.01	0.995	70,539.5
γ_{007}: adjusted effect of block 8	–9.125083	5.077015	–1.80	0.072	170,511.6
γ_{008}: adjusted effect of block 9	–5.606609	5.039964	–1.11	0.266	1,158.8
γ_{009}: adjusted effect of block 10	–5.45257	5.200264	–1.05	0.295	490.4
γ_{0010}: adjusted effect of block 11	–3.632671	6.200963	–0.59	0.558	844.6
γ_{0011}: adjusted effect of block 12	3.840294	3.489978	1.10	0.271	1,252,133.5
γ_{0012}: adjusted effect of block 13	–3.895227	8.790258	–0.44	0.658	67,058.4
γ_{0013}: adjusted effect of block 14	–0.8291687	5.116592	–0.16	0.871	4,514.1
γ_{0014}: adjusted effect of block 15	2.439715	4.440305	0.55	0.583	14,820.9
γ_{0015}: adjusted effect of block 16	–2.132885	5.01863	–0.42	0.671	3,672.6
γ_{0016}: adjusted effect of block 17	3.225902	6.224121	0.52	0.604	3,241.1
Random effects					
r_{jk}: school random effect	8.148972	1.32662	[a]	[a]	16.2
e_{ijk}: student residual	30.62262	0.3785825	[a]	[a]	35.9

a. Not calculated.

Source: Authors' analysis based on data described in text.

Table E2 presents information concerning the pooling of parameter estimates. Specifically, it contains information on the following, for each estimated parameter: the variance information for the pooling (columns 1–3); relative variance increase, which represents the variance increase due to nonresponse; fractions of missing information, which represents the amount of information lost due to nonresponse; and relative efficiency, which represents the efficiency of using the finite (in this case, five) as opposed to the infinite set of imputed data.

Table E2. Pooling of parameter estimates for the impact assessment of Project CRISS on student reading comprehension

	Within-imputation variance	Between-imputation variance	Total variance	Relative variance increase	Fractions of missing information	Relative efficiency
Fixed effects						
γ_{000}: adjusted grand mean for the control condition	3.9901	0.275972	4.32127	0.82997	0.79336	0.984381
γ_{010}: adjusted average treatment effect for Project CRISS	7.15764	0.589359	7.86487	0.098808	0.09358	0.981628
γ_{100}: average effect of pretest score on student outcome	0.000115	0.000018	0.000136	0.182619	0.164324	0.968181
γ_{001}: adjusted effect of block 2	35.3602	0.531618	35.9982	0.018041	0.017876	0.996438
γ_{002}: adjusted effect of block 3	74.0847	2.7341	77.3656	0.044286	0.043268	0.991421
γ_{003}: adjusted effect of block 4	38.4607	2.52681	41.4928	0.078838	0.075542	0.985116
γ_{004}: adjusted effect of block 5	12.0984	6.00051	19.299	0.595172	0.412618	0.923767
γ_{005}: adjusted effect of block 6	78.1744	2.39755	81.0515	0.036803	0.036104	0.992831
γ_{006}: adjusted effect of block 7	74.6314	0.471886	75.1977	0.007587	0.007558	0.998491
γ_{007}: adjusted effect of block 8	25.6512	0.104037	25.7761	0.004867	0.004855	0.999030
γ_{008}: adjusted effect of block 9	23.9088	1.24366	25.4012	0.06242	0.060373	0.988069
γ_{009}: adjusted effect of block 10	24.6004	2.03531	27.0427	0.099282	0.094003	0.981546
γ_{0010}: adjusted effect of block 11	35.8058	2.20514	38.4519	0.073903	0.071015	0.985996
γ_{0011}: adjusted effect of block 12	12.1582	0.018141	12.1799	0.001791	0.001789	0.999642
γ_{0012}: adjusted effect of block 13	76.6719	0.497308	77.2686	0.007783	0.007753	0.998452
γ_{0013}: adjusted effect of block 14	25.4002	0.64942	26.1795	0.030681	0.030197	0.993997
γ_{0014}: adjusted effect of block 15	19.3924	0.269921	19.7163	0.016703	0.016561	0.996699
γ_{0015}: adjusted effect of block 16	24.3554	0.69268	25.1866	0.034129	0.033528	0.993339
γ_{0016}: adjusted effect of block 17	37.3787	1.13412	38.7397	0.03641	0.035725	0.992906
Random effects						
r_{jk}: School random effect	0.885334	0.728822	1.75992	0.98786	0.549355	0.901006
e_{ijk}: Student residual	0.095468	0.039881	0.143325	0.501291	0.368174	0.931415

Source: Authors' analysis based on data described in text.

Endnotes

[1] For school size, measured as total student enrollment in fall 2005, the mean for the three dropout schools was 613.33 (s.d. = 372.87) compared to 480.20 (s.d. = 165.74) for the 49 remaining schools in the original sample, a difference that was nonsignificant ($t = 1.25$, $df = 50$, p-value = 0.22). For the school poverty rate, measured as the U.S. Census percentage of families living at or below poverty in the district geographic area, the mean for the three dropout schools was 0.16 (s.d. = 0.05) compared to 0.15 (s.d. = 0.06) for the 49 remaining schools in the original sample, a difference that was nonsignificant ($t = 0.24$, $df = 50$, p-value = 0.81).

[2] The original design included classroom observational measures in a one-third subsample of schools based on a validated instrument called the Vermont Classroom Observation Tool (VCOT), which measured the application of learning principles similar to the research-based principles in Project CRISS. (See Saginor, 2008—developer of the VCOT and consultant for adapting this instrument to Project CRISS—for a fuller description of the diagnostic classroom observations that were initially contemplated.) The observations were intended as an implementation assessment of classroom practices, in both the treatment and control conditions, that mirror the Project CRISS principles and instructional strategies. After one round of observations, this part of the design was dropped due to high costs and because of restricted variance on the VCOT scales, indicating an inability of the measure to discriminate across different teachers and classrooms. Also included in the original design were principal and teacher questionnaire items in control schools to provide additional treatment-contrast data on professional development activities. These were dropped because of low principal return rates in the first round and because the questions by themselves were vague and did not provide reliable data on counterfactual conditions. In the end, the decision was made to describe implementation in comparison to ideal implementation of Project CRISS rather than use treatment-contrast variables from the control schools.

[3] As a condition for participating in the study, schools had to state at the time of recruitment that they had not received any services from Project CRISS of Lifelong Learning, Inc., during the past five years. This helped ensure that schools selected as controls would have no recent history of Project CRISS services.

[4] The actual number of months for the local facilitator log varied across schools between 8 and 10 months. The local facilitator log was initiated during the first full month following Level I training and the last month of the school year, which was usually June and usually less than a full month. The return rates for local facilitator logs were 92 percent of the monthly logs completed for year 1 and 87 percent completed for year 2.

References

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Barnett, L.G., and Greenough, R. (2004). *Regional needs assessment 2004: northwest educators' priorities for improving low-performing schools*. Portland, OR: Regional Educational Laboratory Northwest.

Barnett, L.G., and Greenough, R. (2005). *Regional needs assessment 2005: identifying priorities for improving low-performing schools*. Portland, OR: Regional Educational Laboratory Northwest.

Biancarosa, G., and Snow, C.E. (2004). *Reading next: a vision for action and research in middle and high school literacy. A report from Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education.

Bloom, B.S., and Broder, L.J. (1950). *Problem solving process of college students: an exploratory investigation* (Supplementary Educational Monograph 73). Chicago, IL: University of Chicago Press.

Bloom, H.S. (2005). Randomizing groups to evaluate place-based programs. In H.S. Bloom (Ed.), *Learning more from social experiments: evolving analytic approaches*. New York, NY: Russell Sage.

Brown, R., El-Dinary, P.B., Pressley, M., and Coy-Ogan, L. (1995). A transactional strategies approach to reading instruction**.** *Reading Teacher, 49*(3), 256–258.

Brown, R., Pressley, M., Van Meter, P., and Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second grade readers. *Journal of Educational Psychology, 88*(1), 18–37.

Bruner, J.S. (1977). *The process of education*. Cambridge, MA: Harvard University Press.

Duffy, G., Roehler, L., Sivan, E., Rackliffe, G., Book, C., Meloth, M., Vavrus, L., Wesselman, R., Putman, J., and Basiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, *22*(3), 347–368.

Duke, N.K., and Pearson, P.D. (2002). Effective practices for developing reading comprehension. In A.E. Farstrup and S.J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed.). Newark, DE: International Reading Association.

Gilmore Research Group. (2006). *Summary of proceedings from the 2006 state education forums*. Portland, OR: Regional Educational Laboratory Northwest.

Kamil, M.L., Borman, G.D., Dole, J., Kral, C.C., Salinger, T., and Torgensen, J. (2008). *Improving adolescent literacy: effective classroom intervention practices* (IES Practice Guide, NCEE 2008-4027). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved September 24, 2010, from http://ies.ed.gov/ncee/wwc/pdf/practiceguides/adlit_pg_082608.pdf

Karlsen, B., and Gardner, E.F. (1996). *Stanford Diagnostic Reading Test, fourth edition (SDRT 4): multilevel norms book/technical information*. San Antonio, TX: Harcourt Brace Educational Measurement.

Konstantopoulos, S., and Hedges, L.V. (2008). How large an effect can we expect from school reforms? *Teachers College Record, 110*(8), 1611–1638.

Palinscar, A.S., and Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117–175.

Paris, S.G., Lipson, M.Y., and Wixson, K.K. (1994). Becoming a strategic reader. In R.B. Ruddell, M.R. Ruddell, and H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed.). Newark, DE: International Reading Association.

Pearson, P.D., and Fielding, L. (1991). Comprehension instruction. In R. Barr, M.L. Kamil, P.B. Mosenthal, and P.D. Pearson (Eds.), *Handbook of reading research* (Vol. 2). New York, NY: Longman.

Pressley, M. (2002). Metacognition and self-regulated comprehension. In A.E. Farstrup and S.J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed.). Newark, DE: International Reading Association.

Raudenbush, S.W., Spybrook, J., Congdon, R., Liu, X., and Martinez, A. (2009). *Optimal Design Software for Multi-level and Longitudinal Research* (Version 2.0) [Software]. Retrieved September 24, 2010, from http://sitemaker.umich.edu/group-based/optimal_design_software

Santa, C.M. (2004). *Project CRISS: evidence of effectiveness*. Unpublished manuscript.

Santa, C.M., Havens, L.T., and Valdes, B.J. (2004). *Project CRISS: creating independence through student-owned strategies* (3rd ed.). Dubuque, IA: Kendall/Hunt.

Saginor, N. (2008). *Diagnostic classroom observation: moving beyond best practice.* Thousand Oaks, CA: Corwin Press.

Simon, H.A. (1979). *Models of thought*. New Haven, CT: Yale University Press.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2008). *What Works Clearinghouse evidence standards for reviewing studies* (Rev. ed.). Retrieved September 24, 2010, from http://ies.ed.gov/ncee/wwc/pdf/study_standards_final.pdf

U.S. Department of Education, National Center for Education Statistics. (2009). *The Nation's Report Card: reading 2009. National Assessment of Educational Progress at grades 4 and 8*

(NCES 2010-458). Retrieved September 24, 2010, from http://nces.ed.gov/nationsreportcard/pdf/main2009/2010458.pdf

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2010). *What Works Clearinghouse intervention report: Project CRISS (Creating Independence through Student-owned Strategies.* Retrieved September 27, 2010, from http://ies.ed.gov/ncee/wwc/pdf/wwc_projectcriss_061510.pdf

Vygotsky, L.S. (1978). *Mind in society: the development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Yoon, K.S., Duncan, T., Lee, S.W.-Y., Scarloss, B., and Shapley, K.L. (2007). *Reviewing the evidence of how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007-033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved February 25, 2011, from http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf