NCEE 2011–4001

U.S. DEPARTMENT OF EDUCATION

The Impact of Collaborative Strategic Reading on the Reading Comprehension of Grade 5 Students in Linguistically Diverse Schools

Final Report

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

REL SOUTHWEST
Regional Educational Laboratory

REL SOUTHWEST
Regional Educational Laboratory
At Edvance Research

The Impact of Collaborative Strategic Reading on the Reading Comprehension of Grade 5 Students in Linguistically Diverse Schools

March 2011

Authors:

John Hitchcock, Ph.D., ICF International and Ohio University

Joseph Dimino, Ph.D., Instructional Research Group

Anja Kurki, Ph.D., American Institutes for Research

Chuck Wilkins, Ph.D., Edvance Research, Inc.

Russell Gersten, Ph.D., Instructional Research Group

Project Officer:

Karen Armstrong
Institute of Education Sciences

NCEE 2011–4001
U.S. Department of Education

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

U.S. Department of Education

Arne Duncan

Secretary

Institute of Education Sciences

John Q. Easton

Director

National Center for Education Evaluation and Regional Assistance

Rebecca A. Maynard

Commissioner

March 2011

This report was prepared for the National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, under contract ED-06C0-0017 with Regional Educational Laboratory Southwest administered by Edvance Research.

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 conflicts of interest

Dr. Joseph Dimino is a coauthor of Collaborative Strategic Reading (CSR), the intervention being evaluated in this study.[1] His sole involvement in the study was to train and monitor coaches and teachers in CSR reading intervention strategies. He was not involved in any data collection or data analysis activities during the study and did not have access to student data. Any royalties he might have received for the use of the intervention in the study were far below the $10,000 limit conventionally used by research organizations such as universities.

The Regional Educational Laboratory Southwest (REL Southwest) contracted with individuals and entities to obtain expert advice and technical assistance for this study. Although these contractors' professional work may not be entirely independent of or separable from the tasks they carried out for REL Southwest, no one from the Instructional Research Group, ICF International, or American Institutes for Research who was associated with this study has financial interests that could be affected by the content of this report.

[1] Contractors carrying out research and evaluation projects for the Institute for Education Studies (IES) frequently need to obtain expert advice and technical assistance from individuals and entities whose other 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.

Acknowledgments

This study was a collaborative effort by many individuals and organizations worthy of acknowledgments. The study authors would like to thank all the district personnel, principals, teachers, and students for their patience during the evaluation and for allowing access to their classrooms several times during the school year. Without their participation, this evaluation would not have been possible.

Through the involvement and guidance of a number of technical experts, the study team was able to clarify and solve many complex technical and analytical questions. The team of experts included individuals from the Institute of Education Sciences (IES), IES's technical advisory group, and David Francis, David Myers, David Chard, Jeremy Kilpatrick, experts for the study's technical working group.

The study authors would also like to thank the many individuals who participated in the study. Gay Lamey, project manager, adeptly directed and participated in all aspects of the study from initial study planning to final report preparation. Norma Reyna, project coordinator, oversaw site-management activities and provided a wide range of assistance throughout the evaluation. Laural Logan-Fain led recruitment efforts during some of the most difficult and challenging situations, such as natural disasters, with remarkable persistence and patience. Meg Grant contributed to the recruitment activities and directed all study dissemination efforts with insightful understanding of constituents. Debrale Graywolf provided comprehensive and essential editorial services and guidance. Lois Gregory, Noelle Howland, and Herb Turner of Analytica, Inc. provided expert quality review and invaluable advice throughout report preparation and multiple review cycles. Mary Jo Taylor, JoAnne Sapp, and Sue Buckley provided perceptive and specialized coaching. Lindsay Anderson, Yibing Li, Anju Sidana, Sarah Steckley, and Monika Townsend of American Institutes for Research showed great flexibility and contributed greatly to student data collection, data processing, and report preparation.

TABLE OF CONTENTS

List of Tables

List of Figures

Abbreviations

CSR	Collaborative Strategic Reading
CSRIVC	Collaborative Strategic Reading Intervention Validity Checklist
ELL	English language learner
ERC	Expository Reading Comprehension observation instrument
FC–ELL	Former and current English language learner
GRADE	Group Reading Assessment and Diagnostic Evaluation
HLM	Hierarchical linear model
IES	Institute of Education Sciences
IVEware	Imputation and Variance Estimation Software
MAR	Missing at random
MCAR	Missing completely at random
RCT	Randomized controlled trial
REL	Regional Educational Laboratory

Glossary

Click and Clunk. One of four primary reading comprehension strategies used in Collaborative Strategic Reading (CSR). The strategy is designed to help students monitor their comprehension by identifying difficult vocabulary as they read and determining the meaning of words they do not understand. Students are taught that when they understand what they are reading, everything *clicks*. A *clunk* is a challenging or unknown word or concept that interrupts their comprehension. (See also *Get the Gist, Preview,* and *Wrap up.*)

Collaborative Strategic Reading. Set of instructional strategies designed to improve the reading comprehension of students with diverse abilities.

Cooperative learning groups. Small instructional groups in which students take on some instructional tasks and learning is considered interdependent (that is, students take on some responsibility for teaching aspects of the curricula to one another).

Departmentalized instruction. School structure in which teachers specialize in subject areas and take on all related teaching tasks. Departmentalized instruction is common in high schools (which have teachers for science, math, social studies, and so forth). In earlier grades a single teacher is usually expected to provide instruction across all curricula.

Get the Gist. One of four primary reading comprehension strategies used in CSR. The strategy teaches students to compose a main idea conveying the essence of a paragraph or short segment of the text. (See also *Click and Clunk, Preview,* and *Wrap up.*)

Metacognitive knowledge. Reader's ability to understand and use the strategies they need to successfully comprehend text.

Preview. One of four primary reading comprehension strategies used in CSR. The strategy requires students to scan text features such as the title, subheadings, and illustrations to identify key words and terms; brainstorm prior knowledge about the topic, such as information acquired from previous lessons or from watching a movie or television program; record brainstorming results in learning logs (journals students keep to record what they learn during the CSR process); and predict what they will learn from reading the selection by generating and recording specific statements in their learning logs. (See also *Click and Clunk, Get the Gist,* and *Wrap up.*)

Reciprocal teaching. Instructional activity that takes place in the form of a dialogue between teacher and students. Reciprocal teaching includes four components: previewing text to obtain a sense of what will be learned when fully immersed in reading, generating questions for oneself about what the text is attempting to convey, clarifying unclear information, and summarizing main points. Students work through these four strategies in groups of 10–12.

Scaffolded instruction. Approach to instruction that provides support and reinforcement for emerging skills. With a student who cannot complete a task independently, the teacher builds in prompts, reminders, and reinforcements to help students master a given skill.

Self-regulatory knowledge. Degree to which readers monitor their understanding of a concept, evaluate, and self-correct.

Think aloud. Teachers verbalize aloud each facet of a comprehension strategy by explaining, step by step, the thought processes they used to generate a main idea, determine the meaning of the word through context, and so forth.

Whole class instruction. Instruction that entails presenting curricula to an entire class at once.

Wrap up. One of four primary reading comprehension strategies used in CSR. The strategy is designed to teach students to identify the most important ideas in a passage and help them understand and remember what they have learned by developing and answering questions and writing a summary. After students have read the text passage, they write questions in their learning logs and then take turns asking and answering them before writing a summary of the passage they have read. (See also *Click and Clunk, Get the Gist,* and *Preview*.)

Zone of proximal development. Point at which a student can apply a skill with teacher support.

Executive summary

Recent findings from an expert panel of reading researchers noted that approximately 8 million adolescents struggle with literacy in middle and high school (Biancarosa and Snow 2006); the "most common problem is that they are not able to comprehend what they read" (p. 3). Before the 1980s, teachers rarely taught reading comprehension (Carlisle and Rice 2002; Durkin 1978). However, over the last 20 years, a large body of research emerged on methods for explicitly teaching reading comprehension to students in the upper elementary grades (Carlisle and Rice 2002). The goal is to teach students to learn from text—to discern which information is critical, integrate such information with what is already known, and draw valid inferences.

Collaborative Strategic Reading (CSR) is a set of instructional strategies designed to improve the reading comprehension of students with diverse abilities (Klingner and Vaughn 1996). Teachers implement CSR at the classroom level using scaffolded instruction to guide students in the independent use of four comprehension strategies; students apply the strategies to informational text while working in small cooperative learning groups. The goals are to improve reading comprehension and conceptual learning so that academic performance also improves. Because CSR involves changes to teachers' instructional practices, regardless of subject matter, it can be used with a variety of curricula and in a variety of settings.

CSR has been implemented in a number of states[1] and has shown the potential to improve reading comprehension for both English language learner (ELL) students and non–ELL students (Klingner, Vaughn, and Schumm 1998). Klingner and Vaughn (2000) suggest that CSR is effective with ELL students because the peer interaction that occurs during cooperative learning is intended to increase students' opportunity to discuss informational text in a non-threatening, low anxiety atmosphere. Cooperative learning permits linguistically diverse students to take advantage of support in their native language from their peers who are bilingual (Cohen 1994a; Garcia 1994; Jacob et al., 1996; Klingner and Vaughn 2000). Research in this area indicates that students who were afforded the opportunity to practice their second language in classrooms that employed cooperative learning demonstrated a broader array of language functions than students in classrooms that were predominantly teacher directed (Long et al. 1976; Long and Porter 1985; Pica et al. 1996).

Prior research on CSR, however, has used only quasi-experimental designs to assess impact, and observed reading skill improvement in those studies may be difficult to attribute to CSR. Because limited English language skills put ELL students at risk for developing academic difficulties (August and Shanahan 2006; Gersten 1996; Moss and Puma 1995), and because this population represents a growing segment of students across the country and in the Southwest Region (National Center for Education Statistics 2004; National Clearinghouse for English Language Acquisition 2008; Ruiz-de-Velasco and Fix 2000), it is important to evaluate whether CSR could be an effective means of improving student reading comprehension in linguistically diverse education contexts.

[1] CSR has been implemented in parts of California, Colorado, Florida, Iowa, Michigan, Montana, New York, Pennsylvania, Texas, and Wisconsin (as part of a statewide initiative).

The current study is a randomized controlled trial (RCT) examining the effect of CSR on student reading comprehension. Within each participating linguistically diverse school, grade 5 social studies classrooms were randomly assigned to either the CSR condition (using CSR when delivering social studies curricula) or to the control condition (a business-as-usual condition). The implementation period was one school year.

This study focused on the following confirmatory research question:

- In linguistically diverse schools, do grade 5 students in CSR classrooms have higher average reading comprehension posttest scores on the Group Reading Assessment and Diagnostic Evaluation (GRADE) than students in control classrooms?

In addition, the study examined three exploratory research questions about CSR's effect on two subgroups of students:

- Do grade 5 former and current ELL (FC–ELL) students in CSR classrooms have higher average reading comprehension posttest scores on the GRADE than FC–ELL students in control classrooms?
- Do grade 5 non–ELL students in CSR classrooms have higher average reading comprehension posttest scores on the GRADE than non–ELL students in control classrooms?
- Does CSR have a differential impact on GRADE reading comprehension posttest scores for grade 5 FC–ELL and non–ELL students?

The intent of these exploratory analyses was to examine whether there is an effect for each subgroup separately, as well as whether there is a differential effect between the subgroups.

Study sample

Recruitment for the study focused on large urban and suburban districts that serve large numbers of ELL students (25 percent or more) in the Southwest Region. Districts serving large numbers of ELL students were targeted to obtain linguistically diverse schools to address the confirmatory research question and a large enough sample of ELL students to address the exploratory research questions. The final analytic sample included 74 classrooms (37 CSR, 37 control) across 26 schools and 5 districts in Oklahoma and Texas. Parent permission was required for students to participate in data collection for this study, and the final analytic sample included 1,355 students (681 CSR, 674 control).

CSR teacher training

This study is an effectiveness trial designed to evaluate CSR as implemented by classroom teachers. In the previous efficacy trial, the developers of CSR delivered the intervention directly to students in their classrooms (Klingner, Vaughn, and Schumm 1998). However, this method of delivery would not be feasible for a large number of classrooms, such as would be required in a district-wide adoption of CSR. Therefore, in this study a CSR codeveloper, Dr. Joseph Dimino, provided an initial two-day training session to teachers; teachers then delivered CSR to students in the classroom. Dr. Dimino also trained coaches,

who provided four follow-up coaching sessions to the teachers throughout the year. CSR training included theoretical foundations of CSR strategies, CSR comprehension strategies, strategies for efficiently integrating CSR into social studies lessons, and logistical issues (forming small groups, behavior management, and paperwork).

Data analysis

Students were pretested prior to implementation and posttested at the end of the school year using the GRADE, a nationally normed and standardized measure of reading comprehension. GRADE was chosen because, like CSR, it focuses on the vocabulary and reading comprehension skills required to understand text, as well as skills such as questioning, predicting, clarifying, and summarizing. CSR's effect on student performance on the reading outcome measure was estimated by comparing students in CSR classrooms with their counterparts in control classrooms within schools, while controlling for student and teacher baseline characteristics. A treatment effect was estimated for each school and the overall effect of CSR was calculated. The analyses were conducted using hierarchical linear modeling, an approach that accounts statistically for the nested data in this study (students nested in classrooms, nested within schools). Students who were enrolled in study classrooms at baseline but had missing pretest or posttest data were included in the analytic sample through the use of multiple imputation, an approach used to correct for missing data.

The effect of CSR on reading comprehension

The primary finding of this study is that CSR did not have a statistically significant impact on student reading comprehension. Nine sensitivity analyses—including alternative statistical approaches, an alternative approach for handling missing data, and different sample specifications—showed that the findings were robust to different analytic approaches.

Three exploratory analyses were also conducted to examine the effects of CSR on FC–ELL and non–ELL students. Statistically significant effects on student reading comprehension were not identified for either subgroup, and no statistically significant differential impacts were identified.

It is often the case that RCTs, because of their greater rigor, do not support the findings of prior quasi-experiments (Glazerman, Levy, and Myers 2002, 2003). With all other design features held constant, randomization yields stronger evidence about program impacts than do quasi-experiments (Boruch 1997; Shadish, Cook, and Campbell 2002).

The current investigation evaluated the impact of CSR in an effectiveness trial designed to approximate a district's implementation of CSR. Data on the fidelity of implementation suggest that professional development was generally delivered according to plan. Data on teacher fidelity of CSR implementation showed that 78.8 percent of teachers reported using CSR two or more times a week, as instructed. However, the single observation conducted for each classroom found that 21.6 percent of CSR teachers were using all five core teacher strategies, which the study defined as full procedural fidelity; 56.8 percent of teachers were observed using three or fewer strategies.

Limitations and future research

This study used a convenience sample of volunteer schools. Its findings apply only to schools in the study sample. A second limitation is that language status for participating students was limited to FC-ELL or non-ELL identification because other language status data—such as level of English proficiency, first language, or any specific ELL programs in which students participated—were not collected. Further limitations are that procedural fidelity was assessed based on a single observation, and that the procedural fidelity instrument used measures whether CSR activities were observed but not how well teachers or students implement the procedures. The impacts observed in this study for varied CSR implementation by study teachers within the schools were not statistically significant. Future research might focus on methods for enhancing CSR implementation within classrooms to determine whether impacts would be larger and statistically significant due to greater fidelity. In addition, future investigations could also consider using more intensive training and/or coaching delivery and investigating CSR impact at different grade levels and in different subject areas.

Chapter 1. Introduction and study overview

This chapter introduces the issues and context that motivate this study. It reviews the relevant literature, overviews the study, and poses the research questions examined.

Motivation for study

About 8 million adolescents struggle with literacy in middle and high school, according to an expert panel of reading researchers (Biancarosa and Snow 2006). "Very few of these older struggling readers need help to read the words on a page; their most common problem is that they are not able to comprehend what they read . . . Obviously, the challenge is not a small one" (p. 3). Until the 1980s, teachers rarely taught reading comprehension (Carlisle and Rice 2002; Durkin 1978). Over the next 20 years, a large body of research emerged on methods for explicitly teaching reading comprehension to students in the upper elementary grades (Carlisle and Rice 2002). The goal of these methods is to teach students to learn from text—to discern which information is critical, integrate such information with what is already known, and draw valid inferences.

Explicit strategy instruction is one set of practices supported by this body of rigorous, albeit often small-scale, efficacy research (Pearson and Dole 1987; Pressley et al. 1992).[3] Explicit strategy instruction is the intentional design and delivery of information by the teacher to students. The teacher models or demonstrates a skill or strategy and provides students with the opportunity to practice and apply the newly learned skills and obtain feedback (National Institute for Literacy 2008).

Research conducted in the 1980s and 1990s supported the utility of explicit strategy instruction (Pearson and Dole 1987; Pressley et al. 1992). A recent large-scale evaluation found that implementation is often weak and erratic, however (James-Burdumy et al. 2009). Teachers appeared reluctant to think aloud for students, model use of comprehension strategies, and encourage students to use inferential clues. Classroom observations indicate that on average teachers used targeted strategies less than once during 10-minute observation intervals (James-Burdumy et al. 2009).

The importance of effective reading comprehension instruction is compounded by the linguistic diversity of many schools, which are composed of both English language learner (ELL) students and non–English language learner (non–ELL) students. Instructional strategies must be responsive to this diversity (Short and Fitzsimmons 2007; Willis 2000). According to results from the *2003–2004 Schools and Staffing Survey* (National Center for Education Statistics 2004), 91.6 percent of schools serving ELL students teach them in the same classrooms as native English speakers, especially in the upper grade levels, when many former ELL students no longer qualify for special services but still require additional support in language (Carlo et al.

[3] Efficacy research attempts to test intervention impacts assuming optimal implementation circumstances. Pearson and Dole (1987) and Pressley et al. (1992) synthesize relevant literature on explicit strategy instruction.

2004). The trend of growing numbers of linguistically diverse schools can be seen across the nation (National Center for Education Statistics 2004).

The five-state area (Arkansas, Louisiana, New Mexico, Oklahoma, Texas) served by the Regional Educational Laboratory Southwest (REL Southwest) expressed a need for more research on the ELL population (Logan-Fain et al. 2008). Four of the five states (all but Louisiana) have ELL enrollment rates that exceed the national average (National Center for Education Statistics n.d.); Texas has the second highest ELL enrollment in the country, after California (National Center for Education Statistics n.d.).

Students who are—or recently were—classified as ELL face a daunting challenge in school; they are expected to master the "double demands" of learning grade-level material and developing proficiency in English (Gersten 1996; Short and Fitzsimmons 2007). Given the large number of students in linguistically diverse settings and the critical need to enhance the comprehension of all students, it would seem important to provide reading comprehension instruction that is effective with heterogeneous groups of ELL students and native English speakers. The National Literacy Panel on Language Minority Children and Youth (August and Shanahan 2006) concluded that non–native English speakers often acquire basic literacy skills (that is, word reading) at rates comparable to native English speakers, but not in the area of comprehension, for which depth of vocabulary knowledge and familiarity with syntax play a large role (Chiappe, Siegel, and Wade-Woolley 2002; Jiménez, García, and Pearson 1995). Grade 5 seems to be a critical year for this type of intervention, because teachers need to prepare students for the demands of middle school, where they are expected to independently acquire information from text (Biancarosa and Snow 2006).

Collaborative Strategic Reading (CSR) literature review

A variety of instructional strategies have been developed in response to the need for effective reading comprehension instruction (Klingner and Vaughn 1996; Palincsar and Brown 1984).[4] Collaborative Strategic Reading (CSR) is a set of instructional strategies designed to help students with diverse abilities acquire and practice comprehension strategies for use with informational text (Klingner and Vaughn 1996). CSR was adapted from reciprocal teaching, an instructional activity that involves a dialogue between teacher and students. In reciprocal teaching, teacher and students take turns assuming an instructional role in leading this dialogue (Palincsar 1986). Reciprocal teaching includes four components: previewing text to obtain a sense of what will be learned when fully immersed in reading, generating questions for oneself about what the text is attempting to convey, clarifying unclear information, and summarizing main points. Students work through these four strategies in groups of 10–12. A meta-analysis of studies related to reciprocal teaching observed positive median effects (Rosenshine and Meister 1994).[5]

---

[4] Examples of instructional strategies developed in response to the need for effective reading comprehension include, but are not limited to, explicit strategy instruction, teacher modeling and opportunities for student practice, summary writing, and think-aloud procedures, in which students make their thinking processes public (Gajria and Salvia 1992; Pearson and Dole 1987).

[5] Rosenshine and Meister (1994) reviewed 16 studies on reciprocal teaching. Eleven used standardized tests as an outcome measure, yielding a median effect size of 0.32; two of these studies identified statistically significant positive effects. Ten of the 16 studies used experimenter-developed tests as an outcome measure, yielding a median effect size of 0.88; eight of these studies identified statistically significant positive effects.

In addition to instructional strategies adapted from reciprocal teaching, CSR includes strategies such as cooperative learning, brainstorming, and student review of what has been learned from reading. (For a detailed discussion, see Klingner et al. 2001.) CSR's systematic set of procedures is designed to help students with diverse abilities acquire and practice comprehension strategies for use with informational text. CSR uses a mix of whole class instruction (instruction that entails presenting curricula to an entire class at once) and small cooperative learning groups, so that a teacher can work with an entire class at the same time. CSR, and the professional development provided to classroom teachers in conjunction with it, is designed to increase teachers' knowledge of reading comprehension and, consequently, affect teacher practice in the classroom. It uses explicit strategy instruction to teach metacognitive and self-monitoring skills that are expected to lead to improved reading comprehension. In theory, such knowledge should help students recognize whether they understand the information they read and take corrective steps when they do not. The intervention developers theorize that teacher and student use of CSR results in improved student reading comprehension, which in turn increases reading achievement. Because CSR involves changes to teachers' instructional practices, it can be used with a wide variety of curricula, regardless of subject matter.

When applying CSR, the classroom teacher initially presents and models the strategies to the entire class. As students become more proficient, they form cooperative learning groups of four to six students (with varying reading abilities) that work together to apply the reading comprehension strategies. Each student in the group is assigned a role, such as leader or timekeeper.

Many researchers suggest that cooperative learning formats may benefit ELL students (Calderón, Hertz-Lazarowitz, and Slavin 1998; Carlo et al. 2004; Saenz, Fuchs, and Fuchs 2005; Vaughn, Cirino et al. 2006; Vaughn, Mathes et al. 2006). Cooperative strategies provide students with an opportunity to talk to peers instead of teachers, and studies show ELL students often benefit from receiving bilingual support from fellow students while communicating in English. For example, Cummins (1984), Hakuta (1990), and Hudelson (1987) reported that comprehension of informational text increased when discussions in the student's native language were used to explain and clarify content. Further, cooperative learning permits linguistically diverse students to take advantage of support in their native language from their peers who are bilingual (Cohen 1994a; Garcia, 1994; Jacob et al. 1996; Klingner and Vaughn 2000). Research also indicates that students who were afforded the opportunity to practice their second language in classrooms that employed cooperative learning demonstrated a broader array of language functions than students in classrooms that were predominantly teacher directed (Long et al. 1976; Long and Porter 1985; Pica et al. 1996). Klingner and Vaughn (2000) suggest that CSR is effective with ELL students because the peer interaction that occurs during cooperative learning is intended to increase students' opportunity to discuss informational text in a non-threatening, low anxiety atmosphere.

CSR was piloted in linguistically diverse classrooms with both ELL and non–ELL students (Klingner and Vaughn 1996). It has been studied for more than a decade (Klingner and Vaughn 1996, 1998, 1999, 2000; Klingner, Vaughn, and Schumm 1998). Most of the research has been case study research without control groups.

One quasi-experimental design suggested that CSR has positive effects. Klingner, Vaughn, and Schumm (1998) explored the efficacy of CSR in five grade 4 social studies classrooms that included both ELL and non–ELL students. Researchers taught students in treatment classrooms (n = 85) to use CSR strategies while reading social studies texts; students in control classrooms

(n = 56) were not taught CSR strategies. Students in both groups received typical social studies instruction for 11 sessions lasting 45 minutes each. ELL students constituted 52 percent of students in treatment classrooms and 48 percent of students in control classrooms. The Gates-MacGinitie Reading Comprehension Test (MacGinitie et al. 2006) was used as both a pretest covariate and an outcome measure. Larger gains in reading comprehension scores among students in treatment classrooms were statistically significant (effect size = 0.43). Positive gains were found for both ELL and non–ELL students, with pretest to posttest change scores favoring CSR students (3.45 for ELL students, 2.22 for non–ELL students).[6]

The Klingner, Vaughn, and Schumm (1998) study used a quasi-experimental design, which does not provide internal validity as strong as that provided by randomized controlled trials (RCTs; Bloom 2005; Boruch 1997; Shadish, Cook, and Campbell 2002).[7] Groups are not formed by random assignment in a quasi-experimental design. Therefore, it is advisable to examine the intervention and control groups to determine whether they are sufficiently similar on observed characteristics. Examination of baseline scores from Klingner, Vaughn, and Schumm (1998) show a 0.11 standardized mean difference favoring the CSR group. Using What Works Clearinghouse (WWC) guidelines, the groups are sufficiently equivalent at baseline to yield a reasonable estimate of CSR's effects.[8]

Klingner, Vaughn, and Schumm (1998) suggest that CSR has positive effects in linguistically diverse classrooms serving both ELL and non–ELL students. Because their quasi-experimental study used a small sample and was conducted within a single school, it is unclear whether CSR is likely to produce a similar effect in wider settings. Moreover, the developers of CSR were directly involved in implementation and provided extensive ongoing support to the teachers, precluding generalization to school settings where support is more limited.

To provide teachers and principals with evidence on the effect of CSR on reading comprehension, a study would need to differ from the previous research in two ways. First, instead of a quasi-experimental design, an RCT would be required to provide more valid causal conclusions about the impact of CSR. Second, the study would need to be an effectiveness trial examining the use of CSR in school settings in which the level of support resembles that provided in most school settings.

Study description

The current study is a multidistrict cluster RCT designed to examine the impact of CSR on reading comprehension scores of grade 5 students in linguistically diverse schools. The study

[6] Standardized effects based on the change scores reported could not be independently estimated, because subgroup standard deviations were not provided as part of the study report (Klingner, Vaughn, and Schumm 1998).

[7] Internal validity is the degree to which one can be confident that a causal relationship exists between two variables (in this case, CSR and the observed impact favoring the CSR group).

[8] The WWC considers an effect size of 0.25 standard deviations in baseline difference between treatment and control groups as a cutoff for determining that a given quasi-experiment does not meet standards for internal causal validity. However, the WWC requires that a quasi-experiment be adjusted for baseline differences if the differences exceed 0.05 standard deviation. Klingner, Vaughn, and Schumm (1998) provide the details needed to calculate a difference-in-difference standardized effect (that is, the pretest to posttest change in the treatment minus the pretest to posttest change in the control divided by the pooled posttest standard deviation). This procedure yields a 0.32 effect size.

was implemented in 74 social studies classrooms in 26 schools across five school districts in Oklahoma and Texas.

Grade 5 was selected for two reasons. First, instruction at this grade tends to focus more on academic content than on basic skill development (Gersten 1996; Gersten and Baker 2000). Second, grade 5 is when ELL students usually transition from intensive bilingual support to English immersion classrooms with relatively little support. Social studies classrooms were selected because previous research (Klingner, Vaughn, and Schumm 1998) indicated that CSR could be used successfully in this content area.

In this study, random assignment was conducted within each participating school; therefore, within each school, grade 5 classrooms were randomly assigned to either the treatment condition (using CSR when delivering social studies curricula) or the control condition (a business-as-usual condition). Random assignment of classrooms provides the basis for interpreting observed posttest differences between CSR and control classrooms as reflecting the impact of CSR. To increase statistical power, students' pretest scores were included in the analysis model as a covariate. (Chapter 2 provides additional information about the analytic model used in this study.)

The teacher professional development provided in this study differs from that used in previous CSR efficacy research. Previous researchers (Klingner, Vaughn, and Schumm 1998; Klingner and Vaughn 2000), who were CSR developers, delivered the majority of intervention components. In this study, a CSR codeveloper trained teachers and coaches, who provided four follow-up coaching sessions with teachers during implementation. The CSR codeveloper did not deliver any intervention components to students. (Chapter 3 provides additional information regarding training and coaching.)

Research questions

This study was designed to answer one confirmatory and three exploratory research questions. The confirmatory question was:

- In linguistically diverse schools, do grade 5 students in CSR classrooms have higher average reading comprehension posttest scores on the Group Reading Assessment and Diagnostic Evaluation (GRADE) than students in control classrooms?

The three exploratory research questions considered the effect of CSR on two subgroups of students: former and current ELL (FC–ELL) students and non–ELL students. FC–ELL students were combined because school practices often require the continuation of language support or other academic support after students exit ELL services (Ragan and Lesaux 2006). FC–ELL students are defined as students currently identified by their schools as having limited English proficiency or identified as such at some point in the past and currently being monitored.[9] By contrast, non–ELL students in this study are students never identified by their schools as having limited English proficiency. This group includes former ELL students who are no longer being monitored.

[9] Schools in the study were drawn from two states (Oklahoma and Texas) that have different ELL exit policies. ELL classification and exit decisions are made primarily at the local level (Ragan and Lesaux 2006). Appendix A provides information on the ELL identification and exit procedures in both states.

This study answers the following exploratory research questions:

- Do grade 5 FC–ELL students in CSR classrooms have higher average reading comprehension posttest scores on the GRADE than FC–ELL students in control classrooms?
- Do grade 5 non-ELL students in CSR classrooms have higher average reading comprehension posttest scores on the GRADE than non-ELL students in control classrooms?
- Does CSR have a differential impact on GRADE reading comprehension posttest scores for grade 5 FC–ELL and non-ELL students?

Although the subgroup impacts examined in the first two exploratory questions are experimental (based on pre–random assignment characteristics), they are secondary to the confirmatory research question because their statistical power is weaker (see appendix B). In addition, only 23 of the 26 participating schools had FC–ELL and non–ELL students in both CSR and control classrooms. Only these 23 schools were included in the subgroup analyses.

Overview of report

Chapter 2 provides an overview of the study design and timeline. It covers such topics as randomization procedures, sample descriptions, data collection procedures, and analysis methods. It provides sufficient information for readers to assess internal and external validity for all analyses described in the report. Chapter 3 examines professional development and presents the data used to assess how well CSR was implemented. Related data and analyses are descriptive; they are presented to provide a context for interpreting the confirmatory findings. Chapter 4 presents the primary confirmatory results regarding the impact of CSR on student achievement outcomes. It also presents the exploratory results on the impact of CSR on FC–ELL and non–ELL student subgroups. Chapter 5 summarizes and discusses the study's findings, identifies the study's limitations, and provides suggestions for future research.

Chapter 2. Study design and methodology

This chapter presents an overview of the study design and timeline. It describes the study sample, data collection instruments, and the data analysis plan.

Overall design

This study is a multidistrict cluster RCT designed to evaluate the effect of CSR on reading comprehension in five school districts in Oklahoma and Texas. The final analytic sample used for the analysis of confirmatory impact included 1,355 students at 26 schools.

Because CSR is designed to be implemented as a classroom-level intervention, random assignment was conducted at the classroom level to ensure that the unit of randomization matched the unit of delivery. Grade 5 social studies classrooms were randomly assigned within linguistically diverse schools to either a treatment condition (using CSR when delivering social studies instruction) or a control condition (using the school's normal instructional procedures when delivering social studies instruction). Each participating school had at least one CSR classroom and one control classroom.

Students' baseline reading comprehension was evaluated using the GRADE.[10] An alternate form of the GRADE was used as an outcome measure to evaluate reading comprehension at the end of the study. Measures used to assess fidelity of implementation included fall and spring teacher surveys and the Collaborative Strategic Reading Intervention Validity Checklist (CSRIVC), an observational checklist. A subscale of the Expository Reading Comprehension (ERC) observation instrument, another observational checklist, was used to examine instructional practices in control classrooms.

The impact of CSR on student reading comprehension was evaluated using hierarchical linear modeling (HLM) (Raudenbush and Bryk 2002).[11] Because the random assignment of classrooms to treatment or control condition was done within schools, school-specific impact estimates were obtained and aggregated to provide the overall impact estimate. Missing data were dealt with using multiple imputation.

[10] The GRADE (American Guidance Service, Inc. 2001) was used in this study both to establish baseline reading achievement and to measure posttest reading achievement. It was chosen because, like CSR, it focuses on the vocabulary and reading comprehension skills required to understand text, as well as skills such as questioning, predicting, clarifying, and summarizing information. A single outcome measure was selected to minimize the testing time required of students and to avoid reduction in statistical power from correcting for multiple comparisons. Further, no measure of social studies knowledge was included due to the fact that the study was conducted across multiple districts and states, each using different curricula.

[11] Hierarchical linear modeling is a regression-based statistical analysis in which the nested structure of the subjects is accounted for appropriately (unlike in ordinary linear regression). In this study, because students are clustered in classrooms that are clustered in schools, hierarchical linear modeling was necessary to obtain correct statistical results.

Study timeline

The study was designed to evaluate the impact of a one-year implementation of CSR on student reading comprehension. However, to enable recruitment of a sample large enough to ensure sufficient statistical power, data were collected over two school years. Data from the two cohorts were combined to form the final study sample. Table 2-1 provides the study timeline for each of the five districts.

Table 2-1. CSR study timeline, by district

(completion dates)

Event	*District 1*	*District 2*	*District 3*	*District 4*	*District 5*
Classrooms randomly assigned	8/07	7/08	8/08	10/08	8/08
Parental consent forms collected	9/07	9/08	9/08	11/08	9/08
CSR professional development training provided to teachers	9/07	8/08	9/08	10/08	9/08
Pretest GRADE administered[a]	9/07	9/08	10/08	11/08	10/08
CSR implementation begun	10/07	10/08	10/08	10/08	10/08
Fall teacher survey administered	12/07	8/08	9/08	10/08	9/08
Follow-up coaching session 1 conducted	10/07	12/08	12/08	12/08	10/08
Follow-up coaching session 2 conducted	11/07	2/09	1/09	1/09	12/08
Follow-up coaching session 3 conducted	2/08	3/09	2/09	3/09	3/09
Follow-up coaching session 4 conducted	3/08	4/09	3/09	3/09	3/09
CSRIVC data collected	4/08	5/09	4/09	5/09	4/09
ERC observations conducted	4/08	5/09	5/09	5/09	5/09
Posttest GRADE administered	5/08	5/09	5/09	5/09	5/09
Spring teacher survey administered	5/08	5/09	5/09	5/09	5/09

a. CSR professional development training was provided to teachers before GRADE pretesting in all five districts. Teachers were instructed to delay implementation of CSR until after baseline data collection was complete. It is nevertheless possible that students could have been exposed to CSR before pretesting. In six schools implementation of CSR began two weeks before pretesting was completed for all students. Some students were therefore exposed to CSR before pretesting. Baseline equivalence analyses conducted for all five districts indicated that there were no statistically significant differences between CSR and control students for pretest GRADE scores.

Source: Authors' analysis of study records.

Study sample

A power analysis conducted for this study indicated that a sample of 26 schools was required for adequate statistical power, under the assumptions that there would be three grade 5 teachers per school and test data available for 16 students per classroom. (See appendix B for the full list of assumptions.) This section describes the eligibility criteria, the recruitment process, and randomization, as well as characteristics of participating districts, schools, teachers, and students.

Participating districts

Three criteria were used to determine district eligibility. First, only medium to large urban and suburban districts (10,000 or more students) in the five states of the REL Southwest Region were considered. The purpose of this criterion was to include only districts in which school size would be likely to accommodate random assignment of multiple classrooms at each site. Second, districts with large percentages of FC–ELL students (25 percent or greater) were targeted. The purpose of this criterion was to sample linguistically diverse schools, to address the confirmatory research question, and to obtain a sufficiently large sample of FC–ELL students, to address the exploratory analyses. Third, the school district had to have at least one school with at least two grade 5 social studies teachers.

A list of districts in the REL Southwest Region that met these criteria was developed using the 2007 Market Data Retrieval system (2010), which uses datasets to compile a description of the characteristics of schools and districts. This system identified 70 potential districts in Arkansas, New Mexico, Oklahoma, and Texas. (No school districts in Louisiana met the size criterion.) These 70 districts were subsequently invited to participate in the study. Forty of the 70 invited districts expressed an interest in learning more about the study; 30 did not respond. No districts in Arkansas or New Mexico responded. Follow-up discussions, by telephone and face-to-face meetings, were held to provide additional study information to interested district and school personnel and to obtain more information about the districts' schools. After these discussions, some districts either declined to participate or were eliminated from consideration because they were not interested in further participation.[12] Seven districts were excluded because they did not meet the third eligibility criterion. This process resulted in a final sample of five participating districts in Oklahoma and Texas.[13]

District characteristics

All five participating study districts were classified as located in an urban area close to a city with a population of 250,000 or more (large city or suburb of a large city). Because of the

[12] Districts cited a variety of reasons for not participating, including participation in other initiatives, concerns over the teacher time required, refusal to randomize at the classroom level, and significant administrative changes, such as the resignation of the district superintendent.

[13] Each participating district was asked to sign a study letter of intent, or memorandum of understanding. In addition to describing data collection activities and expectations, this document granted formal permission to conduct the study in the agreed upon schools. The agreement was signed by the superintendent or another key district official authorized to grant permission to proceed. The study team also worked with each district's Institutional Review Board or research policy group to complete all relevant district-level paperwork and obtain district-level approvals.

eligibility criteria used to identify and select districts, the participating districts are larger than the average district with similar characteristics.[14] They also serve a higher percentage of FC–ELL students (table 2-2).

Table 2-2. Characteristics of urban districts in United States, Oklahoma and Texas, and study sample

Characteristic	*United States*	*Oklahoma and Texas*	*Study sample*
Median number of students per district	3,693.0	12,160.5	40,778.0[a]
Mean percentage of FC–ELL students served in appropriate programs	4.9 (7.5)	8.9 (8.9)	15.7[b] (8.6)
Mean percentage of special education students[c]	13.8 (3.8)	10.7 (2.5)	11.2 (2.2)
Number of districts	2,180[d]	110[e]	5

Note: Figures in parentheses are standard deviations. Districts included in the United States and Oklahoma and Texas comparison groups are those serving grade 5 students in either a large city or in a suburb of a large city.
a. Difference between national and study sample means were statistically significant at the 5% level.
b. Difference between national and study sample medians were statistically significant at the 5% level (Kolmogorov-Smirnow two-sample median test).
c. Special education students are those with a written Individualized Education Program (IEP).
d. National Center for Education Statistics database did not include information for calculating values for FC–ELL or students with IEPs in all schools. U.S. mean was based on n = 1,466 for FC–ELL students and n = 1,793 for students with IEPs.
e. For IEP, n = 109.
Source: Authors' compilation based on data from National Center for Education Statistics (n.d.).

Participating schools and teachers

Schools were recruited from within the participating district sample based on separate school-level eligibility criteria. First, any school in which CSR was already being used was ineligible to participate in the study. Second, schools that used departmentalized instruction for grade 5 social studies were ineligible to participate unless there were at least two social studies teachers who taught different social studies classrooms.[15] Third, schools had to agree to comply with the randomization outcomes, implement CSR in CSR classrooms for the entire school year, and participate in the data collection activities.

Across the five districts, 30 schools met the eligibility criteria and agreed to participate in the study. After completing school-level consent agreements, issues arose in four schools that made them either ineligible or unwilling to participate; these schools were therefore dropped from the study. This occurred after classrooms had been randomly assigned but before pretest data had

[14] A national comparison group of districts serving grade 5 students in either a large city or in a suburb of a large city was used.

[15] Departmentalized instruction refers to a practice in which teachers are responsible for delivering instruction for a single subject within a grade, as is typically the case at the high school level. Of the 26 schools in the final sample, 10 used departmentalized instruction but had at least two grade 5 social studies teachers.

been collected. (See appendix C for more information.) The disqualified schools would not have been deemed eligible to participate had the study team been fully informed of their instructional plans, school closures, or the fact that teachers were unwilling to participate. As dropping these schools entailed simultaneous loss of CSR and control classrooms, it did not pose a major threat to internal validity, as the study design assessed treatment impact within each school and then aggregated data across all sites. Despite a corresponding loss in statistical power, the 26-school sample still provided the study with adequate statistical power. No schools withdrew from the study after pretesting and CSR implementation began.

School characteristics

Table 2-3 compares the participating school sample and comparison groups for the United States as a whole and for Oklahoma and Texas.

Table 2-3. Characteristics of an average urban elementary school in the United States, Oklahoma and Texas, and the study sample, 2006/07

Characteristic	*United States (*n = *17,399)*	*Oklahoma and Texas (*n = *1,751)*	*Study sample (*n = *25)*[a]
Mean number of students			
Per school	544.0(227.1)	614.5 (196.7)	659.5[b] (211.9)
Per teacher	16.4 (3.6)	15.6 (2.0)	16.3 (2.1)
Race/ethnicity (percentage of students)[c]			
American Indian	0.7 (2.0)	1.4 (3.5)	1.2 (2.4)
Asian	6.6 (11.1)	3.9 (5.9)	7.2 (15.4)
Black	21.5 (28.8)	17.2 (20.9)	18.1 (23.8)
Hispanic	28.6 (31.2)	50.5 (32.8)	63.8[b,d] (32.2)
White	41.2 (33.9)	27.0 (27.2)	9.6[b,d] (10.8)
Percentage of students eligible for free or reduced-price lunch	50.0 (32.4)	54.7 (31.2)	76.1[b,d] (27.3)

Note: Figures in parentheses are standard deviations. The national sample size was 17,399 except for *students per teacher* (*n* = 16,881) and *students eligible for free or reduced-price lunch* (*n* = 17,069) comparisons. The Oklahoma and Texas sample size was 1,751, except for the *students per teacher* (*n* = 1,750) and *students eligible for free or reduced-price lunch* (*n* = 1,742) comparisons. Control schools served at least 12 grade 5 students. Some percentages may not total 100 percent, because some data are unreported.

a. Data were only included for study schools with available historical data.

b. Difference between national and study sample means was statistically significant (differences in means *t*-test, $p < 0.05$; all *t*-tests used pooled variances).

c. Students are classified in only one category. Black includes African American, Hispanic includes Latino, Asian includes Native Hawaiian and Other Pacific Islander, and American Indian includes Alaska Native.

d. Difference between regional and study sample means was statistically significant (differences in means *t*-test, $p < 0.05$).

Source: Authors' compilation based on data from National Center for Education Statistics (n.d.).

Randomization of classrooms

Classrooms within schools were randomly assigned after securing school-level participation but before obtaining teacher- and student-level consent. (Details on the random assignment procedures are provided in appendix C.)

Thirty schools with 86 classrooms were included. After random assignment, but before pretesting, two issues arose:

- Four schools and 12 classrooms became ineligible or decided not to participate.[16] These schools and classrooms were dropped from the study.
- Although all districts and schools initially reported that they did not use departmentalized social studies instruction, it was discovered that in 10 schools teachers taught multiple social studies classrooms. For these teachers, one classroom was randomly selected for inclusion in the study. The 25 extra classrooms taught by these study teachers were not included in the sample.[17]

After exclusion of these schools, the baseline sample included 26 schools with 74 classrooms (37 CSR, 37 control).

Teacher consent

Consent forms were distributed to teachers after randomization was complete. CSR and control teachers received different forms because CSR teachers had to consent to more study activities (for example, professional development and coaching) than control teachers. The consent forms provided documentation that the teachers had been informed of, understood, and agreed to the study responsibilities. Teacher consent forms were collected during an initial CSR study orientation.

The baseline sample included 74 grade 5 teachers. The number of teachers in each school ranged from 2 to 5; 16 of the 26 schools had more than two teachers (table 2-4). Each school had at least one CSR and one control teacher (table 2-5). Ten of the 74 teachers taught multiple social studies classes because of departmentalization; however, as discussed previously, only one classroom per teacher was included in the baseline sample.

[16] Reasons schools or classrooms became ineligible or decided not to participate in study include: administrative changes (departmentalizing instruction so that there was only one grade 5 social studies teacher; reorganizing the school's two grade 5 social studies classes based on primary language, so that one classroom conducted instruction in Spanish; or change in classroom instruction strategies), teachers declining to participate and logistical difficulties resulting from issues beyond the school's control.

[17] Teachers with multiple classrooms provided the same instruction to all classrooms (either CSR or control) and did not know which classrooms were randomly selected for inclusion in the sample.

Table 2-4. Distribution of participating grade 5 teachers across schools

Number of grade 5 teachers per school	*Number of schools*	*Total number of teachers*
2	10	20
3	11	33
4–5	5	21
Total	26	74

Source: Authors' analysis of study records.

Table 2-5. Number of participating schools and teachers, by district

		Number of teachers		
District	*Number of schools*	*Total*	*CSR*	*Control*
1	6	19	10	9
2	5	13	7	6
3	4	12	6	6
4	6	16	8	8
5	5	14	6	8
Total	26	74	37	37

Source: Authors' analysis of study records.

Teacher characteristics and baseline equivalence

Descriptive information for participating teachers—such as years of teaching, Spanish fluency, and education—was collected through a teacher survey, and baseline equivalence was tested for these variables (table 2-6). The results indicate equivalence for the CSR and control teachers on all but one measured characteristic: a statistically significantly higher proportion of teachers who reported Spanish as a second language participated in the CSR condition than in the control condition.[18] This difference was statistically controlled for in the analyses by including Spanish fluency as a covariate in the impact models.

Table 2-6. Teacher characteristics and baseline equivalence results

(mean values)

Teacher characteristic	*CSR (n = 37)*	*Control (n = 37)*	p-*value*
Experience, preparation, and professional development			
Years in current school	7.2 (7.4)	4.9 (6.0)	0.13

[18] Because of the small number of teachers who spoke Spanish as a second language in the control group, this category is not reported separately in table 2-6.

Teacher characteristic	*CSR (n = 37)*	*Control (n = 37)*	p-*value*
Years of teaching experience	11.8 (8.0)	11.4 (11.1)	0.80
Paid preparation hours per week	4.1 (1.4)	4.3 (1.2)	0.14
Professional development in reading (hours during previous school year)	16.5 (29.9)	9.8 (15.7)	0.10
Spanish fluency (percent)[a]			
Spanish as a first or second language[b]	37.9 (49.2)	16.2 (37.4)	0.07
Speaks a little Spanish	21.6 (41.7)	40.5 (49.8)	0.07
Speaks no Spanish	40.5 (49.8)	43.2 (50.2)	0.96
Grade level taught			
Taught grade 5 last year	73.0 (45.0)	59.5 (49.8)	0.19
Ever taught grade 5	89.2 (31.5)	76.6 (43.5)	0.11
Education			
Proportion with master's degree	35.1 (48.4)	32.4 (47.5)	0.66

Note: Figures in parentheses are standard deviations. Baseline equivalence was estimated using a two-level hierarchical linear model (HLM) with teachers at level 1 and schools at level 2. The model at level 1 was estimated by regressing the teacher characteristic on the indicator variable for the CSR condition, the only independent variable in the model. Different models were used based on the type of dependent variable. *Spanish fluency*, *education*, and *grade levels taught* were modeled using logistic HLM. Continuous dependent variables (*years teaching experience, years in current school, paid preparation hours per week,* and *professional development in reading*) were modeled through a linear regression HLM. Means and standard deviations are based on the raw data and are not model based. Sample size for each group is 37.

a. Chi-square test for Spanish fluency shows overall significant difference ($\chi 2 = 5.36$, $p = 0.07$). As this analysis was not adjusted for clustering, the *p*-values for each category are also included.

b. To avoid a potential disclosure risk, data for teachers who reported speaking Spanish as a first or second language have been combined.

Source: Authors' analysis of teacher surveys.

Participating students

The primary criterion for a student to be eligible to participate in this study was parental consent. Three additional criteria were used to exclude students from the study: insufficient English fluency (that is, inability to take the GRADE in English); special education needs that removed a student from the regular classroom during social studies instruction; and the need for testing accommodations that prevented the study team from assessing reading comprehension skills using the GRADE.

Student consent

Consent forms and letters describing the study were sent home with students at the beginning of the school year, before pretest data collection. Consent forms were also distributed before posttesting, for students who were present at baseline but had not previously returned consent. The consent forms requested parents' permission for the collection and use of student pretest, posttest, and district-provided demographic data. Parents provided consent only for data collection and use; they were not asked to consent to students participating in CSR. The use or

nonuse of CSR in the classroom was considered an instructional choice no different from any other routine school or teacher decision. For this reason, the consent forms did not inform parents of whether their children were in CSR or control conditions, although this information was provided to any parent who followed up using the provided contact information.

At baseline there were 74 classrooms across 26 schools, which included 1,652 enrolled students (827 CSR, 825 control).[19] Consent was received for 1,351 eligible students at baseline (702 CSR, 649 control). Baseline consent rates were 84.9 percent for CSR students and 78.7 percent for control students. A second round of consent forms was distributed before posttest. Of the 301 students who were present at baseline but did not provide consent before pretest, 56 (14 CSR, 42 control) provided consent during this round. The second round consent rate was 18.6 percent.

At the time parental consent forms were distributed, classrooms had already been randomly assigned to CSR or control groups. Teachers therefore knew their study condition. Because teachers helped collect consent forms, it is possible that CSR teachers were more motivated than control teachers to obtain parental consent, which may result in differential return rates. However, the difference in baseline consent rates was not statistically significant (estimate = 0.30, standard error = 0.25; $p = 0.23$; see appendix D).

Student characteristics and baseline equivalence

Districts provided demographic data on participating students, such as race/ethnicity, and free or reduced-price lunch program status. Districts also provided language status by indicating if a student was classified as FC–ELL or non-ELL; related language status data such as level of English proficiency, first language, or any specific ELL programs in which students participated were not provided. Baseline equivalence was tested for the district-provided variables, as well as for pretest GRADE scores. No statistically significant differences were identified (table 2-7).

Table 2-7. Characteristics of mean participating student and baseline equivalence results

Student characteristic	*CSR*	*Control*	*Total sample*	p-*value*
Gender				
Male	48.3 (50.0)	47.5 (50.0)	47.9 (50.0)	0.92
Female	51.7 (50.0)	52.5 (50.0)	52.1 (50.0)	
Race/ethnicity				
American Indian	1.9 (13.7)	0.8 (8.9)	1.4 (11.6)	0.13
Asian	6.0 (23.8)	5.2 (22.2)	5.6 (23.1)	>0.99
Black	18.7 (39.0)	23.8 (42.6)	21.2 (40.9)	0.52
Hispanic	63.6 (48.1)	61.5 (48.7)	62.6 (48.4)	0.73
White	9.7 (29.7)	8.7 (28.2)	9.2 (28.9)	0.84
Free or reduced-price lunch				
Yes	73.6 (44.1)	72.6 (44.7)	73.07 (44.4)	0.86
No	26.4 (44.1)	27.4 (44.7)	26.93 (44.4)	

[19] There were 507 students in the 25 extra classrooms not included in the baseline sample.

Student characteristic	*CSR*	*Control*	*Total sample*	p-*value*
Special education				
Yes	11.1 (31.4)	11.0 (31.4)	11.05 (31.4)	0.84
No	88.9 (31.4)	89.0 (31.4)	88.95 (31.4)	
Language status				
FC–ELL	48.5 (50.0)	40.1 (49.0)	44.4 (49.7)	0.16
Non–ELL	51.5 (50.0)	59.9 (49.0)	55.6 (49.7)	
Baseline reading proficiency				
GRADE total score	93.9 (12.3)	94.8 (11.8)	94.3 (12.1)	0.40

Note: Figures in parentheses are standard deviations. The number of students in classrooms in the baseline sample for whom there were demographic data ranged from 1,311 to 1,313, depending on the variable. Pretest data was collected for 1,280 students (659 CSR, 621 control). The baseline equivalences were estimated using a three-level hierarchical linear model (HLM) including only an indicator variable for the CSR condition. Different HLMs were used based on the type of dependent variable. *Race/ethnicity* (categories dummy coded), *free or reduced-price lunch, special education,* and *Language status* were modeled using logistic HLM. *Baseline reading proficiency* was modeled using linear regression HLM. Means and standard deviations were calculated from the available data across the schools. A single *p*-value is provided for dichotomous variables (for example, male versus female).
Source: Authors' analysis of student GRADE scores and district demographic data.

Crossovers

Fourteen students in the study changed classrooms, crossing from one study condition to another (eight from the CSR to the control condition and six from the control to the CSR condition). Crossovers occurred in all five participating districts, with no more than two crossovers taking place within a school. These crossovers were analyzed according to an intent-to-treat analysis in which study participants were analyzed by keeping them in the condition to which they were assigned at the start of the study, regardless of the condition they were in at the end. As is common practice in RCTs, a sensitivity analysis was performed to determine whether ignoring crossovers altered the findings (see table 4-2 and appendix E).

Contamination

Contamination is a concern in studies in which treatment and control classrooms are assigned within the same school. It is possible that CSR was used in control classrooms, as control teachers may have been exposed to CSR and incorporated it into their instruction. It is also possible that the interaction of CSR and control students in settings such as recess, other classes, and outside of school could have introduced contamination.

In principle, contamination could lead to biased impact estimates in which observed CSR versus control differences are smaller than they would have been in the absence of contamination. Because CSR theory claims that impacts are realized only if CSR teachers properly apply explicit CSR teaching strategies and cooperative learning techniques, this potential threat to the study's internal validity is low. A fundamental assumption underlying CSR is that learning to use the techniques effectively requires formal professional development, opportunities to apply CSR in classrooms, and coaching. Given these requirements, it is unlikely that teachers in the control classrooms could have implemented the intervention based only on casual conversations with CSR teachers.

Although the risk of contamination was deemed low, both CSR and control teachers were informed about contamination and how it could undermine the study (see appendix F). Control group instruction was observed by study staff (see chapter 3 for the results of these observations), and schools were provided with a reporting mechanism for informing coaches of possible contamination. Neither participating schools nor study staff reported concerns with contamination.

In some cases, midyear administrative changes—in particular when departmentalized instruction occurred in classrooms after implementation of the study had begun—increased the risk of contamination. After consultation with school principals and the CSR teachers where this occurred, the schools and teachers agreed not to use CSR in control classrooms or in other subjects. A sensitivity analysis was conducted to determine whether dropping these classrooms would alter study findings (see chapter 4).

Attrition

Random assignment ensures that any baseline differences between the CSR and control groups occur by chance. Significant baseline differences were accounted for by including appropriate covariates in the model.

Attrition can potentially bias the impact estimate. This section describes attrition at the school, classroom, teacher, and student levels.

School-level attrition

As discussed, four schools dropped out of the study after random assignment but before implementation, and no schools left after implementation began. School-level attrition was therefore 4 of 30 schools (13 percent). Because random assignment occurred within schools, and the impact estimate is estimated within schools and then aggregated, the loss of these schools should not bias the impact estimate.

Classroom-level attrition

The sample recruited for this study included 86 classrooms in 30 schools. Four schools, with 12 classrooms, dropped out of the study after random assignment, leaving a sample of 26 schools, with 74 classrooms. No classrooms left the study after implementation began (although the composition of teachers and students within those classrooms sometimes changed, as detailed below). Classroom-level attrition was therefore 14 percent (12 of 86 classrooms).

Teacher-level attrition

Teacher-level attrition could pose a problem for this study in two ways. First, if a teacher left and the teacher's classroom was subsequently dropped from the study, the impact estimates could be biased. Second, if a CSR teacher who left was replaced by another teacher and the classroom remained in the study, the fidelity of implementation in that classroom could be compromised.

The following steps were taken to prevent teacher attrition:

- A clear explanation of study requirements was provided in the teacher consent form and during the initial teacher training to ensure that both principals and teachers fully understood the burden imposed by study participation.

- Attempts were made to develop teacher surveys that did not excessively burden teachers. Teachers were compensated for the time needed to complete the surveys.
- The importance of participating in a scientific study was emphasized. Participants were informed that the results could be relevant not only for participating teachers but potentially for other teachers in classrooms with high percentages of FC–ELL students.

In contrast to classrooms, teachers became part of the sample only after consenting to participate. The attrition of 4 schools and 12 classrooms occurred before teacher consent. Therefore, there were 74 classrooms in which there were eligible teachers who consented to participate. Overall teacher-level attrition was 1 percent; however, there was no attrition among teachers assigned to the CSR condition.

Student-level attrition

The baseline student sample consisted of the 1,337 (692 CSR, 645 control) eligible, testable students who—at the time of pretesting—were in the final baseline school sample of 26 schools.[20] Of those students, 126 (67 CSR, 59 control) were not present at posttest. Student-level attrition was thus 9 percent (126 of 1,337 students; table 2-8).[21]

Table 2-8. Student-level attrition

	CSR		*Control*		*Total*	
Status	*Number*	*Percent*	*Number*	*Percent*	*Number*	*Percent*
Total enrollment at baseline	827	100.0	825	100.0	1,652	100.0
Eligible at pretest						
Consented	702	84.9	649	78.7	1,351	81.8
Testable	692	83.7	645	78.2	1,337	80.9
Testable at pretest, present at posttest	625	90.3	586	90.9	1,211	90.6
Testable at pretest, not present at posttest	67	9.7	59	8.1	126	9.4

Source: Authors' analysis of study records.

A series of analyses found no statistically significant relationship between student-level attrition and study condition (CSR versus control), and some statistically significant relationships between student-level attrition and demographic variables (see appendix G). These differences are addressed in the multiple-imputation procedures described in the data analysis section of this chapter.

Study sample at each study phase

The participant flow chart (figure 2-1) describes the sample of students at each phase of the study for whom data were collected. At pretest 1,337 students were testable with parent consent (692

[20] Eligible, testable students are those with parental consent who were not excluded due to insufficient English fluency, special education status, or the need for testing accommodations that prevented administration of the GRADE.

[21] The power analysis was based on an assumption of 20 percent student attrition.

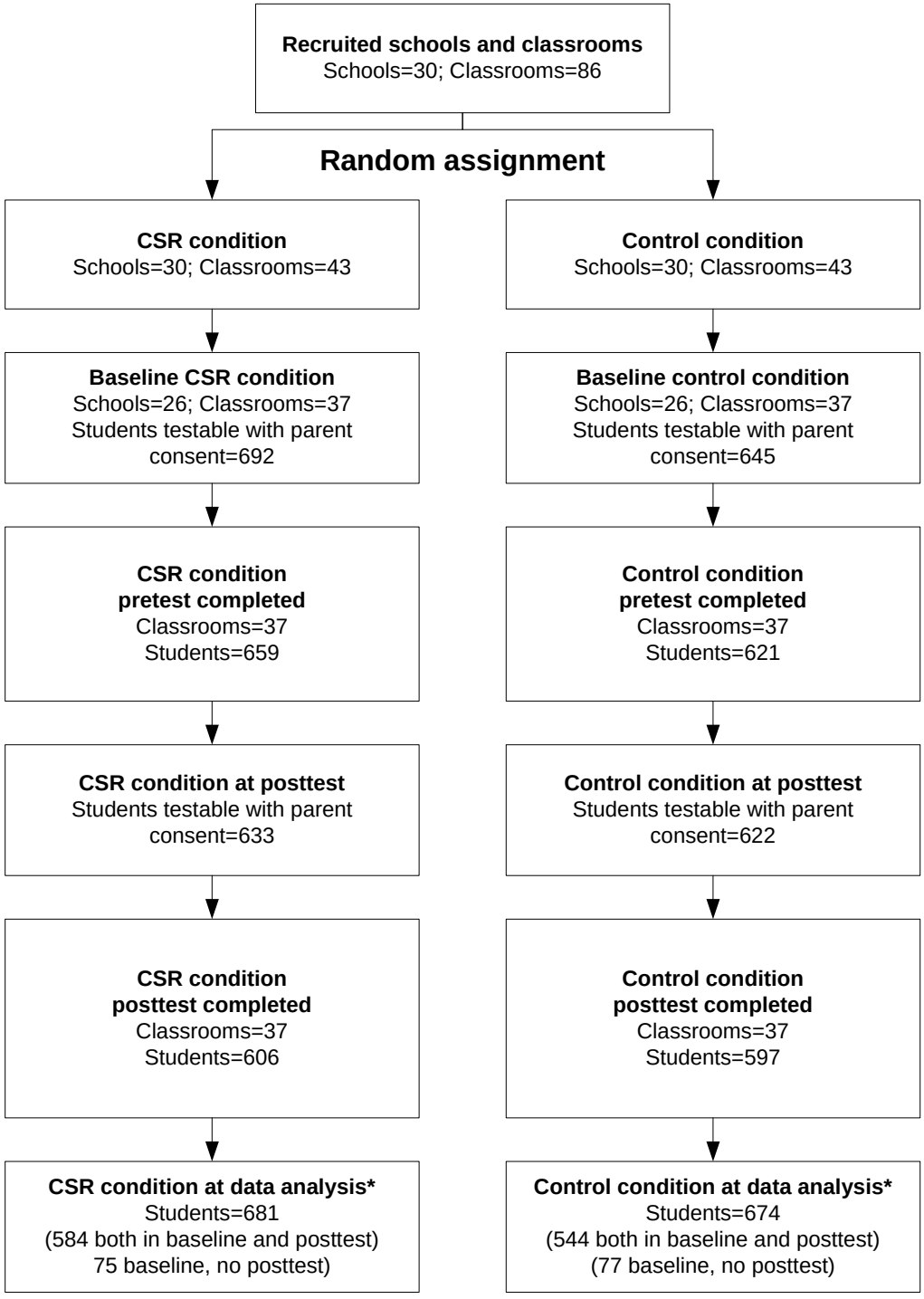
CSR and 645 control). 1,280 of these students were tested (659 CSR and 621 control); this corresponds to a response rate of 95.2 percent CSR students, and 96.3 percent of control students. At posttest 1,255 were testable with parent consent (633 CSR and 622 control). 1,203 of these students were tested (606 CSR and 597 control); this corresponds to a response rate of 95.7 percent CSR students, and 96.0 percent of control students. (Additional response rate data are presented in appendix H.) The final analytic sample includes students for whom multiple imputation results were used to adjust for data missing at either pretest or posttest. The total number of students in the final analytic sample was 1,355 students (681 CSR and 674 control).

Tables 2-9, 2-10, and 2-11 provide additional detail on how the sample changed between baseline and posttest. At pretest the study sample consisted of all students who were enrolled in the classrooms at the beginning of the year, obtained parental consent, and met testing eligibility requirements. By posttest the sample had changed in two ways. First, some students who were part of the sample at pretest had left the classroom or were no longer testable. These students were dropped from the sample. Second, some students who had been enrolled in the classrooms at baseline did not return parental consent forms until after pretesting had occurred or became eligible for testing during the course of the school year. These students were added to the sample at posttest.

Data collection instruments

Several data collection instruments were used (table 2-12). Student reading comprehension at pretest and at posttest was measured using the GRADE. Information about teacher characteristics was collected through a survey administered during the fall. Fidelity of professional development was measured using training observation sheets and coaching logs. Fidelity of implementation was measured through teacher surveys administered in the spring and an observational checklist. Control classroom instruction was examined with an observational instrument.

Figure 2-1. Participant flow chart



Note: After random assignment but before pretesting, the study team discovered that some teachers taught multiple social studies classrooms. In consultation with the study's technical working group, the team decided to randomly select one classroom for inclusion in the sample for each teacher who taught multiple classes. As a result, 25 "extra" classrooms were not included in the sample.
Source: Authors' analysis of study records.

Table 2-9. Number of baseline students enrolled in the classroom

Stage	*CSR*	*Control*	*Total*
Baseline	827	825	1,652
Less students who left classroom before posttest	84	73	157
Posttest	743	752	1,495

Source: Authors' analysis of study records.

Table 2-10. Number of baseline students in the classroom with consent

Stage	*CSR*	*Control*	*Total*
Baseline	702	649	1,351
Less students who left classroom before posttest	71	60	131
Plus students who obtained consent between pretest and posttest	14	42	56
Posttest	645	631	1,276

Source: Authors' analysis of study records.

Table 2-11. Number of baseline students in the classroom eligible for testing

Status	*CSR*	*Control*	*Total*
Baseline	692	645	1,337
Less students who: left classroom before posttest *or* became ineligible for testing at posttest	72	66	138
Plus students who: became eligible for testing at posttest *or* obtained consent between pretest and posttest	13	43	56
Posttest	633	622	1,255

Source: Authors' analysis of study records.

Table 2-12. Data collection instruments used

Instrument	*CSR*	*Control*
Student reading comprehension measure		
GRADE form A (pretest)	✓	✓
GRADE form B (posttest)	✓	✓
Teacher characteristics measure		
Fall survey of teachers	✓	✓
Fidelity of professional development measures		
Training observation sheets	✓	na
Coaching logs (including coaching observation form in the fall and Collaborative Strategic Reading Intervention Validity Checklist [CSRIVC] in the spring)	✓	na

Instrument	*CSR*	*Control*
Fidelity of implementation measures		
Spring survey of CSR teachers	✓	na
Spring survey of control teachers	na	✓
Intervention fidelity checklist (CSRIVC)	✓	na
Classroom instruction measure		
Classroom observation instrument (Expository Reading Comprehension [ERC] subscale)	✓[a]	✓

na is not applicable.
a. ERC data were collected to establish the degree to which teaching behaviors similar to CSR were being used in control classrooms during social studies instruction. These data were collected for both control and CSR classrooms; only the data collected for control classrooms are discussed in the report. The more relevant CSRIVC data were reported for CSR classrooms.
Source: Authors' analysis of study records.

Student reading comprehension measure

The GRADE is a standardized, nationally normed, group-administered reading test developed by American Guidance Service, Inc. (2001). It was chosen as the student reading comprehension measure for this study because it focuses on the same constructs as CSR—namely, the vocabulary, reading comprehension, and metacognitive skills grade 5 students need to understand informational text, such as questioning, predicting, clarifying, and summarizing.

The GRADE yields a total scaled score that includes vocabulary, sentence comprehension, and passage comprehension subscales.[22] It has a mean of 100 and a standard deviation of 15. Item-level missing data are not a concern, because the standardized scoring rules treat any blank item as an incorrect answer. For grade 5 students, estimates of the total scaled score's test-retest and internal consistency reliability range are 0.93–0.97 across the two forms of the GRADE.[23]

The GRADE offers equivalent forms for pretesting (form A) and posttesting (form B). The English-language version of the GRADE was used, and the test was group administered. A make-up day was scheduled for students who were absent or who could not finish the test in the allotted time. To minimize missing data, schools with three or more missing students were given priority in scheduling a make-up day.

The test is untimed; the estimated average administration time for grade 5 students is 60 minutes. The test developers recommend two administration sessions for struggling readers; this recommendation was followed for this study when needed. For example, students who could not complete the test within 90 minutes were given an extra 30 minutes to finish. If this was insufficient, a second testing session was scheduled. The decision to schedule an additional testing session was made by the study team's lead data collector in consultation with classroom teachers. Staff members responsible for GRADE data collection received eight hours of training on administering the measure. Data collection managers were in charge of scheduling and overseeing data collection as well as providing quality control.

[22] Another subscale, listening comprehension, was not pertinent to the study, because CSR was not expected to affect it.
[23] See the GRADE technical manual (American Guidance Service, Inc. 2001) for additional information.

Teacher characteristics measure

A survey was conducted in the fall to collect information about teacher characteristics (appendix I). The paper-and-pencil survey, administered to both CSR and control teachers, took about 20–30 minutes to complete. Items from an earlier study on reading instruction in elementary grades (Garet et al. 2008) were used. The resulting data were used to establish teacher baseline equivalence. The surveys collected information about the following:

- Respondent characteristics (such as years of teaching experience, education, gender, race/ethnicity).
- Approaches to instruction (compartmentalization, team teaching).
- Classroom resources available for instruction.
- Reading and social studies curricula used in the class.
- Professional development teachers received for their business-as-usual reading curricula.
- Time spent on student projects.

Fidelity of professional development measures

Two instruments were used to measure fidelity of professional development.

Training observation sheets

A member of the study team observed the entire initial two-day training sessions conducted in each district and recorded information on time spent on agenda topics. (Chapter 3 presents additional information about the teacher training sessions.) Attendance was also recorded as part of the training observation.

Coaching logs

Coaches assigned to CSR teachers were required to maintain logs of coaching sessions. For the fall coaching sessions, the fall coaching observation form was used (see appendix J). For the spring coaching sessions, the CSRIVC was used. The frequency of coaching sessions and notes from each session were recorded.

Fidelity of implementation measures

Three instruments were used to measure fidelity of implementation. All teachers completed surveys in the spring. The CSRIVC was administered to CSR teachers only. A subscale of the ERC was used in control classrooms.

Teacher surveys administered in the spring

Separate versions of the teacher survey were administered in the spring—one for CSR teachers and one for control teachers. To maximize response rates, teachers were asked to fill out the paper-and-pencil surveys while students were taking the GRADE test. If teachers were not present when the surveys were distributed, they were sent a copy with instructions for completion and a postage-paid return envelope. The surveys took about 20–30 minutes to complete. As in the teacher survey administered in the fall, items from an earlier study on reading instruction in elementary grades (Garet et al. 2008) were used.

The same questions on class structure, class behavior, instructional resources, reading program, and teacher collaboration were included in both versions of the spring survey. The CSR teacher survey also included questions about use of CSR. These items focused on how many times a week teachers used CSR in their classrooms, whether they had met with a CSR coach during the fall semester, and whether they thought CSR had had a positive impact on their students' reading comprehension skills.

The Collaborative Strategic Reading Intervention Validity Checklist (CSRIVC)

The CSRIVC is an observational checklist created by the intervention developers to measure fidelity of CSR implementation (Vaughn et al. 1998). This checklist was used only in CSR classrooms.

The CSRIVC includes 31 items, focused on student behaviors, teacher behaviors, and the classroom setting. While the teacher is using CSR, a classroom observer completes the CSRIVC by noting whether CSR instructional techniques and related student behaviors are evident. The CSRIVC captures the presence or absence of behaviors but not how often or with what quality different aspects of the intervention are implemented. Observers are also asked to evaluate the following:

- Overall functioning of student groups.
- Differences in functioning across groups.
- Adaptations made to the program.
- Portion of the CSR intervention conducted in a whole class rather than small group format, if any.

Previous studies do not report psychometric properties of the CSRIVC or identify explicit criteria for judging high versus low implementation fidelity. However, the CSRIVC focuses on the aspects of the intervention deemed important by the developers. In this study, the CSRIVC was used both by CSR coaches to help them identify CSR elements teachers needed to improve and by a separate group of trained observers to document fidelity of implementation.[24] Fidelity observations using the CSRIVC were conducted a single time by a single observer; therefore, interrater reliability data are not available for these observations.

Classroom instruction measure

One instrument was used to categorize and code teachers' comprehension and vocabulary instruction.

The Expository Reading Comprehension (ERC) observation instrument

The ERC is a time-based observation instrument that tallies the frequency of use of a set of instructional behaviors that reading experts believe are important for reading comprehension. It was developed to systematically categorize and code the content and quality of a teacher's comprehension and vocabulary instruction.[25] In this study, a subscale of the ERC was used to

[24] CSR coaches served as observers for CSRIVC fidelity observations only for teachers they did not coach.

[25] The ERC protocol was developed for a recent reading comprehension intervention study funded through the Institute of Education Sciences (James-Burdumy et al. 2009).

document group instruction and explicit reading instruction strategies used in control classrooms. ERC subscale interrater reliability data was assessed by including a second observer in 20 percent of the observations; average interrater reliability (as measured by the percentage agreement between observers for the ERC protocol) was 97 percent (standard deviation = 4.12), ranging from 88 to 100 percent (see appendix N for more information).

Data analysis

This section provides an overview of the confirmatory impact model, the sensitivity analyses, and the exploratory impact model. It also describes the approach to missing data. (See appendix E for detailed descriptions of the confirmatory, exploratory, and sensitivity models.) [26]

Confirmatory impact model

One confirmatory research question is examined in this study:

- In linguistically diverse schools, do grade 5 students in CSR classrooms have higher average reading comprehension posttest scores on the GRADE than students in control classrooms?

Given the nested data structure in this study (students nested within classrooms nested within schools), the primary analytic approach was the HLM method (Raudenbush and Bryk 2002). This method was chosen over traditional regression because the results, particularly the standard errors, are more accurate for nested data. Covariates can also be included to address a lack of baseline equivalence, because they are theoretically relevant or because they increase the statistical precision of the parameter estimates. The data analyses included baseline covariates at the student or teacher level, which was necessary because the Spanish fluency of CSR and control teachers differed at baseline (see table 2-7). The confirmatory impact model included two covariates for students: the pretest GRADE score measured at baseline and students' language status (FC–ELL or non–ELL). The pretest GRADE score was strongly correlated to the outcome measure (the posttest GRADE score). It was included to improve the precision of the analysis. The language status variable was included because the study was designed to examine the effect of CSR in linguistically diverse classrooms. In addition, these subgroups had significantly different pretest GRADE scores and were examined separately in the exploratory analyses.[27]

The effect of CSR on student performance on the reading outcome measure was estimated by comparing students in CSR classrooms with their counterparts in control classrooms within schools, while controlling for student and teacher baseline characteristics. A treatment effect was estimated for each school; the overall effect of CSR was estimated as a weighted average of the school effects. The number of students in the study sample in a given school relative to the total number of students in the study yielded a proportion that was used to weight the school-specific estimates to obtain the combined treatment effect.[28]

---

[26] Before the confirmatory hypothesis was tested, preliminary data analyses were conducted, including analyses of outliers and missing data. All data were independently entered twice and compared, in order to ensure the accuracy of data entry. Any discrepancies were verified by examining the original records.

[27] The multiply imputed estimate of the mean baseline GRADE scores is 91.7 for the FC-ELL subgroup and 96.4 for the non-ELL subgroup. This difference is statistically significant (estimate = 12.47, SE = 0.70, p-value <0.01).

[28] The number of students ranged from 21 to 93 per school in the full sample, from 2 to 65 in the FC–ELL student sample, and from 10 to 73 in the non–ELL student sample.

Schools in this sample were not randomly sampled from a larger population of schools. Therefore, the confirmatory impact model used a *schools as fixed effects* approach (students nested within classrooms nested within schools), meaning that the results of this analysis can be applied only to the schools in the sample; they are not generalizable to a broader population.

The impact of CSR was estimated using an intent-to-treat approach, meaning that students were analyzed according to the classroom to which they were initially assigned (even if the student crossed over from one condition to another). The estimation of the effect of CSR is based on 10 multiply imputed datasets (see appendix K). The magnitude of the effect is gauged by the size of the Hedges' *g* effect (the mean difference between groups divided by the pooled standard deviation of posttest scores).

Sensitivity analyses

The primary confirmatory analysis was supplemented by sensitivity analyses to explore the robustness of findings to various analytic assumptions and choices made (appendix E). The sensitivity analyses conducted were as follows:

- *Equal weighting:* The confirmatory impact analysis estimated the intervention effect by aggregating the school-specific estimates using weighting relative to student sample size in schools to account for differences in sample sizes across schools. To evaluate the robustness of the findings to this choice, a sensitivity analysis was conducted in which the effect of the CSR intervention was estimated using a model that gave equal weights to school-specific estimates.
- *Random effects:* The confirmatory impact analysis used a fixed effects model in which school-specific treatment effects were estimated and then aggregated to obtain the overall impact estimate. To evaluate the robustness of the findings to this choice, a sensitivity analysis was conducted in which the confirmatory impact analysis was estimated using a model that included schools as random, rather than fixed, effects.
- *List-wise deletion:* The confirmatory impact analysis used multiple imputation as the missing data approach (as described in the missing data approach section of this chapter). To evaluate the robustness of the findings to this choice, a sensitivity analysis was conducted in which the confirmatory impact analysis was estimated using list-wise deletion.
- *Exclusion of crossover students:* In order to obtain appropriate intent-to-treat impact estimates, students who changed condition during the school year were analyzed in the confirmatory impact analysis based on their initial classroom assignment. To evaluate the robustness of the findings to this choice, a sensitivity analysis was conducted in which crossover students were excluded from the student sample.
- *Exclusion of schools with increased contamination risk:* The school sample used for the confirmatory impact analysis included study schools with an increased risk of contamination as a result of the midyear departmentalization of instruction. To evaluate the robustness of the findings to this choice, a sensitivity analysis was conducted in which these schools were excluded from the school sample.
- *Unconditional effect of treatment:* Baseline covariates were included to increase power; language covariates were included to account for baseline differences. To evaluate the

robustness of the findings to this choice, a sensitivity analysis was conducted in which the impact model was run without including any covariates.

- *Heterogeneity of slopes (using 26 schools):* The confirmatory impact analysis assumed that the effect of the CSR did not vary depending on students' pretest scores. To test this assumption, a sensitivity analysis was conducted using a model that included a pretest*CSR interaction term to test this term's significance.
- *Heterogeneity of slopes (using 23 schools):* The same sensitivity analysis was conducted using the 23 schools that represent the subset of the school sample in which FC–ELL and non–ELL students were located in both CSR and control classrooms.
- *Cohort effect:* Data used for the confirmatory impact analysis were collected over a two-year period using two cohorts of schools. The first cohort included 6 schools; the second included 20 schools. To evaluate whether the treatment effect differs between cohorts, a sensitivity analysis was conducted using a model that estimates a differential effect by study cohort.

Exploratory analyses

Three exploratory questions were examined in this study:

- Do grade 5 FC–ELL students in CSR classrooms have higher average reading comprehension posttest scores on the GRADE than FC–ELL students in control classrooms?
- Do grade 5 non–ELL students in CSR classrooms have higher average reading comprehension posttest scores on the GRADE than non–ELL students in control classrooms?
- Does CSR have a differential impact on GRADE reading comprehension posttest scores for grade 5 FC–ELL and non–ELL students?

The first two exploratory questions were examined by estimating the effect of CSR on FC–ELL and non–ELL students. CRS and control groups were examined for both FC–ELL students and non–ELL students while controlling for student and teacher baseline characteristics. School-specific treatment effects were estimated for each school. The overall effect of CSR was estimated giving equal weights to the school effects. Equal weighting was used instead of weighting by the proportion of study sample students to make results consistent across subgroups and yield an estimate of the treatment effect for each subgroup in an average school.

The third exploratory research question used the same HLM, but instead of estimating the average effect of CSR for FC–ELL and non–ELL students, the difference in CSR's impact for these two groups was estimated. (See appendix E for detailed descriptions of the exploratory models.)

Approach to missing data

A wide variety of approaches can be used to deal with missing data. Two common approaches are list-wise deletion and multiple imputation. List-wise deletion assumes that the data are missing completely at random (MCAR), which means that the probability of being missing is unrelated to either observed or unobserved variables, including the treatment indicator variable.

However, the probability of being missing was significantly related to some student characteristics, such as reading comprehension (pretest GRADE scores), special education status, and language status. Moreover, there were statistically significant differences in the level of

missing pretest GRADE scores for the CSR and control groups (see appendixes G and K). Because the assumption of MCAR is not supported, list-wise deletion is not appropriate.

A weaker assumption about missing data is that the data are missing at random (MAR), meaning that the probability of being missing does not depend on unobserved variables after accounting for observed variables. When this assumption is reasonable, an appropriate method for dealing with missing data is multiple imputation. Although the MAR assumption is not directly testable, a variety of observed variables could be included in the imputation model. Therefore, multiple imputation was selected as the missing data approach for this study. (See chapter 4 for the findings from a sensitivity analysis conducted using list-wise deletion.)

The procedure was conducted by using imputation and variance estimation software (IVEware), which applies multivariate stochastic sequential regression-based multiple imputation (Raghunathan, Solenberger, and van Hoewyk 2002). Multiple imputation was conducted separately for the CSR and control groups; 10 multiply imputed datasets were created for the CSR and control conditions and then combined to create a final analytic dataset. The multiple imputation models included student-level variables for race/ethnicity, language status, gender, free or reduced-price lunch status, and pretest and posttest scores. The teacher-level variables were Spanish fluency, teaching experience (total number of years), and class size. Dummy indicator variables were included to account for the nested structure of the data (students nested within teachers, teachers nested within schools). (See appendix K for a detailed description of the multiple imputation procedures.)

The multiply imputed datasets used to analyze confirmatory impact included all students who were enrolled in study classrooms at baseline, had parental consent, and were testable (that is, not excluded because of insufficient English fluency, special education status, or the need for testing accommodations that prevented administration of the GRADE); present at baseline but provided consent for posttesting only; and had at least one valid GRADE test score (pretest or posttest). Multiple imputation was used to estimate posttest GRADE scores for 152 students (75 CSR, 77 control) and pretest GRADE scores for 75 students (22 CSR, 53 control). *ELL* and *free or reduced-price lunch* status were missing for 31 students (18 CSR, 13 control); *gender*, *race/ethnicity*, and *special education* status were missing for 30 students (17 CSR, 13 control).

Chapter 3. Implementation of the Collaborative Strategic Reading intervention

This study is an effectiveness trial designed to evaluate CSR as implemented by classroom teachers. In the previous efficacy trial (Klingner, Vaughn, and Schumm 1998), the developers of CSR delivered the intervention directly to students in the classrooms. This method of delivery would not have been feasible for the large number of classrooms required in a district-wide study. Therefore, in this study, teachers received an initial two-day training session provided by a codeveloper of CSR, Dr. Joseph Dimino, and then delivered CSR directly to students in the classroom.

This chapter describes the teacher professional development provided and discusses the cost of implementation. It also examines the fidelity of implementation, including fidelity of professional development, and provides some evidence as to whether CSR strategies were present in control classrooms.

Professional development for participating teachers

CSR teachers participated in an initial two-day training session at the beginning of the school year. They received four follow-up coaching sessions during the school year.

Initial training

Initial two-day training sessions were held for each district at the beginning of the school year, before pretesting. The training was identical in all five districts, with the exception of minor adaptations made to address teacher questions (see table 3-1).

The training agenda included the introductory overview session provided for both CSR and control teachers and four CSR teacher training segments: the theoretical foundations of CSR, CSR comprehension strategies, implementation practices and logistics, and coaching.[29] Dr. Dimino, a coprincipal investigator and codeveloper of CSR, provided this training.[30] Coaches attended the training session to meet the CSR teachers they would be coaching. Throughout the training, Dr. Dimino emphasized that CSR strategy instruction could be integrated into social studies lessons without significantly decreasing the amount of time allocated to instruction.

Training segment 1: Theoretical foundations of CSR

Training segment 1 focuses on metacognition and its components: *metacognitive knowledge* (readers' ability to understand and use the strategies they need to successfully comprehend text) and *self-regulatory knowledge* (the degree to which readers monitor, evaluate, and self-correct) (Billingsley and Wildman 1990; Duffy 2002). The training also covers three scaffolding steps used in CSR for instructional delivery: teacher modeling, teacher assistance, and independent practice.[31] It discusses the point at which a student can apply a skill with teacher support, known

[29] After attending the brief introduction that explained the study and emphasized the need to avoid contamination, control teachers were dismissed; only the CSR teachers attended the actual CSR training.

[30] See the disclosure section of the report for additional information about Dr. Dimino's involvement with CSR.

[31] Scaffolded instruction is a system for providing support and reinforcement for emerging skills.

as the zone of proximal development (Vygotsky 1978). In practice, students use metacognitive abilities to detect when comprehension has broken down and to select and implement the appropriate CSR comprehension strategy.

Training segment 2: CSR comprehension strategies

Training segment 2 introduces the four CSR reading comprehension strategies: *Preview, Click and Clunk, Get the Gist,* and *Wrap up.* The first strategy, *Preview,* includes several steps, which require students to:

- Scan text features, such as the title, subheadings, and illustrations to identify key words and terms.
- Brainstorm prior knowledge about the topic, such as information acquired from previous lessons or from watching a movie or television program.
- Record brainstorming results in learning logs (journals students keep to record what they learn during the CSR process).
- Predict what they will learn from reading the selection by generating and recording specific statements in their learning logs.

The second CSR comprehension strategy, *Click and Clunk,* is designed to help students monitor their comprehension by identifying difficult vocabulary as they read and to provide students with a system for determining the meaning of words they do not understand. Students are taught that when they understand what they are reading, everything is *clicking*; *clunking* occurs when students encounter a challenging or unknown word or concept that interrupts their comprehension (that is, a *clunk*). After reading each paragraph or text passage, students pause to identify clunks and determine their meaning. When a clunk is identified, students deploy four *fix-up strategies* in a sequence until they have "declunked" the word and determined its meaning:

- Reread the sentence containing the clunk, looking for clues to the word's meaning.
- Reread the sentence before and after the sentence containing the clunk, looking for clues to the word's meaning.
- Look for prefixes or suffixes that might help determine a word's meaning.
- Look for smaller words within the word (for example, *music* in the word *musician*).

If the clunk is still not understood after using all four fix-up strategies, students request a definition of the word from the teacher.

The third CSR reading comprehension strategy is *Get the Gist.* Its purpose is to teach students to compose a main idea conveying the essence of a paragraph or short segment of the text. Specifically, students are taught to:

- Identify the most important *who* or *what* in the section of the text they have just read.
- Identify the critical information about that *who* or *what*.
- Synthesize the information.
- Write a main idea in a complete sentence of 10 words or fewer.

The fourth CSR reading comprehension strategy, *Wrap up,* is designed to teach students to identify the most important ideas in the passage they have read and to help them understand and

remember what they have learned by developing and answering questions and writing a summary. This strategy is used after students have read the text passage. Students write questions in their learning logs and then take turns asking and answering questions before writing a summary of the portion of text they have read.

When CSR practices are fully implemented in classrooms, students collaborate in small cooperative learning groups to apply these explicit reading comprehension strategies while reading and discussing text. During the professional development, CSR trainers provided teachers with opportunities to practice each strategy, using examples developed for adults.

Training segment 3: Implementation practices and logistics

Training segment 3 provides teachers with practical information and procedures related to the logistics of implementing CSR in the classroom, with particular attention to the small, cooperative learning groups in which students work during implementation of CSR.[32] (See appendix M for a list of discussion topics.)

Because the cooperative learning groups should be composed of students with varying abilities, it is not uncommon for groups to be made up of at least one student who reads below grade level. Therefore, teachers were taught how to provide accommodations for each segment of the CSR process to optimize the ability of these students to benefit from the group and become contributing members. For example, for the preview reading comprehension strategy, teachers were taught that students who have difficulty reading can copy words from the reading selection to indicate what they think they might learn. If the story is about insects, they can copy words such as *spiders, ants,* and *bugs* in their learning logs.

Training segment 4: Coaching and materials

Training segment 4 provides coaches an opportunity to begin building rapport with teachers by interacting with them and providing assistance. Time is provided for teachers to meet with their coaches to remedy any confusion on the content and for coaches to offer tips for preparing and managing materials. Coaches are also expected to discuss how and when CSR would be integrated into weekly schedules and to schedule dates for the first follow-up coaching session.

Coaches and follow-up coaching sessions

It is critical that all coaches understand the importance of integrating the tenets of explicit comprehension instruction into the social studies lessons. Therefore, coaches were selected based on their knowledge of explicit reading comprehension instruction, including the phases of scaffolded instruction (modeling, teacher-assisted, and independent); metacognition; and self-regulatory knowledge. Minimum formal qualifications for the coaching positions were a bachelor's degree, an elementary education or reading endorsement, three years of experience teaching reading, and experience conducting classroom observations.[33]

Three coaches participated in the study. Two had master's degrees, and one had a doctoral degree. All three coaches worked as researchers, either part time or full time, and each had more than 20 years of teaching experience.[34]

---

[32] Appendix L details the four-step procedure for assigning students to cooperative learning groups.
[33] The minimum qualifications were selected by Dr. Dimino, a codeveloper of CSR.
[34] Dr. Dimino knew the three coaches who participated in the current study; he recruited them and ensured that they

The coaches were trained at the same time by Dr. Dimino, in a session that was not attended by external observers. They participated in the initial two-day teacher training session and four follow-up coaching sessions. Coaches were assigned to teachers based on geography and case load; except for rare instances in which a coach was ill, the same coach provided all follow-up coaching sessions to their assigned teachers. [35] Coaches were responsible for maintaining coaching logs (described in chapter 2), in which they recorded details such as the dates of and notes from the follow-up coaching sessions. No data were collected to evaluate consistency across coaches.

The follow-up coaching sessions consisted of two components. First, coaches observed teachers during a social studies class period (of about 45 minutes). Based on their observations, they provided feedback in the form of "kudos" (praise) or "food for thought," which included specific recommendations for the area or areas of instruction that needed to be modified. Second, the coach and teacher met later the same day to discuss what occurred during the observed lesson. At that meeting, the coach gave the written kudos and food for thought to the teacher.

Four follow-up coaching sessions were provided (two in the fall and two in the spring). The fall coaching sessions focused on whole class comprehension instruction. Coaches used the fall coaching observation form to provide specific recommendations for the area or areas of instruction that needed to be modified (appendix J). The spring coaching sessions focused on behaviors and instruction when students were working in small cooperative groups. During the spring coaching sessions, coaches were instructed to use the CSRIVC checklist measure to generate feedback.

Fall coaching sessions: explicit comprehension instruction

During the fall coaching sessions, coaches observed CSR teachers teaching reading comprehension strategies. Coaches made detailed observation notes and then completed the fall coaching observation form (which served as the coaching log for the fall coaching sessions). The form contained items related to whole-class instruction under the categories of critical procedural behaviors (instructional behaviors teachers should exhibit during the modeling and teacher-assisted phases of instruction) and critical scaffolding features (features added to instruction during the teacher-assisted phase, during which the teacher gradually releases responsibility to the students).

Key critical procedural behaviors pertaining to the modeling phase include the following (see appendix M for a complete list of procedures):

- The teacher explains the strategy.
- The teacher explains when the strategy is used.
- The teacher explains why the strategy is important.
- The teacher models how to perform the strategy.

---

met the formal qualifications.

[35] Coaching assignments were based on geography. As a result, a potential coach effect is confounded with the district effect. Therefore, coach and district effects cannot be distinguished and whether the effectiveness of CSR varied by coach cannot be estimated.

The critical scaffolding features of the teacher-assisted phase included the following:

- The teacher provides specific feedback for correct responses by communicating clearly what the student did correctly.
- The teacher elaborates on students' responses to clarify or extend them.
- When responses are incorrect, vague, or insufficient, the teacher asks further questions to help students generate a correct response.
- When students are not responding or are responding incorrectly, the teacher adjusts instruction by re-teaching or clarifying instruction.

Spring coaching sessions: CSR cooperative groups

During the spring coaching sessions, coaches observed and recorded student behaviors while students worked in cooperative groups. The CSRIVC served as the coaching log for the spring coaching sessions. Student behavior items on the CSRIVC addressed procedures students should implement during group work (for example, writing and sharing predictions and brainstorms, gists, and summaries, and determining the meaning of words). Teacher behavior items on the CSRIVC included items such as pre-teaching vocabulary, monitoring groups, providing sufficient time for group work, and conducting a class debriefing. Classroom setting items addressed the general classroom climate and organization.

Cost of implementation

The cost of implementing CSR included two components: the cost of professional development and costs that occurred throughout the school year during implementation. The costs for professional development included the costs of the CSR trainers (including fees, travel, lodging, and meals); teacher stipends; substitutes; training materials; and facility rentals. If CSR training was conducted in the summer or over the weekend, teachers received stipends, at an average hourly summer/weekend rate of $30. If CSR training was conducted during regular school days, requiring the school to hire a substitute for the teachers attending training, the school was reimbursed for the cost of the substitute, at an average daily rate of $90. Costs that occurred throughout the school year included the cost of coaching.

The average cost of CSR was calculated to be about $6,176 per school ($n$ = 26), $4,340 per CSR teacher ($n$ = 37), and $149 per student ($n$ = 1,077). These calculations are based on all of the students at baseline (n = 1,077) in classrooms in which study teachers used CSR. They include the "extra" classrooms taught by teachers in schools with departmentalized instruction (even though only one classroom was randomly selected for each of these teachers to be part of the analytic sample), as well as students who did not have parental consent to participate in the data collection but who still received CSR in their social studies classroom (n = 184).

District costs to implement CSR would likely differ from the cost of implementation in this study for several reasons. First, while five separate trainings were conducted for this study, a district could reduce costs by training all teachers at one time. Second, while stipends were used to compensate teachers for training time in the current study, a district could reduce costs by conducting training during teacher in-service days. Third, coaching costs associated with travel to the district could be reduced by training teachers and/or administrators to serve as coaches.

However, district implementation costs could be higher than in this study if the district opted to send teachers to off-site training and accrued travel costs as a result.

Assessment of fidelity

Fidelity was assessed for both the professional development provided to teachers and the implementation of CSR. Fidelity of professional development was assessed using training observation sheets, coaching logs (including fall coaching observation forms and the CSRIVC), and data from the spring survey of teachers. The fidelity of CSR implementation was assessed using data from the spring survey of teachers and the CSRIVC.

Fidelity of professional development

The fidelity of both the initial training and the follow-up coaching was assessed.

Initial training

Teacher training occurred before pretesting (timing varied, based on the district's schedule [see table 2-1]). Ninety-two percent of teachers attended both days of training. Some teachers could not be present because of medical reasons; therefore, they received a DVD of the training and accompanying materials and a follow-up call from their assigned coach. As a result, all CSR teachers were trained.

Training observation sheets, completed by data collection team members, were used to collect data about teacher attendance and content coverage during the five initial two-day training sessions (one training session per district). Data obtained from the training observation sheets indicated that all content was delivered. Average instructional training time (excluding logistics, such as completing fall and signing in, and breaks) was 7 hours and 57 minutes, with a standard deviation of 41 minutes (table 3-1). Although formal targets were not established, average training time per segment met the intended training intensity. Total training time ranged from 427 minutes to 537 minutes. Average total training time (475.8 minutes) amounted to just 79 percent of the planned delivery time (10 hours). Additional time was allotted to allow for teacher questions and elaboration, as necessary.

Table 3-1. Instructional time by topic during initial two-day training

(minutes)

Statistic	*Study overview*	*CSR training segment 1: Theoretical foundations*	*CSR training segment 2: Comprehension strategies*	*CSR training segment 3: Implementation practices and logistics*	*CSR training segment 4*[a]		*Total training time*
					Coaching	*Materials*	
Mean	23.2 (4.4)	55.6 (8.1)	223.8 (11.7)	45.2 (24.5)	93.8 (39.0)	28.2 (13.1)	475.8 (41.1)
Minimum	18	47	212	26	61	13	427
Maximum	29	67	243	77	160	49	537

Note: Figures in parentheses are standard deviations.

a. CSR training segment 4 consisted of two distinct sections, coaching and material usage.

Source: Authors' analysis of study records.

Follow-up coaching sessions

Coaches maintained coaching logs that documented training dates and notes about teachers' use of CSR (see table 2-1). These data were collected using fall coaching observation forms for the fall follow-up coaching sessions and the CSRIVC for the spring follow-up coaching sessions. Analysis of the teaching episode and the kudos and food for thought statements that were generated from the analysis were recorded. Throughout the year, Dr. Dimino monitored the coaching logs and contacted the coaches within 24 hours if there were any questions or concerns.

Data from the coaching logs were checked for accuracy against data from the CSR spring teacher surveys administered at the end of the school year. This survey was administered to teachers concurrent with posttest student data collection. It included questions about the number of follow-up coaching sessions CSR teachers received.

The mean number of follow-up coaching sessions reported by CSR teachers who answered the questions about frequency of CSR usage was 4.1 sessions (standard deviation = 1.2). Given that only four coaching sessions were scheduled per teacher, this number may indicate that some teachers included phone contacts or other brief conversations with coaches that did not constitute full follow-up coaching sessions. In most cases, follow-up coaching sessions were delivered as planned.[36]

Fidelity of CSR implementation

Fidelity of CSR implementation was evaluated for two criteria: frequency of teacher use of CSR and procedural fidelity in CSR classrooms. It is important to acknowledge that available data are limited to the percentage of teachers and students who engaged in specific CSR behaviors, and that this information is based on self-reported data (frequency of teacher use of CSR) and single observations (procedural fidelity). The available measure for observing the use of CSR (the CSRIVC) focuses only on whether a procedure was used, and not whether it was used well. The result is a simple, descriptive approach to measuring fidelity that provides a snapshot of intervention implementation.

Frequency of teacher use of CSR

Data from the CSR teacher spring survey were used to determine how frequently teachers used CSR during social studies instruction. Use of CSR two or more times a week was targeted. About 85 percent of CSR teachers reported that frequency of use, with about 15 percent of CSR teachers reporting using CSR less than twice a week (table 3-2).

Table 3-2. Teacher reports of use of CSR in classroom

Frequency	*Percent of teachers (*n=*33[a])*
Once a week	15
At least twice a week	85

a. The total number of participating CSR teachers was 37; 4 CSR

[36] Reasons for not completing all four follow-up coaching sessions included maternity leave and a school-wide administrative decision for the school to shift away from social studies instruction toward a focus on the content covered by state testing.

teachers did not respond to the survey, reducing the sample size to 33.
Source: Authors' analysis of CSR teacher survey administered in the spring.

Procedural fidelity in CSR classrooms

The CSR developers created the CSRIVC to evaluate procedural fidelity in CSR classrooms (Vaughn et al. 1998); it should therefore have strong content validity. The study team adapted the CSRIVC for use in this study by both coaches when engaged in coaching activities and by observers. CSR coaches, who functioned as trained observers for teachers they were not coaching, conducted observations in late spring (April or early May), when all aspects of the intervention should have been implemented. As described in chapter 2, the CSRIVC was used to record information about the presence or absence of CSR activities. The measure assesses teachers' instructional behaviors (that is, teachers' use of core CSR strategies) as well as the degree to which students are using CSR procedures. The observations lasted a full class period, averaging 45 minutes. All 37 CSR classrooms were observed a single time by a single observer.[37]

To evaluate the fidelity of CSR implementation, the study team developed two subindexes based on items specific to teacher strategies and student strategies.[38] The teacher strategies index consisted of five items:

1. Conduct a whole class preview to introduce the lesson.
2. Pre-teach vocabulary.
3. Circulate among groups and monitor students' use of strategies during group work.
4. Circulate among groups and reinforce content learning during group work.
5. Provide ongoing instruction in the comprehension strategies (for example, model or coach students on how to use a strategy).

The study team designated the use of all five of these core strategies as full procedural fidelity. The CSRIVC findings indicate that 56.7 percent of teachers used three or fewer strategies and just 21.6 percent used all five (table 3-3).

Table 3-3. Use of core teacher strategies by CSR teachers

Number of core teacher strategies observed	*Number of teachers*	*Percentage of teachers*[a]
0 or 1[b]	6	16.2
2 or 3[b]	15	40.5
4	8	21.6
5	8	21.6

[37] For CSR teachers with multiple classrooms, one classroom was randomly selected for inclusion in the study, but all of these teachers' classrooms were observed using the CSRIVC. Data are presented only for the classrooms randomly selected for inclusion in the study.

[38] Not all of the 31 CSRIVC items were included in the two indexes; only items that address the core elements of CSR and are directly related to implementation of comprehension strategies or student use of the strategies were included.

a. Number of teachers (n = 37) divided by total number of observations.
b. Tallies were aggregated to ensure confidentiality.
Source: Authors' analysis of Collaborative Strategic Reading Intervention Validity Checklist data.

The student strategies index consisted of seven items:

1. Record brainstorms in learning log.
2. Record predictions in learning log.
3. Write *clunks* in learning log after reading each paragraph or section of a selection.
4. Use *fix-up strategies* if no one in the group knows what the clunk means.
5. Write a *gist* in learning log after reading each paragraph or section of a selection.
6. Generate own questions as part of a wrap-up after completing entire reading assignment and write them in learning log.
7. Write summaries in learning log as part of the wrap-up after reading a selection.

Findings for the student strategies index—developed from seven CSRIVC items that reflect core student CSR strategies (see chapter 2)—indicate that students in 13 of the 37 CSR classrooms were observed using all of the core CSR strategies; students in another 10 CSR classrooms used 6 out of 7 strategies. Students in five classrooms were observed using five strategies; in nine classrooms, students used between 0 and 4 of the student strategies (table 3-4). Note that student use of core strategies is also a reflection of teacher implementation, as student use of these strategies should be influenced by how well teachers taught, guided, and supported the use of these strategies.

Table 3-4. Use of core student strategies in CSR classrooms

Number of core student strategies observed	*Classrooms (*n = *37)*	
	Number	*Percent*
0–4	9	24.3
5	5	13.5
6	10	27.0
7	13	35.1

Source: Authors' analysis of Collaborative Strategic Reading Intervention Validity Checklist data.

Control group observations

The ERC subscale was designed to document if small group instruction and explicit instruction strategies were present during a given observation. Observations of control classrooms[39] were conducted because the combined use of these approaches could be indicative of CSR use in control classrooms. Trained observers conducted a single observation in each control classroom using the ERC subscale. (A description of the ERC observer training and the ERC interrater reliability rates is provided in appendix N.) The observations, which took place during social studies instruction, were completed at the end of each treatment year. If use of small group instruction was noted, observers were asked to record if this approach was paired with explicit

[39] These observations are consistent with recommendations by Lipsey and Cordray (2000) and O'Donnell (2008).

instruction. Observers were also asked to record if there was any evidence of CSR being used by small groups.

Of the 37 classrooms, 22 did not use small group instruction, indicating that CSR was absent from these classrooms (at least during the time of observation), and observers did not further note if any explicit instruction strategies were used in those 22 classrooms. Of the 15 classrooms that used some form of small group instruction, the following four strategies were observed in at least one of the classrooms:

- Teachers explicitly assigned roles to the students.
- Teachers gave students explicit directions for implementing their roles.
- Students were expected to consistently implement their roles.
- When groups were given a written assignment, students were expected to complete each part of the assignment independently.

The last strategy, pertaining to written assignments, was observed in the most classrooms (n = 7). Other strategies included in the ERC subscale were not observed in any of the 15 control classrooms that used small group instruction. In terms of the key observational item—whether actual CSR activities were present—no observers reported any evidence the intervention was used.

Chapter 4. Impact results

This chapter presents empirical evidence on whether CSR led to higher posttest GRADE scores for grade 5 students in social studies classrooms in linguistically diverse schools. The confirmatory research question is discussed first, followed by the results of a series of sensitivity analyses conducted to determine the robustness of these findings to alternative model specifications, missing data approaches, and two different sample specification analyses (crossover and contamination analyses and data collection cohort analysis). The chapter closes with a discussion of the findings of the exploratory research questions.

Confirmatory research question

The confirmatory research question addressed in this study was:

- In linguistically diverse schools, do grade 5 students in CSR classrooms have higher average reading comprehension posttest scores on the GRADE than students in control classrooms?

The confirmatory impact analysis used a two-level HLM to estimate the effect of CSR on a reading outcome measure (the composite score of the GRADE) by comparing students in CSR classrooms with their counterparts in control classrooms in the same school. A treatment effect was estimated for each school, and the overall effect of the intervention was estimated as a weighted average of the school effects.[40]

The baseline equivalence findings presented in chapter 2 showed no differences between students in CSR and control conditions. Equivalence tests estimated by using the analytic (multiply imputed) datasets also showed no statistically significant differences between study conditions.[41] Pretest GRADE scores and language status were, nevertheless, included in the confirmatory model to increase statistical power and to account for the difference in pretest GRADE scores between FC–ELL and non–ELL students. Because a greater proportion of teachers assigned to the CSR condition reported speaking Spanish as a second language than teachers assigned to the control condition (see the *Teacher characteristics and baseline equivalence* section of chapter 2), three dummy indicator variables—little Spanish ($Spanish_2$), Spanish as a second language ($Spanish_3$), and Spanish as a first language ($Spanish_4$)—were added at the teacher level to control for these differences, using no Spanish ($Spanish_1$) as the reference group.

Results

Student scores on the GRADE reading measure were used to examine the impact of CSR on student reading comprehension. The analysis focused on a reading comprehension score that includes subscale scores for vocabulary, sentence comprehension, and passage comprehension. The mean difference in the GRADE test score of students in CSR and control classrooms was

[40] The number of students in the study sample in a given school relative to the total number of students in the study yielded a proportion that was used to weight the school-specific estimates to obtain the combined treatment effect. See appendix E for a detailed description of the analytic models.

[41] See appendix O for descriptive statistics regarding the GRADE score; see table P-1 in appendix P for baseline equivalence results for the multiple-imputed sample.

0.66, which translates into a not statistically significant ($p = 0.11$) effect size (Hedges' g) of 0.05 (table 4-1; see appendix Q for full analytic output).

Table 4-1. Estimated impact of CSR intervention on reading comprehension

Outcome	*Mean CSR*	*Mean control*	*Mean estimated impact*	*p-value*	*95% confidence interval*	*Effect size (Hedges' g)*
GRADE composite score	98.67 (12.06)	98.01 (11.71)	0.66 (0.40)	0.11	(–0.15, 1.44)	0.05

Note: Figures in parentheses are standard deviations (adjusted for multiple imputation) for CSR and control groups and standard errors for estimated impact. Results are based on 10 multiply imputed datasets. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,355 students (681 CSR, 674 control). CSR and control group means are regression adjusted for students' pretest GRADE score and language status differences, as well as for differences in teachers' Spanish fluency.
Source: Authors' analysis of student GRADE scores.

The impact presented in table 4-1 reflects an average impact across all schools. The number of students in the study sample in a given school relative to the total number of students in the study yielded a proportion that was used to weight the school-specific estimates. The impact, however, was not uniform across schools, indicating that CSR was more effective in some schools than in others (F-test estimate 1.68, p-value = 0.02).[42] (See appendix Q, table Q-2 for more information.)

Sensitivity analyses

Nine sensitivity analyses were conducted. Table 4-2 presents the results of analyses 1–6; table 4-3 presents the results of analyses 7–9.

The impact estimate for each of these analyses was not statistically significant. Across all of the sensitivity analyses, the results were consistent with the confirmatory impact analysis. Thus, it can be reasonably concluded that the confirmatory impact analysis provides results that are robust with respect to the particular set of analytic decisions for which sensitivity analyses were conducted. (See appendix Q for full analytic output.)

Table 4-2. Results of sensitivity analyses of estimated impact of CSR

Model	*Type of sensitivity analysis*	*Mean CSR*	*Mean control*	*Mean estimated impact*	p-*value*	*95% confidence interval*	*Effect size (Hedges' g)*
2[a]	Equal weighting	98.57 (12.06)	98.01 (11.71)	0.55 (0.43)	0.20	–0.30, 1.40	0.05
3[a]	Random effects	99.08 (12.06)	98.27 (11.71)	0.80 (0.46)	0.09	–0.17, 1.77	0.07
4[b]	List-wise deletion	98.20 (11.99)	97.70 (11.00)	0.49 (0.43)	0.26	–0.40, 1.38	0.04
5[c]	Exclusion of crossover students	98.69 (12.06)	98.02 (11.74)	0.66 (0.41)	0.11	–0.14, 1.46	0.06

[42] Only schools with more than two classrooms contribute to the estimation of variation in the intervention effect.

Model	*Type of sensitivity analysis*	*Mean CSR*	*Mean control*	*Mean estimated impact*	p-*value*	*95% confidence interval*	*Effect size (Hedges'* g)
6[d]	Exclusion of schools with increased contamination risk	98.99 (12.10)	98.21 (11.81)	0.77 (0.42)	0.07	–0.06, 1.59	0.06
7[e]	Unconditional treatment effect	98.23 (12.06)	98.11 (11.72)	0.11 (1.08)	0.92	–2.00, 2.29	0.01

Note: Figures in parentheses are standard deviations (adjusted for multiple imputation) for CSR and control groups and standard errors for estimated impact. Results are based on 10 multiply imputed datasets. CSR and control group means are regressions adjusted for differences in students' pretest GRADE scores and language status (except in model 3), as well as differences in teachers' Spanish fluency.

a. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,355 students (681 CSR, 674 control).
b. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,123 students (543 CSR, 580 control).
c. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,341 students (673 CSR, 668 control).
d. Analysis included 24 schools, 69 teachers (34 CSR, 35 control), and 1,282 students (631 CSR, 651 control).
e. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,355 students (681 CSR, 674 control).

Source: Authors' analysis of student GRADE scores, study records, and teacher survey.

Table 4-3. Results of sensitivity analyses of estimated differential impact of CSR

Model	*Type of sensitivity analysis*	*Mean estimated differential impact*	*95% confidence interval*	p-*value*
8[a]	Heterogeneity of slopes	0.01(0.05)	–0.08, 0.11	0.77
8[b]	Heterogeneity of slopes	0.00(0.05)	–0.09, 0.09	0.96
9[c]	Cohort effect	–4.25(3.60)	–11.54, 3.04	0.25

Note: Figures in parentheses are standard errors.
a. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,355 students (681 CSR, 674 control).
b. Analysis included 23 schools, 66 teachers, and 1,199 students (279 FC–ELL CSR, 256 FC–ELL control; 322 non–ELL CSR, 342 non–ELL control).
c. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,355 students (681 CSR, 674 control).
Source: Authors' analysis of student GRADE scores, study records, study records, and teacher surveys.

Exploratory research questions

The three exploratory research questions examined whether CSR is effective for FC–ELL and non–ELL students taught together in heterogeneous classrooms.

Two types of analyses were conducted to examine the impact of the CSR intervention on reading comprehension for these subgroups. First, the study team estimated the effects of CSR for each subgroup. (See appendix E for model specifications.) Second, the study team examined whether CSR had a differential effect on the subgroups—that is, whether it worked better for FC–ELL or non–ELL students. These analyses are exploratory because of their reduced statistical power. They are based on the subset of the school sample in which FC–ELL and non–ELL students were present in both treatment and control classrooms. Equally weighted estimates of the effects of CSR for each subgroup are presented.

The results indicate an estimated effect of the CSR intervention for FC–ELL students of 0.94 points on the GRADE; the result was not statistically significant ($p = 0.47$). The total score

difference between CSR and control group FC-ELL students of 0.94 in the GRADE test translates to an effect size of 0.08. For non–ELL students, the estimated effect of the CSR intervention was 0.69 points on the GRADE; the result was not statistically significant (p = 0.22). This total score difference between CSR and control group non-ELL students of 0.69 on the GRADE translates to an effect size of 0.06. These results indicate no differential effect for the FC–ELL and non–ELL subgroups (the estimate for the interaction term was 0.25, with a standard error of 1.13 and a p-value of 0.82), suggesting that CSR did not work any better (or worse) for FC–ELL students than for non–ELL students (table 4-4; see appendix Q for full analytic output).

Table 4-4. Effect of CSR on GRADE composite scores

Student group	*CSR*	*Control*	*Estimated impact*	p-*value*	*95% confidence interval*	*Effect size (Hedges'* g*)*
FC–ELL	98.49 (11.60)	97.55 (10.98)	0.94 (1.30)	0.47	(–1.62, 3.51)	0.08
Non–ELL	98.69 (12.06)	98.00 (11.94)	0.69 (0.57)	0.22	(–0.42, 1.81)	0.06

Note: Data are mean values; figures in parentheses are standard deviations for CSR and control groups and standard errors for estimated impact. Analysis included 23 schools, 66 teachers, and 1,199 students (279 FC–ELL CSR, 256 FC–ELL control; 322 non–ELL CSR, 342 non–ELL control). CSR and control group means were regressions adjusted for differences in pretest GRADE scores, language status, as teachers' Spanish fluency.

Source: Authors' analysis of student GRADE scores.

Summary

CSR did not have a statistically significant effect on students' reading comprehension scores as measured by the GRADE (effect size = 0.05). This result is robust to changes in estimation methods and the primary missing data approach; results from additional sensitivity analyses accounting for crossovers and potential study cohort differences are similar. The effect of CSR is not uniform, however, and the variation in treatment effect is statistically significant across schools. The CSR effects for the FC–ELL and non–ELL subgroups separately were not statistically significant, and CSR did not have statistically significant differential effects on the two subgroups.

Chapter 5. Summary of key findings and study limitations

This chapter summarizes the study's main findings. It also identifies the study's limitations and offers recommendations for future research.

Did CSR improve reading comprehension?

This study was a rigorous evaluation of CSR in which grade 5 social studies classrooms were randomly assigned within each participating linguistically diverse school to either the CSR condition or the business-as-usual control condition. The outcome measure in this study was the GRADE, a nationally normed test with scaled scores reported with a mean of 100 and standard deviation of 15.

The study was powered for a minimal detectable effect size of 0.20 standard deviations. The primary finding is that CSR did not have a statistically significant impact on student reading comprehension. The mean GRADE reading posttest score for students in CSR classrooms in linguistically diverse schools was 98.67, compared with 98.01 for students in control classrooms (a 0.66 point difference, which was not statistically significant). The effect of CSR was not uniform, however, and the variation in treatment effect is statistically significant across schools. Nine sensitivity analyses showed that this finding was robust to various analytic decisions and assumptions, including alternative statistical models and an alternative approach for handling missing data.

Three exploratory analyses were conducted to examine the effects of CSR on FC–ELL and non–ELL students separately and to determine whether differential effects were observed. For FC–ELL students, the average posttest GRADE score was 98.49 in CSR classrooms and 97.55 in control classrooms (a 0.94 difference, which was not statistically significant). For non–ELL students, the average posttest GRADE score was 98.69 in CSR classrooms and 98.00 in control classrooms (a 0.69 difference, which was not statistically significant). CSR did not have a statistically significant differential effect on these two subgroups (estimate for the interaction term = 0.25, standard error = 1.13, $p = 0.82$).

Fidelity of implementation was evaluated for both the professional development delivered to teachers and the implementation of CSR by those teachers in the classroom. Teacher professional development was delivered as planned. Teachers were instructed to use CSR two or more times a week; 79 percent of teachers reported doing so. During a single classroom observation, 8 of 37 CSR teachers in the study (21.6 percent) were observed using all five core teaching strategies, another 8 were observed using at least four of the five strategies, and 21 were observed using three or fewer strategies. During the classroom observation, students in 13 of the 37 CSR classrooms (35.1 percent) were observed using all seven core student strategies; in 9 of the observations, none of the strategies was observed.

Contribution of this study

A quasi-experimental efficacy study described in chapter 1 concluded that CSR had positive impacts on all students exposed to the intervention and that these impacts were stronger for ELL students (Klingner, Vaughn, and Schumm 1998).[43] The current study did not find statistically significant impacts.

Several factors may account for the different conclusions of the two studies. First, randomized controlled trials often fail to support the findings of quasi-experiments (Glazerman, Levy, and Myers 2002, 2003), because holding all other design features equal, randomization yields stronger internal validity about program impacts than quasi-experiments (Boruch 1997; Shadish, Cook, and Campbell 2002). Second, this study was an effectiveness trial whereas the previous study was an efficacy trial. In the current study, CSR was implemented using an approach designed to approximate real-world implementation: teachers were trained to deliver CSR to their own students. By contrast, in the earlier study, the CSR developers delivered CSR directly to participating students, a method of delivery that would not be feasible for large-scale implementation. Third, in the current study, CSR was implemented in 26 schools across five districts located in two states; the previous study was conducted in a single school.

Limitations of this study

This study used a convenience sample of volunteer schools rather than a randomly selected sample. Its findings therefore apply only to schools in the sample. Generalizing to other schools with characteristics similar to the participating schools should be done cautiously.

A second limitation of this study is that language status for participating students was limited to FC-ELL or non-ELL identification because other language status data—such as level of English proficiency, first language, or any specific ELL programs in which students participated—were not collected.

A third limitation is that procedural fidelity was assessed based on a single classroom observation. The CSRIVC data collected therefore apply only to the specific point in time when the observations were made; the resulting data may not generalize to the rest of the school year. This is, of course, an inherent concern with any snapshot of classroom implementation.

Finally, the CSRIVC measures whether CSR activities were observed but does not measure how well teachers or students implement the procedures. In addition, the CSRIVC does not offer a priori cutpoints for distinguishing between strong and weak implementation fidelity. Therefore, it is not possible to say how low CSRIVC scores would need to be for fidelity of implementation to be classified as "poor." The CSRIVC scores in this study can be used only to indicate that some elements of CSR were not implemented during the observed class time, which is to be expected in an effectiveness trial.

[43] Klingner, Vaughn, and Schumm (1998) did not distinguish between former and current ELL students.

Future research

Future research on CSR might focus on enhancing the fidelity of CSR implementation within classrooms to determine whether impacts that are larger and statistically significant can be measured. Future investigations could also consider using more intensive training and/or coaching delivery and investigating CSR impact at different grade levels and in different subject areas.

Appendix A. Identification and exit criteria for English language learner students in Oklahoma and Texas

In this study, student English language learner (ELL) status was provided by participating districts. Identification and exit policies for ELL students vary widely at the district level (Ragan and Lesaux 2006). For this reason, only a general overview of procedures used in Oklahoma and Texas follows.

General procedures in Oklahoma

Oklahoma uses the following criteria to designate a student as an ELL or limited English proficient student (McGavock 2009):

- The home language survey indicates that a language other than English is spoken more often than English in the home.
- A language other than English is spoken in the home and the child has a less than satisfactory score on a qualifying test.
- The student performed below various thresholds on a screening measure and on end-of-year assessments.

According to Oklahoma's *Bilingual and English Language Learner/Limited English Proficient Technical Assistance Guide* (McGavock 2009), all ELL and limited English proficient students must be assessed annually until they demonstrate English proficiency, as determined by a cutoff score on a state measure, Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs). Districts are expected to assist ELL and limited English proficient students until they reach such proficiency. The proficiency score is informed by the annual measureable achievement objectives of Title III in the 2001 No Child Left Behind Act.

General procedures in Texas

Texas uses the following steps to designate a student as an ELL:

Step 1: The school determines whether a language other than English is spoken in the home and by the student, usually using the state's home language survey, which is completed by parents or guardians. If the completed survey indicates that only English is spoken in the home, the student is not designated as an ELL. If it reveals that a language other than English is spoken in the home, the next steps are followed.

Step 2: Students are identified as bilingual if English is spoken at home along with another language. They are rated as learning English as a second language if only a foreign language is spoken at home. In kindergarten and first grade, English proficiency is tested using an oral language proficiency test approved by the Texas Education Agency. This test is supplemented with a norm-referenced standardized achievement test in grades 2–12.

Step 3: Based on various test score cutpoints, students are rated as beginning, intermediate, advanced, or advanced high in listening, speaking, reading, and writing (Texas Education Agency 2008). After proficiency is assessed, a placement designation is made by a Language Proficiency Assessment Committee, and parents are notified. Parental permission is required before a placement decision can be implemented (Texas Education Agency 2008–09a).

Step 4: Program exit decisions are based on performance on the Texas Assessment of Knowledge and Skills, the Texas English Language Proficiency Assessment System, and, depending on grade level, a Texas Education Agency–approved oral language proficiency test, as well as on teacher evaluations (Texas Education Agency 2008–09b).

Appendix B. Assumptions used to determine statistical power and observed power

The minimal detectable effect size represents the smallest true program impacts, in standard deviation units, that can be detected with high probability (Bloom 2005). All else being equal, the smaller the effect size to be detected, the larger the study sample must be. The minimal detectable effect size selected should be large enough that the detectable impact is important yet small enough to be feasible given the intervention.

Based on previous literature syntheses, information from the What Works Clearinghouse, feedback from the study's technical working group members (a group with considerable experience with randomized controlled trials), and other experts in the field, it appeared that an effect size of 0.20 was appropriate for the primary confirmatory hypotheses in this study of Collaborative Strategic Reading (CSR). Given that the Group Reading Assessment and Diagnostic Evaluation (GRADE) is a nationally normed test in which scaled scores are normally distributed with a mean of 100 and a standard deviation of 15, an effect size of 0.20 corresponds to a scaled score difference of 3 points on the GRADE.

Assumptions made for the power analysis of the full analytic sample

The power analysis assumes a design in which classrooms are randomly assigned within sites (for this study, sites are schools) and site effects are treated as fixed. The power calculations are derived from the following assumptions:

- Number of schools and classrooms: 26 schools and 74 classrooms (37 CSR, 37 control).
- Desired statistical power: 80 percent.
- Statistical significance level: 0.05 (two-tailed); no adjustment for multiple comparisons.
- Number of grade 5 teachers per school—assume an average of three per school.[44]
- Number of students per classroom: 20, with 80 percent posttest response rates (16 students per classroom provide both pretest and posttest data).
- Proportion of classrooms in CSR condition: 50 percent under a balanced sample allocation.
- School level: modeled as fixed effects.
- Intraclass correlation assumed 0.15 (actual 0.16 at the school level, 0.09 at the classroom level).
- Exploratory power of the pretest: $r = 0.84$; $R^2 = 0.71$.
- Number of school districts: 5.

[44] It was reasonable to assume minimal teacher turnover because CSR was implemented for one academic year (fall semester/spring semester) in participating schools. Most teacher turnover takes place between school years (in effect, during summer); teacher turnover that takes place during a school year is typically due to events such as pregnancy or illness. In addition, replacements for CSR condition teachers lost to attrition would have received CSR training similar to that received by the original study teachers.

Assumptions made for power analysis of subsamples

The power calculations are based on the following assumptions. Statistical power for former and current English language learner (FC–ELL) students and for non–ELL students is presented separately.

- Number of schools and classrooms: 23 schools and 66 classrooms (33 CSR, 33 control).
- Desired statistical power: 80 percent.
- Statistical significance level: 0.05 (two-tailed).
- Number of grade 5 teachers per school: 3.
- Number of students per classroom: 20, with 80 percent posttest response rates (16 students per classroom provide both pretest and posttest data). On average, 43 percent of students were FC–ELL students (seven students per classroom). However, because there was a wide distribution in the number of FC–ELL students across classrooms, a conservative assumption of four FC–ELL students per classroom was used when estimating power for the FC–ELL analysis. For the non–ELL analysis, an assumption of eight students per classroom was used.
- Proportion of classrooms in CSR condition: 50 percent under a balanced sample allocation.
- School-level: modeled as fixed effects.
- Intraclass correlation assumed 0.15 (actual 0.16 at the school level, 0.09 at the classroom level).
- Explanatory power of the pretest: $r = 0.85$, $R^2 = 0.72$ for non–ELL student sample; $r = 0.80$, $R^2 = 0.64$ for FC–ELL student sample, with resultant error reduction.
- Number of school districts: 5.

Target sample size

Table B-1 displays findings from the power analyses incorporating the above assumptions.

Table B-1. Minimal detectable effect size for 66 classrooms in 23 schools

Model/sample	R^2	*Minimal detectable effect size*
Fixed effect, FC–ELL student sample	0.64	0.25
Random effects, non–ELL student sample	0.72	0.19

Source: Authors' power analyses.

Observed minimal detectable effect size

The actual observed minimal detectable effect size in the full sample and the subsamples was calculated as follows:

$$\text{Minimal detectable effect size} = \text{factor}(\alpha, \beta, df) * \sqrt{\text{Var(impact)}} / \sigma$$

where *Var(impact)* is the variance of the impact estimate; σ is the pooled standard deviation of the outcome measure; *Factor (*α, β, df) is a constant that is a function of the significance level (α), statistical power (β), and number of degrees of freedom (df). (*Factor(.)* becomes larger as the significance level falls and as the power level rises.)

The observed minimal detectable effect size for the full sample analysis is 0.10 (0.13 for the non–ELL student subgroup and 0.34 for the FC–ELL student subgroup).

Appendix C. Random assignment

The design for this study used schools as a blocking factor, with classrooms randomly assigned within each school. Therefore, each site had at least one Collaborative Strategic Reading (CSR) classroom and one control classroom. This appendix details the random assignment process and changes to the school and classroom samples after random assignment was completed.

The random assignment of classrooms to either the CSR or control condition was done by a statistician who was removed from all aspects of intervention delivery. Each school had two to four eligible classrooms, which were randomly assigned to a study condition. The random assignment procedure was done independently in each of the five school districts in the study. For each district, a Microsoft Excel® file was created, with a list of schools and classrooms within those schools. Using Excel's rand () function, a random number between 0 and 1 was assigned to each classroom in the district. The classrooms within each school were then sorted according to the assigned random numbers and then assigned to either the CSR or the control condition.

Within each school, principals were asked to provide a list of the teachers and students assigned to each grade 5 classroom. The student rosters for these classrooms were prepared by the schools using their regular classroom assignment procedures before random assignment. Each participating teacher was notified by the study team of the assigned condition of his or her classroom before participating in the initial two-day teacher training. This information was also shared with principals and district staff.

The random assignment procedure was conducted separately for each school. For schools with an even number of classrooms, the first half (those with the smaller random numbers) were assigned to the CSR group and the second half to the control condition. If a school had an odd number of classrooms, the extra classroom was assigned to either the CSR or the control condition using a process that ensured that the total number of CSR and control classrooms in the district was as equal as possible. This three-step process was as follows:

- *Step 1*: For all schools in which random assignment had already occurred, the total number of CSR and control classrooms was calculated. Whether there was an equal number of CSR and control classrooms or whether there was one more CSR classroom or more control classroom was noted.
- *Step 2:* For schools in which there was an odd number of *Z* classrooms, the first *Z*/2–0.5 classrooms were assigned to the CSR condition and the last *Z*/2–0.5 classrooms were assigned to the control condition. For example, if there were five classrooms, the first two would be assigned to the CSR group and the last two to the control group. This left one extra classroom unassigned to a study condition.
- *Step 3:* The unassigned classroom for the school was assigned based on the results of step 1. If more control than CSR classrooms had already been assigned, the unassigned classroom was placed in the CSR condition. If more CSR than control classrooms had already been assigned, the unassigned classroom was placed in the control condition. If the number of CSR and control classrooms already assigned was equal, the unassigned classroom was randomly assigned to a condition, using the following procedure. The unassigned classroom

was given a new random number ranging from 0 to 1. If the number was less than 0.5, the classroom was assigned to the CSR group; if the number was more than 0.5, the classroom was assigned to the control group. This process ensured that the overall number of CSR and control classrooms for a district was the same or at most differed only by one classroom.

Appendix D. Analysis of consent rate at baseline

For binary outcomes such as consent status, a hierarchical generalized linear model can assume a Bernoulli sampling model and a logit link function. The following three-level model examines whether study condition predicts students' consent status at baseline (1 = consent, 0 = no consent).

At the student level:

$$\text{Prob(Consent at baseline)} = 1|\text{CSR} = \varphi$$

$$\text{Log}[\varphi/(1-\varphi)] = \eta_{ijk}$$

$$\eta_{ijk} = \pi_{0jk} + e$$

where

- $\eta_{ijk} = \log(\varphi / 1 - \varphi)$ (the log of the odds of a student having a parental consent at baseline)
- π_{0jk} is the average log odds of consent of students in class j in school k
- e is an error term $e \sim 1/(\varphi\ (1-\varphi))$, where the probability of $(Y = 1|\ \text{CSR}) = \varphi$.

The classroom average outcome estimated from this model (in effect, the level 1 intercept π_{0jk}) was modeled as varying randomly across classrooms and as a function of the study condition at level 2, the classroom level:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}*(CSR)_{jk} + r_{0jk}$$

where

- β_{00k} is the average log odds of consent across all classrooms in school k
- CSR is a school-mean centered indicator variable for the intervention
- β_{01k} is the difference in consent rates in Collaborative Strategic Reading (CSR) and comparison classrooms in school k
- r_{0jk} is a random error associated with classroom j in school k on classroom average log odds of consent; $r_{0jk} \sim N(0, \tau_{00\,k})$.

In the level 3 model, the teacher average outcome (β_{00k}) was modeled as a random effect, on the assumption that the classroom average may differ systematically across schools. The difference in the log odds of consent between CSR and control classrooms was assumed to be constant across schools:

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

$$\beta_{01k} = \gamma_{010}$$

where

- γ_{000} is the average log odds of consent across all schools
- u_{00k} is a random error associated with school k on school average log odds of consent; $u_{00k} \sim N(0, \tau_{000})$

- γ_{010} is the average CSR effect across all schools.

The key parameter of interest is γ_{010}, the overall difference between students in CSR and in control classrooms in the log odds of having consent at baseline. The results indicate that the relationship between study condition and parental consent status at baseline is not statistically significant (tables D-1 and D-2).

Table D-1. Returned parental consent forms in CSR and control groups

Group	*Number*	*Percent*
CSR	702	84.9
Control	649	78.6
Total	1,351	81.7

Source: Authors' analysis of study records.

Table D-2. Three-level hierarchical generalized linear model of student consent

Item	*Value*
CSR indicator	0.30
Standard error	0.25
p-value	0.23

Source: Authors' analysis of study records.

.

Appendix E. Estimation methods

This appendix presents the estimation models used in the confirmatory, sensitivity, and exploratory analyses briefly described in chapter 2.

Model 1: Specifications of confirmatory impact analysis

The model used to estimate the effect of Collaborative Strategic Reading (CSR) acknowledges potential systematic variation in the treatment effect across schools. It included pretest Group Reading Assessment and Diagnostic Evaluation (GRADE) and language status variables at the student level. At the teacher level, it included indicator variables for teachers' Spanish fluency, in order to account for significant differences at baseline. Spanish fluency was measured by four indicator variables (no Spanish = $Spanish_1$, little Spanish = $Spanish_2$, Spanish as a second language = $Spanish_3$, and Spanish as a first language = $Spanish_4$); $Spanish_1$ was the omitted category.

The level 2 equation omitted the level 2 intercept; it therefore included all 26 school indicator variables and 26 CSR*school indicator interaction terms. The omission of the intercept in the level 2 equation and inclusion of all 26 school indicator variables and CSR* school interaction terms provided estimates of the effect of CSR for each school separately. The overall effect is the sum of all 26 school-specific CSR effects (from the *CSR*school* interaction terms), weighted by the proportion of students in the sample in each school.

At level 1 (the student level):

$$Y_{ij} = \beta_{0j} + \beta_{1j}*(Pretest)_{ij} + \beta_{2j}*(Language\ Status)_{ij} + r_{ij}$$

where

- Y_{ij} is the outcome for student *i* in classroom *j*
- $(Pretest)_{ij}$ is the pretest score for student *i* in classroom *j*, grand–mean centered
- $(Language\ Status)_{ij}$ is the language status variable (FC–ELL = 1; non–ELL = 0) for students in classroom *j*, grand–mean centered
- β_{0j} is the average outcome of students in classroom *j*
- β_{1j} is the association between pretest GRADE scores and student outcomes (the expected change in achievement outcome when the pretest GRADE score increases by one unit, holding language status constant)
- β_{2j} is the association between language status and student outcomes (the expected difference in achievement outcome between former and current English language learner (FC–ELL) students and non–ELL students, holding the pretest GRADE score constant)
- r_{ij} is a random error associated with student *i* in classroom *j*; $r_{ij} \sim N(0, \sigma^2)$.

At level 2 (the classroom level):

$$\beta_{0j} = \gamma_{01}(Spanish_1)_j + \gamma_{02}(Spanish_2)_j + \gamma_{03}(Spanish_3)_j +$$
$$\delta_k \sum_{k=1}^{26} School_k + \lambda_k \sum_{k=1}^{26} School_k * CSR_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$

where

- *CSR* is an indicator variable for the intervention condition (1 = *CSR* classroom, 0 = control classroom)
- $Spanish_1$, $Spanish_2$, and $Spanish_3$ are indicator variables for teachers' Spanish fluency, grand–mean centered
- $School_k$, k = 1, 2, . . . , 26, are dummy indicator variables representing the 26 schools in the sample
- $School_k*CSR_j$ are 26 intervention-by-school interaction terms
- γ_{01} is the association between teachers speaking a little Spanish and the average student outcome, holding other level 2 covariates constant
- γ_{02} is the association between teachers speaking Spanish as a second language and the average student outcome, holding other level 2 covariates constant
- γ_{03} is the association between teachers speaking Spanish as a first language and the average student outcome, holding other level 2 covariates constant
- δ_k, k = 1, 2, …, 26, represents the 26 school intercepts
- λ_k, k = 1, 2, …, 26, represents the effect of CSR for each of the 26 schools
- γ_{10} is the fixed effect representing the average pretest across all classrooms
- γ_{20} is the fixed effect representing the average language status across all classrooms; and u_{0j} is a random error associated with classroom j on classroom average outcome; $u_{0j} \sim N(0, \tau_{00})$.

The reduced form equation is

$$Y_{ij} = \gamma_{01} * (Spanish_1)_j + \gamma_{02} * (Spanish_2)_j + \gamma_{03}(Spanish_3)_j + \delta_k \sum_{k=1}^{26} (School)_k +$$
$$\lambda_k \sum_{k=1}^{26} (School)_k * (CSR)_j + \gamma_{10} * (Pretest)_{ij} + \gamma_{20} * (Language\ Status)_{ij} + r_{ij} + u_{0j}.$$

Of primary interest is the combined effect of the λ_k coefficients for the *School*CSR* indicator interaction terms, which weighted school-specific estimates by the proportion of students in the sample in each school. The combined effect represents the intervention's main effect on the outcome across all schools. A statistically significant positive value of the combined effect would indicate that students in CSR classrooms demonstrate higher levels of reading

comprehension than their counterparts in control classrooms. School-specific effects were weighted as follows:

$$\text{Combined estimate} = \sum_{i=1}^{26} w_i b_i$$

where $\sum_{i=1}^{26} w_i = 1$, w_i is the weight, and b_i is the school specific estimate of CSR's effect. The standard error of the combined estimate is

$$\sqrt{\sum_{n=i}^{26} SEb_i^2 w_i^2 + \sum_{n=1}^{26} 2\text{Cov} b_i b_j w_i w_j} \ .$$

The school-specific weights used in different sensitivity analyses are shown in table E-1.

Table E-1. School-specific weights used in confirmatory impact and sensitivity analyses using weights

		Sensitivity analysis				
School	*Confirmatory impact analysis*	*Equal weighting*	*List-wise deletion*	*Exclusion of crossover students*	*Exclusion of schools with increased risk of contamination*	*Unconditional treatment effect*
1	0.058	0.0384	0.057	0.057	0.061	0.058
2	0.030	0.0384	0.027	0.030	0.032	0.030
3	0.047	0.0384	0.046	0.048	0.050	0.047
4	0.031	0.0384	0.028	0.031	0.033	0.031
5	0.043	0.0384	0.043	0.043	0.045	0.043
6	0.034	0.0384	0.032	0.034	0.036	0.034
7	0.035	0.0384	0.022	0.036	0.037	0.035
8	0.041	0.0384	0.041	0.040	0.043	0.041
9	0.031	0.0384	0.037	0.031	0.033	0.031
10	0.044	0.0384	0.048	0.045	0.047	0.044
11	0.038	0.0384	0.045	0.038	0.040	0.038
12	0.025	0.0384	0.028	0.025	0.027	0.025
13	0.060	0.0384	0.059	0.060	0.063	0.060
14	0.066	0.0384	0.066	0.066	0.070	0.066
15	0.039	0.0384	0.045	0.040	na	0.039
16	0.036	0.0384	0.039	0.036	0.038	0.036
17	0.028	0.0384	0.030	0.027	0.030	0.028
18	0.041	0.0384	0.038	0.042	0.044	0.041
19	0.015	0.0384	0.015	0.015	na	0.015
20	0.031	0.0384	0.036	0.031	0.033	0.031
21	0.031	0.0384	0.033	0.031	0.033	0.031
22	0.026	0.0384	0.026	0.025	0.027	0.026
23	0.069	0.0384	0.070	0.069	0.073	0.069
24	0.028	0.0384	0.021	0.028	0.030	0.028
25	0.039	0.0384	0.037	0.039	0.041	0.039

		Sensitivity analysis				
School	*Confirmatory impact analysis*	*Equal weighting*	*List-wise deletion*	*Exclusion of crossover students*	*Exclusion of schools with increased risk of contamination*	*Unconditional treatment effect*
26	0.033	0.0384	0.031	0.032	0.035	0.033

Na is not applicable.
Note: Sample size is 26 for all sensitivity analyses except exclusion of schools with increased risk of contamination.
Source: Authors' analysis of study records.

Sensitivity analyses—alternative model specifications

Eight sensitivity analyses were run on the confirmatory model. Each is described below.

Model 2: Equal weighting (sensitivity analysis 1)

The confirmatory impact model (model 1) was estimated by equally weighting each school, regardless of the proportion of students in the sample located in the schools.

Model 3: Random effects (sensitivity analysis 2)

At level 1 (the student level):

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}*(Pretest)_{ijk} + \pi_{2jk}*(Language\ Status)_{ijk} + e_{ijk}$$

where

- Y_{ijk} is the outcome for student *i* in classroom *j* in school *k*
- *Pretest* is the pretest score of student *i* in classroom *j* in school *k*, grand–mean centered
- *Language Status* (FC–ELL=1, non–ELL=0) is the language status of student *i* in classroom *j* in school *k*, grand–mean centered
- π_{0jk} is the average outcome of students in classroom *j* in school *k*
- π_{1jk} is the relationship of pretest with outcome of students in classroom *j* in school *k*
- π_{2jk} is the relationship of language status with outcome of students in classroom *j* in school *k*
- e_{ijk} is the random error associated with student *i* in classroom *j* in school *k*; $e_{ijk} \sim N(0, \sigma^2)$.

The classroom average outcomes estimated from this model (level 1 intercept π_{0jk}) were modeled as varying randomly across classrooms and as a function of the intervention at level 2 and teachers' Spanish fluency dummy indicators. The level 1 slopes (π_{1jk}) and (π_{2jk}) were modeled as a fixed effect at level 2:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}*(CSR)_{jk} + \beta_{02k}\,(Spanish_1)_{jk} + \beta_{03k}\,(Spanish_2)_{jk} + \beta_{04k}\,(Spanish_3)_{jk} + r_{0jk}$$

$$\pi_{1jk} = \beta_{10k}$$

$$\pi_{2jk} = \beta_{20k}$$

where

- *CSR* is a group–mean centered indicator variable for the intervention (CSR = 1, control = 0) for classroom *j* in school *k*

- $Spanish_1$, $Spanish_2$, and $Spanish_3$ are grand–mean centered indicator variables for teachers' Spanish fluency for classroom *j* in school *k*
- β_{00k} is the average student outcome across all classrooms in school *k*, adjusted for the pretest GRADE score, students' language status, CSR, and teachers' Spanish fluency indicators
- β_{01k} is the difference between student outcomes in CSR and control classrooms (in effect, the intervention effect) in school *k*, holding other level 2 covariates constant
- β_{02k} is the relationship between teachers speaking little Spanish and student outcomes across all classrooms in school *k*, holding other level 2 covariates constant
- β_{03k} is the relationship between teachers speaking Spanish as a second language and student outcomes across all teachers in school *k*, holding other level 2 covariates constant
- β_{04k} is the relationship between teachers speaking Spanish as a first language and student outcomes across all teachers in school *k*, holding other level 2 covariates constant
- β_{10k} is the average relationship between pretest scores and student outcomes across all classrooms in school *k*, holding other level 2 covariates constant
- β_{20k} is the average relationship between language status and student outcomes across all classrooms in school *k*, holding other level 2 covariates constant
- r_{0jk} is the random error associated with classrooms *j* in school *k* on average student outcomes; $r_{0jk} \sim N(0, \tau_{00k})$.

In the level 3 model, both average classroom outcomes and the CSR effect within each school (β_{00k} and β_{01k}) were modeled as random effects, on the assumption that both average classroom achievement and the CSR effect differ systematically across schools. In addition, the classroom average outcome and the CSR effect within each school were assumed to be potentially affected by the school district (effect coded indicator variables for school districts). The effects of pretest and indicator variables about teachers' Spanish fluency were fixed at their respective grand means at the school level:

$$\beta_{00k} = \gamma_{000} + \gamma_{001}\text{District1} + \gamma_{002}\text{District2} + \gamma_{003}\text{District3} + \gamma_{004}\text{District4} + u_{00k}$$

$$\beta_{01k} = \gamma_{010} + \gamma_{011}\text{District1} + \gamma_{012}\text{District2} + \gamma_{013}\text{District3} + \gamma_{014}\text{District4} + u_{01k}$$

$$\beta_{02k} = \gamma_{020}$$

$$\beta_{03k} = \gamma_{030}$$

$$\beta_{04k} = \gamma_{040}$$

$$\beta_{10k} = \gamma_{100}$$

$$\beta_{20k} = \gamma_{200}$$

where

- γ_{000} is the average student outcome across all schools (in effect, the grand mean)
- $\gamma_{001–}\gamma_{004}$ are the relationships between school districts and the average student outcome across all schools
- γ_{010} is the average CSR effect across all schools

- γ_{011}– γ_{014} are the relationships between school districts and the average CSR effect across all schools
- u_{00k} is the random error associated with school *k* on school average student outcome
- $u_{01k}$is the random error associated with school *k* on the CSR effect

$$\begin{pmatrix} u_{00k} \\ u_{01k} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{000} & \tau_{001} \\ \tau_{100} & \tau_{111} \end{pmatrix}\right]$$

- γ_{020} is the relationship of teachers speaking a little Spanish with the average student outcome across all schools
- γ_{030} is the relationship of teachers speaking Spanish as a second language with the average student outcome across all schools
- γ_{040} is the relationship of teachers speaking Spanish as a first language with the average student outcome across all schools
- γ_{100} is the average relationship of pretest scores with the student outcome across all schools
- γ_{200} is the average relationship of language status with the student outcome across all schools.

Of primary interest is the intervention's average effect on the outcome across all schools after holding constant the covariates in the model (γ_{010}). The interpretation of the intervention's effect, however, needs to be qualified if the effect varies significantly across schools (that is, τ_{111} is statistically significant). Detection of significant variation suggests that the intervention does not have a common effect across all schools.

Model 4: List-wise deletion (sensitivity analysis 3)

Model 1 was estimated with list-wise deletion rather than multiply imputed datasets.

Model 5: Exclusion of crossover students (sensitivity analysis 4)

Model 1 was estimated with a sample that excluded students who crossed over from the CSR to control or the control to CSR condition.

Model 6: Exclusion of schools with increased risk of contamination (sensitivity analysis 5)

Model 1 was estimated with a sample that excluded schools in which a CSR teacher started teaching social studies to control students in the middle of the school year (as a result of unexpected restructuring of instruction) or in which departmentalization of instruction increased the risk of contamination.

Model 7: Unconditional treatment effect (sensitivity analysis 6)

The following model was used to estimate the effect of CSR. It acknowledges potential systematic variation in the treatment effect across schools. The level 2 equation for the level 1 intercept omitted the level 2 intercept; it therefore included all 26 school indicator variables and 26 *School*CSR* indicator interaction terms. The omission of the intercept and inclusion of all 26 school indicator variables and *School*CSR* interaction terms provided separate estimates of the effect of CSR for each school. The overall effect is the sum of all 26 school-specific CSR effects

(from the *School*CSRI* interaction terms), weighted by the proportion of students in the sample in each school.

At the student level:

$$Y_{ij} = \beta_{0j} + r_{ij}$$

where

- Y_{ij} is the outcome for student *i* in classroom *j*
- β_{0j} is the average outcome of students in classroom *j*
- r_{ij} is a random error associated with student *i* in classroom *j*; $r_{ij} \sim N(0, \sigma^2)$.

At the classroom level:

$$\beta_{0j} = \delta_k \sum_{k=1}^{26} School_k + \lambda_k \sum_{k=1}^{26} School_k * CSR_j + u_{0j}$$

where

- *CSR* is the indicator variable for intervention condition (1 = CSR classroom, 0 = control classroom)
- $k = 1, 2, \ldots, 26$, are dummy indicator variables representing the 26 schools in the sample
- $School_k*CSR_j$ are 26 intervention-by-school interaction terms
- δ_k, $k = 1, 2, \ldots, 26$, represents the 26 school intercepts
- λ_k, $k = 1, 2, \ldots, 26$, represents the effect of CSR on each of the 26 schools
- u_{0j} is the random error associated with classroom *j* on classroom average outcome; $u_{0j} \sim N(0, \tau_{00})$.

The reduced form equation is

$$Y_{ij} = \delta_k \sum_{k=1}^{26} (School)_k + \lambda_k \sum_{k=1}^{26} (School)_k * (CSR)_j + r_{ij} + u_{0j} .$$

Of primary interest is the combined effect of λ_k coefficients for the *School*CSRI* indicator interaction terms, in which school-specific estimates were weighted by the proportion of students in the sample in each school. The combined effect represents the intervention's main effect on the outcome across all schools. A statistically significant positive value of the combined effect would indicate that students in CSR classrooms demonstrate higher levels of reading comprehension than their counterparts in control classrooms.

Model 8: Heterogeneity of treatment slopes regarding pretest covariate (sensitivity analyses 7 and 8)

Whether the effect of CSR varied depending on the level of the pretest GRADE scores was tested through a model that specifically estimated the significance of the *CSR*Pretest* interaction. This model included the same student- and teacher-level covariates as the confirmatory impact model (the GRADE pretest score, students' language status, indicator variables for teachers' Spanish fluency). The analysis was conducted with all 26 schools and

with a sample that included the 23 schools that could be included in the FC–ELL/non–ELL subgroup analyses.[45]

At the student level:

$$Y_{ij} = \beta_{0j} + \beta_{1j}*(Pretest)_{ij} + \beta_{2j}*(Language\ Status)_{ij} + r_{ij}.$$

The level 2 model not only models the level 1 intercept as the outcome, it also allows the effect of the treatment to differ by the level of the pretest GRADE score:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(CSR)_j + \gamma_{02}(Spanish_1)_j + \gamma_{03}(Spanish_2)_j + \gamma_{04}(Spanish_3)_j + \delta_k \sum_{k=2}^{26} School_k + \lambda_k \sum_{k=2}^{26} School_k * CSR_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}*(CSR)_j + \gamma_{12}(Spanish_1)_j + \gamma_{13}(Spanish_2)_j + \gamma_{14}(Spanish_3)_j + \theta_k \sum_{k=2}^{26} School_k + \vartheta_k \sum_{k=2}^{26} School_k * CSR_j$$

$$\beta_{2j} = \gamma_{20}$$

where γ_{10} is the influence of the pretest covariate across all classrooms (that is, the grand pretest slope), and γ_{11} is the differential treatment effect based on the GRADE pretest covariate.

This model was estimated by using effect-coded school indicator variables to directly estimate the *CSR*Pretest* interaction term. The parameter estimate of interest in determining whether the treatment effect varies by the pretest covariate is γ_{11}, which represents the intervention's main effect on the pretest slope. A statistically significant γ_{11}would indicate that CSR has a significant differentiated effect, suggesting that students with different levels of the pretest GRADE covariate scores benefit differently from CSR.

Model 9: Cohort effect (sensitivity analysis 9)

Six schools participated in the study during year 1; another 20 schools participated during year 2. Sensitivity analysis was conducted to see whether the effect of CSR differed for schools that participated in year 1 and year 2 (differential treatment effect).

At the student level:

$$Y_{ij} = \beta_{0j} + \beta_{1j}*(Pretest)_{ij} + \beta_{2j}*(Language\ Status)_{ij} + r_{ij.}$$

As level 1 is identical to the confirmatory impact model, the focus here is on level 2, the classroom level:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(CSR)_j + \gamma_{02}*(Year)_j + \gamma_{03}*(Year)_j*(CSR)_j + \gamma_{04}*(Spanish_1)_j + \gamma_{05}*(Spanish_2)_j + \gamma_{06}*(Spanish_3)_j + \sum_{k=3}^{26} \delta_k * (School_k)_j + \sum_{k=3}^{26} \lambda_k * (School_k)_j * CSR_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

[45] Three schools were missing FC–ELL or non–ELL students in both control and treatment conditions.

where

- *CSR* is an indicator variable for the intervention condition (1 = CSR teachers, 0 = control teachers)
- $Year_j$ is an indicator variable for the data collection cohort (1 = first year, 0 = second year) in classroom *j*
- $Spanish_1$, $Spanish_2$, and $Spanish_3$ are indicator variables for teachers' Spanish fluency, grand–mean centered
- $School_k$, k = 3, …, 26, are 24 effect coded dummy indicator variables representing the 26 schools in the sample, with $school_1$ and $school_2$ the omitted reference categories for year 1 and year 2; school indicator variables are effect coded
- $(School_k)_j*CSR_j$, k = 3, …, 26, are 24 effect coded *School*CSR* interactions in classroom *j*, with $school_1$ and $school_2$ the omitted reference categories for year 1 and year 2; school indicator variables are effect coded
- $(Year)_j*(CSR)_j$ are *Data Collection Year/Cohort*CSR* interaction terms in classroom *j*
- γ_{00} is the average student outcome across all teachers, adjusted for covariates
- γ_{01} is the effect of CSR on average student outcomes
- γ_{02} is the relationship between *Data Collection Year/Cohort 1* and average student outcomes, holding other level 2 covariates constant
- γ_{03} is the relationship between *CSR* in *Data Collection Year/Cohort 1* and average student outcomes, holding other level 2 covariates constant
- γ_{04} is the relationship between teachers speaking a little Spanish and average student outcomes, holding other level 2 covariates constant
- γ_{05} is the relationship between teachers speaking Spanish as a second language and average student outcomes, holding other level 2 covariates constant
- γ_{06} is the relationship between teachers speaking Spanish as a first language and average student outcomes, holding other level 2 covariates constant
- δ_k, k = 3, …, 25, represent the 24 fixed school effects
- λ_k, k = 3,…25, represent the 24 *School*CSR* interaction terms
- γ_{10} and γ_{20} are the fixed effects representing the average slopes across all teachers
- u_{0j} is a random error associated with teacher *j* on teacher average outcome; $u_{0j} \sim N(0, \tau_{00})$.

The model provided three estimates of interest: γ_{03} provided an estimate of the differentiated CSR effect between study years; γ_{01} provided an estimate of the effect of CSR for schools participating in year/cohort 2 of the study; and the combined effect of γ_{01} and γ_{03} provided an estimate of the effect of CSR for schools participating in year/cohort 1 of the study.

Model specifications for exploratory analyses

A hierarchical linear model (HLM) that allowed the treatment effect to vary by students' language status was used to test the effectiveness of CSR for FC–ELL and ELL subgroups and to determine whether CSR had a differential effect on FC–ELL and non–ELL students. This analysis included the 23 schools with FC–ELL and non–ELL students in both CSR and control classrooms. The model included the same variables as the main impact model (pretest GRADE score, students' language status, indicator variables for teachers' Spanish fluency).

At the student level:

$$Y_{ij} = \beta_{0j} + \beta_{1j}*(Pretest)_{ij} + \beta_{2j}*(Language\ Status)_{ij} + r_{ij}.$$

Because level 1 is identical to the confirmatory impact model, the focus here is on the terms in level 2, which models the level 1 intercept as the outcome and allows the effect of CSR to vary by students' language status:

$$\beta_{0j} = \gamma_{01}(Spanish_1)_j + \gamma_{02}(Spanish_2)_j + \gamma_{03}(Spanish_3)_j + \delta_k \sum_{k=1}^{23} School_k + \lambda_k \sum_{k=1}^{23} School_k * CSR_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{21}(Spanish_1)_j + \gamma_{22}(Spanish_2)_j + \gamma_{23}(Spanish_3)_j + \theta_k \sum_{1}^{23} School_k + \pi_k \sum_{k=1}^{23} School_k * CSR_j + u_{2j}$$

where

- *CSR* is an indicator variable for the intervention condition (1 = CSR teachers, 0 = control teachers)
- $Spanish_1$, $Spanish_2$, and $Spanish_3$ are indicator variables for teachers' Spanish fluency proficiency, grand–mean centered
- $School_k$, $k = 1, 2, \ldots, 23$ are 23 dummy indicator variables representing the 23 schools in the sample
- $School_k*CSR_j$ are intervention-by-school interaction terms
- γ_{01} is the association between teachers speaking a little Spanish and average student outcomes, holding other level 2 covariates constant
- γ_{02} is the association between teachers speaking Spanish as a second language and average student outcomes, holding other level 2 covariates constant
- γ_{03} is the association between teachers speaking Spanish as a first language and average student outcomes, holding other level 2 covariates constant
- δ_k, $k = 1, 2, \ldots, 23$, represents the 23 school intercepts for the non–ELL students
- λ_k, $k = 1, 2, \ldots, 23$, represents the CSR effect on the non–ELL students in each of the 23 schools

- γ_{10} is the fixed effects representing the average pretest score across all classrooms
- γ_{20} is the fixed effects representing the average language status across all classrooms
- u_{0j} is a random error associated with classroom *j* on classroom average outcome; $u_{0j} \sim \text{N}(0, \tau_{00})$
- γ_{21} is the association between teachers speaking a little Spanish and the average achievement gap between FC–ELL and non–ELL students, holding other level 2 covariates constant
- γ_{22} is the association between teachers speaking Spanish as a second language and the average achievement gap, holding other level 2 covariates constant
- γ_{23} is the association between teachers speaking Spanish as a first language and the average achievement gap, holding other level 2 covariates constant
- θ_k is the difference in the FC–ELL/non–ELL student achievement gap for each of the 23 schools
- π_k is the difference in the FC–ELL/non–ELL student achievement gap between CSR and control classrooms in each of the 23 schools
- u_{2j} is a random error associated with classroom *j* on the achievement gap between FC–ELL and non–ELL students,

$$\begin{pmatrix} u_{0j} \\ u_{2j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{02} \\ \tau_{20} & \tau_{22} \end{pmatrix}\right].$$

The model provided three combined estimates of interest:

- The combined effect of the λ_k coefficients for the *School*CSR* indicator interaction terms capture the intervention's main effect on the outcome for the non–ELL students across all schools. A statistically significant positive value of the combined effect would indicate that the non–ELL students in the CSR classrooms demonstrate higher levels of reading comprehension than their counterparts in control classrooms.
- The combined effect of π_k coefficients for the *School*CSR* indicator by language status interaction terms represents the differentiated effect between FC–ELL and non–ELL students. A statistically significant positive value of the combined effect would indicate that FC–ELL students benefit more than non–ELL students from CSR.
- The combined effect of λ_k and π_k coefficients represents the main effect of CSR on FC–ELL students. A statistically significant positive value of the combined effect would indicate that the FC–ELL students in the CSR classrooms demonstrate higher levels of reading comprehension than their counterparts in control classrooms.

Effect size calculations from multiply imputed datasets

Effects sizes were calculated using Hedges' *g*, the mean difference between groups divided by the multiply imputed student-level pooled standard deviation of posttest scores.

In general, $G = \frac{\lambda}{\sqrt{\frac{(n_1 - 1)S_1^{\,2} + (n_2 - 1)S_2^{\,2}}{n_1 + n_2 - 2}}}$

Normally, $S_1^{\,2}$ and $S_2^{\,2}$ are obtained by simply calculating the raw student-level variance for the CSR and control groups. However, because multiple imputation was used, it was necessary to obtain the multiply imputed student-level variances.

To calculate these estimates, the multiply imputed standard error was first calculated, using the formula $SE_{MI} = \sqrt{\left(\frac{M+1}{M}\right)V_B + V_W}$

Where M = number of imputed datasets; V_B = between-imputation variance; and V_W = within-imputation variance.

The desired student level variances were calculated next (accounting for multiple imputation), $S_1^{\,2} = n_1(SE_1)^2$ and $S_2^{\,2} = n_2(SE_2)^2$. These variances were then entered into the first equation above to obtain the correct effect size.

Appendix F. Frequently asked questions about contamination

The following answers to frequently asked questions (FAQs) were provided to participating teachers in both Collaborative Strategic Reading (CSR) and control classrooms.

What is contamination and why should I care? Contamination occurs when important aspects of CSR being taught in this training make their way into the control classrooms. If control teachers start to use CSR then we won't be able to clearly determine whether or not CSR works.

What are the chances contamination can occur in the CSR study? Since each school will have both "treatment" (i.e., CSR) teachers and control teachers, contamination is certainly a risk. However, by being aware of the potential for contamination, and keeping alert, it is fairly easy for teachers to help minimize this risk.

What, specifically, might contamination look like in this study? What should I watch for and what is okay? Contamination will occur in this study if specific details about how to use CSR are shared with control teachers, either through conversations or sharing of material. Effective use of CSR requires training and coaching, and therefore casual discussion about CSR in staff meetings and the like are not a concern. However, treatment teachers should not discuss details of the method with control teachers or school staff. Also, no materials from the professional development (e.g., handouts, the CSR book) should be shared with the control teachers. It is important that these materials not be casually left in common meeting places.

What should I do if I think that contamination is happening? If you find that a control teacher is looking at materials or if you learn of a case where a CSR treatment teacher gives specific directions to a control teacher on how to use the techniques, you should inform <insert project manager contact information> as soon as possible. That way we can take appropriate steps to minimize any contamination and deal with the issue.

What is the study team doing to prevent contamination and deal with it if it occurs? Our primary goal is to make all of the administrators, CSR treatment teachers, and control teachers understand how important it is to avoid contamination so that it doesn't occur in the first place. Teachers are devoting time and effort to use CSR, and we all want to make sure that the study can scientifically determine whether CSR is truly going to be beneficial to the students. The most important step we can take is to make sure that everyone understands how vital it is to avoid contamination.

Should contamination occur, we will first take steps to stop it from continuing. We will also evaluate how serious the contamination is and determine what we can do statistically or from a research point of view to deal with it so that study results will still be valid. Again, the faster we become aware of any contamination issue, the easier it will be to minimize its effects. You are our first line of defense in not letting contamination happen and making us aware of it if it does happen.

So what is the bottom line? Treatment teachers and administrators who have access to training should not share CSR documents with control teachers. Likewise, control teachers should not ask for these materials and should not try to use them when teaching.

If you have any questions about this document now or at any point in the future, please call:
<Name>, CSR Project Manager <contact information>

Appendix G. Attrition analyses

Attrition analyses explored whether study condition (Collaborative Strategic Reading [CSR] classroom or control classroom), pretest Group Reading Assessment and Diagnostic Evaluation (GRADE) scores, or demographic information (race/ethnicity, gender, language status, free or reduced-price lunch status, special education status) were related to students leaving study schools after pretesting. Attrition analyses were conducted using a sample of students eligible at baseline for pretesting (n = 1,337). For binary outcomes such as sample membership at posttest, a hierarchical generalized linear model can assume a Bernoulli sampling model and a logit link function.

The following two-level hierarchical general linear model examined whether study condition predicted sample membership at posttest for students present at baseline. Given a Bernoulli sampling model and a logit link function, model A1 was specified at the student level:

$$\text{Prob(present at posttest)} = 1|\beta = \varphi$$

$$\text{Log}[\varphi/(1-\varphi)] = \eta_{ij}$$

$$\eta_{ij} = \pi_{0j} + e$$

where

- η_{ij} = log (φ /(1 – φ)) (the log of the odds of not being in the posttest sample)
- π_{0j} is the average log odds of a student attriting in classroom j
- e is an error term, $e \sim 1/(\varphi(1-\varphi))$, where the probability of (Y = 1| β) = φ.

The classroom average outcome estimated from this model (in effect, the level 1 intercept π_{0j}) was modeled as varying randomly across classrooms:

$$\pi_{0j} = \beta_{00} + \beta_{01}*(CSR)_j + r_{0j}$$

where

- β_{00} is the average log odds of a student attriting across all classrooms
- *CSR* is an indicator variable for the intervention in classroom j (CSR = 1, control = 0)
- β_{01} is the difference in log odds of a student attriting in a CSR classroom and a control classroom (in effect, intervention effect)
- r_{0j} is a random error associated with classroom j on classroom average log odds of a student attriting, $r_{0j} \sim N(0, \tau_{00})$.

The key parameter of interest is β_{01}, the difference in the log odds for students in CSR and control classrooms to remain in the posttest sample. Two additional models were estimated to examine the robustness of this parameter. Model A2 included additional student-level demographic variables (race/ethnicity, gender, language status, free or reduced-price lunch status, special education status). Model A3, conducted as a sensitivity analysis, included the *CSR* indicator and the demographic variables and their interactions (table G-1). Additional sensitivity analyses were conducted by analyzing the same model using logistic regressions.

Table G-1. Results of attrition analysis using two-level hierarchical generalized linear model

Analysis	*Coefficient*	p-*value*
Model A1: Two-level HLM		
Study condition	0.05 (0.21)	0.79
Model A2: Two-level HLM with student covariates		
Study condition	0.10 (0.24)	0.68
Pretest GRADE score	–0.02 (0.01)	0.03
Special education status	1.29 (0.28)	< 0.01
Hispanic	–0.39 (0.28)	0.17
Free or reduced-price lunch status	–0.36 (0.26)	0.18
Gender	0.12 (0.23)	0.61
Language status	0.51 (0.29)	0.08
Model A3: Two-level HLM with student covariates and interaction terms		
Study condition	–4.93 (2.10)	0.02
Pretest GRADE score	–0.04 (0.02)	0.01
Special education status	1.12 (0.41)	0.01
Hispanic	–0.59 (0.40)	0.14
Free or reduced-price lunch status	–1.16 (0.36)	< 0.01
Gender	0.08 (0.34)	0.81
Language status	0.25 (0.41)	0.55
Pretest GRADE score*Study condition	0.04 (0.02)	0.09
Special education status*Study condition	0.32 (0.57)	0.57
Hispanic*Study condition	0.37 (0.56)	0.51
Free or reduced-price lunch status*Study condition	1.69 (0.55)	< 0.01
Gender*Study condition	0.12 (0.47)	0.80
Language status*Study condition	0.57 (0.58)	0.33

Note: Figures in parentheses are standard errors.

Source: Authors' analysis of student GRADE scores and study records.

The attrition analysis including only the *CSR* indicator variable (model A1) showed that the study condition was not significantly related to student attrition (estimate = 0.05, standard error = 0.21, $p = 0.79$). To explore whether student demographics were associated with student attrition, another model (model A2) was examined that included GRADE scores, special education status, language status, gender, free and reduced-price lunch status, and whether a student was Hispanic.[46] The intervention indicator (that is, study condition) was not statistically significant in this model (estimate = 0.10, standard error = 0.24 $p = 0.68$). In contrast, special education status was positively associated with attrition (estimate = 1.28, standard error = 0.28, $p < 0.01$), and the GRADE scores were negatively associated with attrition (estimate = -0.02, standard error = 0.01, $p = 0.03$), implying that students with special education status and lower GRADE scores attrited more often.

Results from models A1 and A2 consistently showed that the study condition was not related to attrition. Model A3, which included *CSR*demographic* interactions, shows that special education baseline students (estimate = 1.12, standard error = 0.41, $p = 0.01$) attrited more often than non–special education students. Students in CSR classrooms (estimate = –4.93, standard error = 2.10, $p = 0.02$); students with free or reduced-price lunch status (estimate = –1.16, standard error = 0.36, $p < 0.01$); and students with higher pretest GRADE scores (estimate = –0.02, standard error = 0.01, $p = 0.02$) were less likely to attrite than their counterparts in control classrooms. In contrast, students receiving a free or reduced-price lunch were more likely to attrite in the CSR condition than in the control condition (estimate = 1.69, standard error = 0.55, $p < 0.01$). These analyses suggest that there is nontrivial attrition and that the assumption that the missing data are missing completely at random (MCAR) appears to be too strong. Consequently, multiple imputation was selected as the missing data approach (missing imputation procedures are described in appendix K).

[46] As the majority of students in the sample were Hispanic, Hispanic versus non-Hispanic made the most sense for examining race/ethnicity.

Appendix H. Response rates for demographic data

The response rates for teacher and student data collection are presented in tables H-1 and H-2. More than 95 percent of testable students took the Group Reading Assessment and Diagnostic Evaluation (GRADE) at both pretest and posttest.

Table H-1. Response rates (percent) for teacher data collection, by school term and data collection instrument

Instrument/characteristic	*Fall 2007*	*Spring 2008*	*Fall 2008*	*Spring 2009*
Teacher survey				
	100	94.7	100	100
Collaborative Strategic Reading Intervention Validity Checklist (CSRIVC)				
	na	100	na	100
Expository Reading Comprehension (ERC) observation instrument				
	na	100	na	100

na is not applicable.
Source: Authors' analysis of teacher surveys and data collected from the CSRIVC and ERC.

Table H-2. Pretest and posttest response rates for student data collection

	Pretest GRADE			*Posttest GRADE*			*Pretest and posttest GRADE*		
Item	*CSR*	*Control*	*Total*	*CSR*	*Control*	*Total*	*CSR*	*Control*	*Total*
Number of consenting students	702	649	1,351	645	631	1,276	643	597	1,240
Number of testable students	692	645	1,337	633	622	1,255	634	596	1,230
Number of tested students	659	621	1,280	606	597	1,203	584	544	1,128
Response rates (percent)									
Consenting students	93.9	95.7	94.7	94.0	94.6	94.3	90.8	91.1	91.0
Testable students	95.2	96.3	95.7	95.7	96.0	95.9	92.1	91.3	91.7

Source: Authors' analysis of student Group Reading Assessment and Diagnostic Evaluation (GRADE) scores and study records.

Appendix I. Fall and spring teacher surveys

The following surveys are provided in their entirety:

- Collaborative Strategic Reading for Fifth Graders, Fall 2007 Teacher Survey
- Collaborative Strategic Reading for Fifth Graders, Spring 2008 Teacher Survey, Version A
- Collaborative Strategic Reading for Fifth Graders, Spring 2008 Teacher Survey, Version B

COLLABORATIVE STRATEGIC READING FOR FIFTH GRADERS

FALL 2007

TEACHER SURVEY

Paperwork Burden Statement

Dear Teacher:

The Collaborative Reading Strategies Study (CSR Study) is a groundbreaking study designed to test an innovative method for teaching reading comprehension in the fifth grade. Your participation is important and appreciated, but you do have the right to skip any question that you do not wish to answer. Below are answers to some general questions concerning this survey.

What is the purpose of this survey?

The purpose of this survey is to collect background information, such as years of teaching experience, about the teachers participating in the study.

Who is conducting this survey?

The CSR Study was commissioned by the Department of Education's Institute of Education Sciences, and is administered by Regional Educational Laboratory Southwest, American Institutes for Research, Caliber Associates, and RG Research Group.

Why should you participate in this survey?

Policymakers and educational leaders rely on findings from studies like the CSR Study to inform their decisions. The current study will fill a critical gap in the reading research literature as to what is effective for improving reading achievement, especially for children who are English language learners. Your participation in the study will help us to find out whether CSR is an effective solution.

Will your responses be kept confidential?

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 or individual. We will not provide information that identifies you or your district to anyone outside the study team, except as required by law.

How will your information be reported?

The information you provide will be combined with the information provided by other teachers in statistical reports. No information that links your name, address, or telephone number with your responses will be included in any reports related to the study.

Where should you return your completed survey?

Please return the completed survey to the person who gave you the survey.

Who can you contact about the survey?

If you have any questions about the survey, you can ask the person who gave you the survey, or you can contact the coordinator of data collection, <Name and contact information of data collection coordinator>.

Thank you for your cooperation in this very important effort!

Background Information

Education and Preservice Training

1. Have you earned any of the following degrees, certificates or credentials? (Check no or yes in each row, and write in the major code from Table 1 and year if applicable.)

Degree	Earned	Major code/s (from table 1)	Year/s
a. Bachelor's degree	1☐ No 2☐ Yes➔		
b. Master's degree	1☐ No 2☐ Yes➔		
c. Educational specialist or professional diploma (at least one year beyond master's level)	1☐ No 2☐ Yes➔		
d. Certificate of advanced graduate studies	1☐ No 2☐ Yes➔		
e. Doctorate or professional degree (Ph.D., Ed.D., M.D, L.L.B, J.D, D.D.S)	1☐ No 2☐ Yes➔		

Table 1. Major field of study codes

Major code	Major field
01	Elementary education
02	Secondary education
03	Special education
04	Arts/music
05	English/language arts
06	English as a second language

07	Foreign languages
08	Mathematics
09	Computer science
10	Natural sciences
11	Social sciences
12	Education administration
13	Reading specialist
14	Other

Certification

2. Which of the following describes the teaching certificate you currently hold in this state?

- ☐ Regular or standard state certificate or advanced professional certificate
- ☐ Probationary certificate (issued after satisfying all requirements except the completion of a probationary period)
- ☐ Provisional or other type of certificate given to persons who are still participating in what the state calls an "alternative certification program"
- ☐ Temporary certificate (requires some additional college coursework, student teaching, and/or passage of a test before regular certification can be obtained)
- ☐ Waiver or emergency certificate (issued to persons with insufficient teacher preparation who must complete a regular certification program in order to continue teaching)
- ☐ I do not have any of the above certifications in this state.

Full-Time/Part-Time

3. How do you classify your position at THIS school, that is, the activity at which you spend most of your time during this school year? *Mark (X) only one box.*

- ☐ Regular full-time teacher
- ☐ Regular part-time teacher
- ☐ Itinerant teacher (i.e., you provide instruction at more than one school)
- ☐ Long-term substitute (i.e., you fill the role of a regular teacher on a long-term basis, but you are still considered a substitute)

Experience

4. How many years of teaching experience do you have: (Write in number of years. Count the current year as one full year.)

	Number of years
a. Teaching in total	☐ years
b. Teaching fifth grade	☐ years
c. Teaching social studies	☐ years
d. Teaching at this school	☐ years

5. What grade/s have you taught in the past? Please circle all that apply.

K	1	2	3	4	5	6	7	8
___	Secondary/high school							
___	None: This is my first year teaching.							

6. **What grade did you teach last year?**

K	1	2	3	4	5	6	7	8
___	I did not teach elementary school last year.							

Preparation Time

7. How many hours per week do you have designated as paid preparation periods?

☐

Demographics

8. What is your gender?

- ☐ Male
- ☐ Female

9. Are you of Hispanic origin?

- ☐ Yes
- ☐ No

10. What is your race? Mark (X) one or more races to indicate what you consider yourself to be.

- ☐ White
- ☐ Black or African American
- ☐ Asian
- ☐ Native Hawaiian or Other Pacific Islander
- ☐ American Indian or Alaska Native

11. Which of the following describes your oral Spanish language fluency?

- ☐ I do not speak any Spanish
- ☐ I speak a little Spanish
- ☐ Fluent: Spanish is my second language
- ☐ Fluent: Spanish is my home or first language

Professional Development Experiences

Types of Professional Development

In answering the following items, consider all the professional development activities related to *reading instruction* in which you have participated during the 2006–2007 school year and the summer 2007. *Professional development* refers to a variety of activities intended to enhance your professional knowledge and skills, including teacher networks, coursework, institutes, workshops, committee work, coaching, and mentoring. Workshops are short-term learning

opportunities that can be located in your school or elsewhere. Institutes are longer-term professional learning opportunities, for example, of a week or longer in duration.

12. Did you participate in any professional development related to *reading instruction* during the summer of 2007 (including Collaborative Strategic Reading training)?

☐ Yes

☐ No ➔ *Skip to question 14*

13. During the summer of 2007, what is the total number of hours you spent in the following professional development activities?

Write the total number of <u>hours</u> you spent in these activities. Mark "0" if you participated in none.

	Summer of 2007
	Number of hours
a. Attended short, stand-alone training or workshop in reading (half-day or less).	
b. Attended longer institute or workshop in reading (more than half-day).	
c. Attended a college course in reading (include any courses you are currently attending).	
d. Received coaching or mentoring related to reading instruction.	
e. Acted as a coach or mentor related to reading instruction.	
f. Other informal professional development (e.g., participate in teacher study group, network or collaboration supporting professional development in reading, participated in committee or task force related to reading, visited or observed reading instruction in other schools).	

14. Did you participate in any professional development related to *reading instruction* during the 2006–2007 school year?

☐ Yes

☐ No ➔ *Skip to question 16*

15. During the 2006–2007 school year, what is the total number of hours you spent in the following professional development activities?

Write the total number of <u>hours</u> you spent in these activities. Mark "0" if you participated in none.

	2006–2007 school year
	Number of hours
a. Attended short, stand-alone training or workshop in reading (half-day or less).	
b. Attended longer institute or workshop in reading (more than half-day).	
c. Attended a college course in reading (include any courses you are currently attending).	
d. Received coaching or mentoring related to reading instruction.	
e. Acted as a coach or mentor related to reading instruction.	
f. Other informal professional development (e.g., participate in teacher study group, network or collaboration supporting professional development in reading, participated in committee or task force related to reading, visited or observed reading instruction in other schools).	

16. During the 2006–2007 school year and summer of 2007, how much did the professional development in which you participated emphasize the following reading/language arts/English topics? ***(Circle one number in each row.)***

	Topic in reading/language arts/English	Not an emphasis	Minor emphasis	Moderate emphasis	Major emphasis
a.	Alphabetic code (e.g., letter sounds, spelling patterns)	1	2	3	4
b.	Vocabulary (e.g., definitions, synonyms, suffixes, etc.)	1	2	3	4
c.	Fluent reading of text (e.g., awareness of text, pace, accuracy, etc.)	1	2	3	4
d.	Comprehension of text (e.g., text elements, strategies, main idea, etc.)	1	2	3	4
e.	How to use a reading program or curriculum	1	2	3	4
f.	How to organize small group instruction	1	2	3	4
g.	How to diagnose reading problems	1	2	3	4
h.	Analyzing and interpreting student achievement data	1	2	3	4
i.	How to interpret and use assessment data to guide instruction	1	2	3	4
j.	How to teach reading to ELL students	1	2	3	4

You are done with the survey. Thank you.

COLLABORATIVE STRATEGIC READING FOR FIFTH GRADERS

SPRING 2008

TEACHER SURVEY, VERSION A

Paperwork Burden Statement

Dear Teacher:

The Collaborative Reading Strategies Study (CSR Study) is a groundbreaking study designed to test an innovative method for teaching reading comprehension in the fifth grade. Your participation is important and appreciated, but you do have the right to skip any question that you do not wish to answer. Below are answers to some general questions concerning this survey.

What is the purpose of this survey?

The purpose of this survey is to collect information about the classrooms of the teachers participating in the study.

Who is conducting this survey?

The CSR Study was commissioned by the Department of Education's Institute of Education Sciences, and is administered by Regional Educational Laboratory Southwest, American Institutes for Research, Caliber Associates and RG Research Group.

Why should you participate in this survey?

Policymakers and educational leaders rely on findings from studies like the CSR Study to inform their decisions. The current study will fill a critical gap in the reading research literature as to what is effective for improving reading achievement, especially for children who are English language learners. Your participation in the study will help us to find out whether CSR is an effective solution.

Will your responses be kept confidential?

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 or individual. We will not provide information that identifies you or your district to anyone outside the study team, except as required by law.

How will your information be reported?

The information you provide will be combined with the information provided by other teachers in statistical reports. No information that links your name, address, or telephone number with your responses will be included in any reports related to the study.

Where should you return your completed survey?

Please return the completed survey to the person who gave you the survey.

Who can you contact about the survey?

If you have any questions about the survey, you can ask them from the person who gave you the survey, or you can contact the coordinator of data collection <Name and contact information of data collection coordinator>.

Thank you for your cooperation in this very important effort!

Classroom Context

Class Structure

1. Which category best describes the class(es) for which you provide reading and social studies instruction?

- ☐ I have a self-contained classroom. I am the main reading and social studies teacher for students enrolled in the classroom. (A few students may get extra help from a reading specialist or other teacher.)
- ☐ I team teach with another teacher. We share joint responsibility for reading and social studies instruction for *all* our students.
- ☐ I team teach with another teacher. I have responsibility for reading and social studies instruction for *some* of the students, and my team teacher has responsibility for reading and social studies instruction for *other* students. (That is, we are regularly responsible for *different groups* of students within the class.)
- ☐ I do not teach reading.
- ☐ I do not teach social studies.
- ☐ Other. *Please specify*.

2. A student teacher is defined as someone who is not yet licensed or certified to teach, who is completing a degree program that will lead to a degree in teaching and licensure or certification, and who will take over the teaching of your class for a period of time. Please check the statement/s that best describe your experience with student teachers this school year and fill in the number of weeks a student teacher taught in your classroom.

- ☐ I had a student teacher in my classroom in the fall. S/he took over all teaching for ____ weeks.
- ☐ I have or will have a student teacher in my classroom this spring. S/he will take over all teaching for ____ weeks.
- ☐ I did not have a student teacher in my classroom during the fall and I will not have a student teacher in my classroom this school year.

3. Think about the students for whom you currently provide reading instruction. Do you expect to provide reading instruction for most of these students for the full school year, or does your school periodically re-assign students to classes for reading over the school year?

- ☐ I expect to continue to provide reading instruction for most of the students in my current class all year.
- ☐ My school periodically re-assigns students in my current class for reading.

4. In answering 4a–4g, include ALL of the students to whom you teach reading, whether you teach reading on your own in a self-contained classroom, to a group that includes students from other classes, or to more than one group of students.

	Enter # below
a. What is the total number of students to whom you currently teach reading?	
b. How many of your reading students receive intervention services in reading from you or another teacher or tutor? **Reading Intervention** is a program designed for **struggling readers** to be used only with struggling readers in addition to the core reading program.	
c. How many of your students are reading at or above grade level?	
d. How many of your students are reading one year below grade level?	
e. How many of your students are reading two or more years below grade level?	
f. How many of your students are English Language Learners (ELL), also referred to as Limited English Proficient (LEP), English for Speakers of Other Languages (ESOL), or English as a Second Language (ESL) students? (e.g., Spanish, Russian, Chinese)?	
g. How many of your students are special education students with Individualized Education Plans (IEP's) who receive special education services in reading?	

Class Behavior

5. At this point in the school year, how would you rate the behavior in your class? *Please mark (x) one box.*

- ☐ Group misbehaves very frequently and is almost always difficult to handle
- ☐ Group misbehaves frequently and is often difficult to handle
- ☐ Group misbehaves occasionally
- ☐ Group behaves well
- ☐ Group behaves exceptionally well

Instructional Resources

6. In a typical week, do you have paid and/or volunteer aides or specialists assist in your class in the following ways? *Please indicate separately whether paid and/or volunteer aides assist you in your classroom. If you do not have aides or specialists in a particular category, you only need to check the relevant box ☐ in the right column.*

	Paid aide/specialist	**Volunteer** aide/specialist	Do ***not*** have this type of aide or specialist assistance
Regular Aides			
Working directly with children on instructional tasks	☐	☐	☐
Doing non-instructional work (e.g., photocopying, preparing materials, etc.)	☐	☐	☐
Pulling children out of class to work on instructional tasks	☐	☐	☐
Special Education Aides or Specialists			
Working directly with children on instructional tasks	☐	☐	☐
Doing non-instructional work (e.g., photocopying, preparing materials, etc.)	☐	☐	☐
Pulling children out of class to work on instructional tasks	☐	☐	☐
ESL or Bilingual Education Aides or Specialists			
Working directly with children on instructional tasks	☐	☐	☐
Doing non-instructional work (e.g., photocopying, preparing materials, etc.)	☐	☐	☐
Pulling children out of class to work on instructional tasks	☐	☐	☐
Reading Specialists Who Work Primarily with Students			
Working directly with children on instructional tasks	☐	☐	☐
Doing non-instructional work (e.g., photocopying, preparing materials, etc.)	☐	☐	☐
Pulling children out of class to work on instructional tasks	☐	☐	☐

Reading Program

7. What reading program (such as Open Court Reading) or curriculum do you use in your classroom? Please check all that apply.

- ☐ Basal reader, such as *Open Court* or *Houghton Mifflin*
- ☐ School-wide literacy model, such as Success for All or Balanced Literacy
- ☐ Teacher (i.e., self) developed reading curriculum
- ☐ District or school developed literature-based (i.e., trade books) curriculum
- ☐ Other (please specify) _____

8. What kind of professional development have you received regarding the reading program or curriculum that you are currently using?

- ☐ None
- ☐ Start-up training only
- ☐ Ongoing support only (e.g., mentoring, coaching, consultation)
- ☐ Start-up training plus ongoing support

9. What social studies program or curriculum do you use in your classroom? Please check all that apply.

- ☐ Social studies textbook

Title of the text book:_____

If a social studies textbook is used, is it supplemented with trade books?
- ☐ Yes
- ☐ No

- ☐ Teacher (i.e., self) developed social studies curriculum
- ☐ District developed social studies curriculum
- ☐ Social studies curriculum consisting primarily of trade books
- ☐ Other (please specify) _____

10. How many weeks are spent on student/class projects?

_____ weeks

Use of Collaborative Strategic Reading (CSR) Program

11. How many times per week do you use Collaborative Strategic Reading (CSR) approach and instructional practices that you learned about in the beginning of 2007 school year?

☐ I did not use CSR approach in my classroom
☐ 1 time a week
☐ 2 times a week
☐ 3 times a week
☐ 4 times a week
☐ 5 times a week
☐ 6+ times a week

12. Did you meet with your CSR coach during the fall of 2007?

☐ Yes

☐ No. Please skip to question 14.

13. How many times and approximately for how long did you meet with the CSR coach?

_____times for approximately _____minutes.

14. In your opinion, is CSR having a positive effect on the reading skill growth, particularly reading comprehension skills, of your students?

☐ Yes

☐ No

Collaboration among Teachers

We would like to learn about teachers' experiences collaborating with other teachers in their schools. Please think about both formal activities at your school intended to encourage collaboration and informal conversations you have with other teachers.

15. **Not including the current school year and not including student teaching, how many years have you been a teacher? If this is your first year teaching, answer "zero."**

    _____ years

16. **Not including the current school year and not including student teaching, how many years have you taught in your current school? If this is your first year in this school, answer "zero."**

    _____ years

17. **Some teachers work independently while other teachers prefer to get input from other teachers. Would you say you get…**

    ☐ No input

    ☐ Minimal input

    ☐ Moderate input

    ☐ A great deal of input

18. **How comfortable are you receiving advice from other teachers?**

    ☐ Not at all comfortable

    ☐ Slightly comfortable

    ☐ Moderately comfortable

    ☐ Completely comfortable

19. How comfortable are you offering advice to other teachers?

- ☐ Not at all comfortable
- ☐ Slightly comfortable
- ☐ Moderately comfortable
- ☐ Completely comfortable

20. How supportive are other teachers at your school when you need help or advice with teaching?

- ☐ Not at all comfortable
- ☐ Slightly comfortable
- ☐ Moderately comfortable
- ☐ Completely comfortable

21. How receptive are other teachers at your school when you need help or advice with teaching?

- ☐ Virtually no teachers are receptive
- ☐ Some teachers are receptive, but a majority are not
- ☐ A majority of teachers are receptive, but some are not
- ☐ Nearly every teacher is receptive

22. In general, how often do you participate in any organized group activities or meetings involving other teachers at your school…

	Number of times per week	Number of times per month	Number of times per year
…that primarily focus on administrative issues, such as schedules, upcoming events, and teachers work assignments?			
…that primarily focus on issues pertaining to student instruction/behavior?			

23. Think of changes that you have made *over the past year* that were due to a suggestion from another teacher in your school OR due to having observed another teachers in your school.

Do NOT include changes that were due to a principal, or to someone outside of your school, that you were required to make, or that occurred as a regular part of the school calendar (for example, changes that always occur when switching from fall to spring semesters).

Mark all that apply.

Changes in…	**Mark all that apply**
…classroom materials that you use such as	
Handouts	☐
Books	☐
Hands-on learning materials	☐
Computer software	☐
Assessments (tests)	☐
Behavior charts	☐
Parent communication product (for example, daily reports)	☐
Other (please describe)	
… how you teach lessons that you've taught in the past	☐
… curriculum that involve teaching new lessons	☐
… the homework you assign to students	☐
… how you handle behavior problems involving an individual student	☐
… your overall approach to managing student behavior in your class	☐
… classroom management unrelated to discipline	☐
… strategies for communicating with parents	☐
… the classroom setting (physical environment)	☐
… your own understanding of materials/procedures that you currently use	☐
… your own understanding of the content of what you teach	☐
… your approach to teaching specific groups of students (for example, students who are less proficient in English than they are in another language)	☐
… your approach to any aspect of extra-curricular activities that you might be involved with (for example, coaching, tutoring or helping in an after school program).	☐

COLLABORATIVE STRATEGIC READING FOR FIFTH GRADERS

SPRING 2008

TEACHER SURVEY, VERSION B

Paperwork Burden Statement

Dear Teacher:

The Collaborative Reading Strategies Study (CSR Study) is a groundbreaking study designed to test an innovative method for teaching reading comprehension in the fifth grade. Your participation is important and appreciated, but you do have the right to skip any question that you do not wish to answer. Below are answers to some general questions concerning this survey.

What is the purpose of this survey?

The purpose of this survey is to collect information about the classrooms of the teachers participating in the study.

Who is conducting this survey?

The CSR Study was commissioned by the Department of Education's Institute of Education Sciences, and is administered by Regional Educational Laboratory–Southwest, American Institutes for Research, Caliber Associates and RG Research Group.

Why should you participate in this survey?

Policymakers and educational leaders rely on findings from studies like the CSR Study to inform their decisions. The current study will fill a critical gap in the reading research literature as to what is effective for improving reading achievement, especially for children who are English language learners. Your participation in the study will help us to find out whether CSR is an effective solution.

Will your responses be kept confidential?

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 or individual. We will not provide information that identifies you or your district to anyone outside the study team, except as required by law.

How will your information be reported?

The information you provide will be combined with the information provided by other teachers in statistical reports. No information that links your name, address, or telephone number with your responses will be included in any reports related to the study.

Where should you return your completed survey?

Please return the completed survey to the person who gave you the survey.

Who can you contact about the survey?

If you have any questions about the survey, you can ask them from the person who gave you the survey, or you can contact the coordinator of data collection, <Name and contact information of data collection coordinator>.

Thank you for your cooperation in this very important effort!

Classroom Context

Class Structure

1. Which category best describes the class(es) for which you provide reading and social studies instruction?

- ☐ I have a self-contained classroom. I am the main reading and social studies teacher for students enrolled in the classroom. (A few students may get extra help from a reading specialist or other teacher.)
- ☐ I team teach with another teacher. We share joint responsibility for reading and social studies instruction for *all* our students.
- ☐ I team teach with another teacher. I have responsibility for reading and social studies instruction for *some* of the students, and my team teacher has responsibility for reading and social studies instruction for *other* students. (That is, we are regularly responsible for *different groups* of students within the class.)
- ☐ I do not teach reading.
- ☐ I do not teach social studies.
- ☐ Other. *Please specify.*

2. A student teacher is defined as someone who is not yet licensed or certified to teach, who is completing a degree program that will lead to a degree in teaching and licensure or certification, and who will take over the teaching of your class for a period of time. Please check the statement/s that best describe your experience with student teachers this school year and fill in the number of weeks a student teacher taught in your classroom.

- ☐ I had a student teacher in my classroom in the fall. S/he took over all teaching for ____ weeks.
- ☐ I have or will have a student teacher in my classroom this spring. S/he will take over all teaching for ____ weeks.
- ☐ I did not have a student teacher in my classroom during the fall and I will not have a student teacher in my classroom this school year.

3. Think about the students for whom you currently provide reading instruction. Do you expect to provide reading instruction for most of these students for the full school year, or does your school periodically re-assign students to classes for reading over the school year?

- ☐ I expect to continue to provide reading instruction for most of the students in my current class all year.
- ☐ My school periodically re-assigns students in my current class for reading.

4. In answering 4a-4g include ALL of the students to whom you teach reading, whether you teach reading on your own in a self-contained classroom, to a group that includes students from other classes, or to more than one group of students.

	Enter # below
a. What is the <u>total number</u> of students to whom you currently teach reading?	
b. How many of your reading students receive <u>intervention services</u> in reading from you or another teacher or tutor? **Reading Intervention** is a program designed for **struggling readers** to be used only with struggling readers in addition to the core reading program.	
c. How many of your students are reading <u>at or above</u> grade level?	
d. How many of your students are reading one year below grade level?	
e. How many of your students are reading two or more years below grade level?	
f. How many of your students are <u>English Language Learners</u> (ELL), also referred to as Limited English Proficient (LEP), English for Speakers of Other Languages (ESOL), or English as a Second Language (ESL) students? (e.g., Spanish, Russian, Chinese)	
g. How many of your students are <u>special education students</u> with Individualized Education Plans (IEP's) who receive special education services in reading?	

Class Behavior

5. At this point in the school year, how would you rate the behavior in your class? *Please mark (x) one box.*

- ☐ Group misbehaves very frequently and is almost always difficult to handle
- ☐ Group misbehaves frequently and is often difficult to handle
- ☐ Group misbehaves occasionally
- ☐ Group behaves well
- ☐ Group behaves exceptionally well

Instructional Resources

6. In a typical week, do you have paid and/or volunteer aides or specialists assist in your class in the following ways? *Please indicate separately whether paid and/or volunteer aides assist you in your classroom. If you do not have aides or specialists in a particular category, you only need to check the relevant box ☐ in the right column.*

	Paid aide/specialist	**Volunteer** aide/specialist	Do ***not*** have this type of aide or specialist assistance
Regular Aides			
Working directly with children on instructional tasks	☐	☐	☐
Doing non-instructional work (e.g., photocopying, preparing materials, etc.)	☐	☐	☐
Pulling children out of class to work on instructional tasks	☐	☐	☐
Special Education Aides or Specialists			
Working directly with children on instructional tasks	☐	☐	☐
Doing non-instructional work (e.g., photocopying, preparing materials, etc.)	☐	☐	☐
Pulling children out of class to work on instructional tasks	☐	☐	☐
ESL or Bilingual Education Aides or Specialists			
Working directly with children on instructional tasks	☐	☐	☐
Doing non-instructional work (e.g., photocopying, preparing materials, etc.)	☐	☐	☐
Pulling children out of class to work on instructional tasks	☐	☐	☐
Reading Specialists Who Work Primarily with Students			
Working directly with children on instructional tasks	☐	☐	☐
Doing non-instructional work (e.g., photocopying, preparing materials, etc.)	☐	☐	☐
Pulling children out of class to work on instructional tasks	☐	☐	☐

Reading Program

7. What reading program (such as Open Court Reading) or curriculum do you use in your classroom? Please check all that apply.

- ☐ Basal reader, such as *Open Court* or *Houghton Mifflin*
- ☐ School-wide literacy model, such as Success for All or Balanced Literacy
- ☐ Teacher (i.e., self) developed reading curriculum
- ☐ District or school developed literature-based (i.e., trade books) curriculum
- ☐ Other (please specify) _____

8. What kind of professional development have you received regarding the reading program or curriculum that you are currently using?

- ☐ None
- ☐ Start-up training only
- ☐ Ongoing support only (e.g., mentoring, coaching, consultation)
- ☐ Start-up training plus ongoing support

9. What social studies program or curriculum do you use in your classroom? Please check all that apply.

☐ Social studies textbook

Title of the text book:_____

If a social studies textbook is used, is it supplemented with trade books?
☐ Yes
☐ No

☐ Teacher (i.e., self) developed social studies curriculum

☐ District developed social studies curriculum

☐ Social studies curriculum consisting primarily of trade books

☐ Other (please specify) _____

10. How many weeks are spent on student/class projects?

_____ weeks

Collaboration among Teachers

We would like to learn about teachers' experiences collaborating with other teachers in their schools. Please think about both formal activities at your school intended to encourage collaboration and informal conversations you have with other teachers.

11. Not including the current school year and not including student teaching, how many years have you been a teacher? If this is your first year teaching, answer "zero."

____ years

12. Not including the current school year and not including student teaching, how many years have you taught in your current school? If this is your first year in this school, answer "zero."

_____ years

13. Some teachers work independently while other teachers prefer to get input from other teachers. Would you say you get…

- ☐ No input
- ☐ Minimal input
- ☐ Moderate input
- ☐ A great deal of input

14. How comfortable are you receiving advice from other teachers?

- ☐ Not at all comfortable
- ☐ Slightly comfortable
- ☐ Moderately comfortable
- ☐ Completely comfortable

15. How comfortable are you offering advice to other teachers?

- ☐ Not at all comfortable
- ☐ Slightly comfortable
- ☐ Moderately comfortable
- ☐ Completely comfortable

16. How supportive are other teachers at your school when you need help or advice with teaching?

- ☐ Not at all comfortable
- ☐ Slightly comfortable
- ☐ Moderately comfortable
- ☐ Completely comfortable

17. How receptive are other teachers at your school when you need help or advice with teaching?

- ☐ Virtually no teachers are receptive
- ☐ Some teachers are receptive, but a majority are not
- ☐ A majority of teachers are receptive, but some are not
- ☐ Nearly every teacher is receptive

18. In general, how often do you participate in any organized group activities or meetings involving other teachers at your school…

	Number of times per week	Number of times per month	Number of times per year
…that primarily focus on administrative issues, such as schedules, upcoming events, and teachers work assignments?			
…that primarily focus on issues pertaining to student instruction/behavior?			

19. Think of changes that you have made *over the past year* that were due to suggestion from another teacher in your school OR due to having observed another teachers in your school.

Do NOT include changes that were due to a principal, or to someone outside of your school, that you were required to make, or that occurred as a regular part of the school calendar (for example, changes that always occur when switching from fall to spring semesters).

Mark all that apply

Changes in…	**Mark all that apply**
…classroom materials that you use such as	
Handouts	☐
Books	☐
Hands-on learning materials	☐
Computer software	☐
Assessments (tests)	☐
Behavior charts	☐
Parent communication product (for example, daily reports)	☐
Other (please describe)	
… how you teach lessons that you've taught in the past	☐
… curriculum that involve teaching new lessons	☐
… the homework you assign to students	☐
… how you handle behavior problems involving an individual student	☐
… your overall approach to managing student behavior in your class	☐
… classroom management unrelated to discipline	☐
… strategies for communicating with parents	☐
… the classroom setting (physical environment)	☐
… your own understanding of materials/procedures that you currently use	☐
… your own understanding of the content of what you teach	☐
… your approach to teaching specific groups of students (for example, students who are less proficient in English than they are in another language)	☐
… your approach to any aspect of extra-curricular activities that you might be involved with (for example, coaching, tutoring or helping in an after school program).	☐

Appendix J. Fall coaching observation form

Background information

Coach	**Today's date** _/ /_____ mm dd yyyy	**Strategy** (Underline) Preview Click and Clunk Get the Gist Wrap up
Teacher	**Start time** a.m. p.m.	**Instructional phase** (Underline) Modeling Teacher-assisted
School	**End time** a.m. p.m.	**Number of days** _____
District		
State		

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

Any special circumstances that interrupted instruction? (Please explain below.)

Note to observer: Focus on primary teacher for rating purposes. If a student teacher is leading class, please do not observe and reschedule the observation.

Time	Field notes	CPB / CSF evidence	Food for thought	Notes

CSR Coaching Feedback

Teacher: _____ School: **_____District**: **_____**
Date: **_____**

Kudos

Kudos for:

Food for thought

Consider the following:

Appendix K. Multiple imputation

Multiple imputation is a Monte Carlo technique in which missing values are replaced by $m > 1$ simulated versions. By convention, generating 5–10 datasets is considered sufficient for obtaining parameter estimates that are close to being fully efficient (Little and Rubin 1987; Rubin 1987; Schafer 1997; Schafer and Graham 2002). In this study, m was set at 10.

The overall estimate and its standard error are derived from the m multiply imputed datasets. The overall estimate is the average of individual estimates from the imputed datasets:

$$\overline{Q} = \frac{1}{m}\sum_{j=1}^{m}\hat{Q}_j \ .$$

However, the overall standard error has to be adjusted for within-imputation and between-imputation variance. It is the square root of the total variance, which combines within-imputation and between-imputation variance:

$$T = \overline{U} + \left(1 + \frac{1}{m}\right)B \ .$$

Within-imputation variance is calculated by

$$\overline{U} = \frac{1}{m}\sum_{j=1}^{m}U_j$$

where $U_j = \mathrm{var}(\hat{Q}_j)$.

Between imputation variance is calculated by

$$B = \frac{1}{m-1}\sum_{j=1}^{m}\left(\hat{Q}_j - \overline{Q}\right)^2 \ .$$

The descriptive analysis presented in chapter 2 and appendix D suggested that student characteristics may explain some of the missing patterns on the pretest and posttest; special education status or lower baseline Group Reading Assessment and Diagnostic Evaluation (GRADE) scores were associated with student attrition. The assumption of missing completely at

random (MCAR) therefore cannot be made. A more appropriate assumption is that missing patterns can be predicted by observed covariates or that missing at random (MAR) better describes the type of missingness (Rubin 1987). The MAR assumption implies that missing pretest or posttest scores do not depend on unobserved covariates, after controlling for observed ones. MAR, however, cannot be tested, and it is not possible to test whether missingness on pretests or posttests depends on the values that are missing (the not missing at random assumption).

Multiple imputation was implemented using the multivariate stochastic sequential regression-based multiple imputation method (Raghunathan et al. 2001) by applying Imputation and Variance Estimation Software (IVEware).

Sample selection criteria

Chapter 2 includes a detailed explanation of the sample composition over time. The multiply imputed datasets used for the confirmatory impact analysis included all students who met the following criteria:

- Were enrolled in study classrooms at baseline, had parental consent, and were testable (that is, not excluded because of insufficient English fluency, special education status, or the need for testing accommodations that prevented administration of the GRADE).
- Were present at baseline but provided consent for posttesting only.
- Had at least one valid GRADE test score (pretest or posttest).

As a result, the sample used for imputation included 1,355 students.

Missing data

A statistically significant difference was observed between Collaborative Strategic Reading (CSR) and control groups in the pretest GRADE total scores (table K-1). Higher levels of missing data were expected for pretest and posttest scores than for demographic variables because of attrition and because students who were present at baseline and received consent during the school year (after pretesting was complete) were included in the analysis.

The models described in appendix G for use in attrition analyses were used to conduct missing data analyses, with different outcome variables. Missing data analyses explored whether study condition, pretest GRADE scores, or available demographic information (race/ethnicity, gender, students' English language learner (ELL) status, free or reduced-price lunch status, special education status) were related to missing posttest GRADE scores. Missing data analyses were conducted on students who were part of the analytic sample (n = 1,355). A binary outcome variable was created to indicate whether or not a student had nonmissing posttest GRADE scores (yes = 1, no = 0).

Table K-1. Rates of student-level missing data

Student characteristic	*CSR* (n = *681)*		*Control* (n = *674)*		*Total* (n = *1,355)*		*Chi-squared* p-*value*
	Number	*Percent*	*Number*	*Percent*	*Number*	*Percent*	
Pretest GRADE total score	22	3.23	53	7.86	75	5.53	0.00
Posttest GRADE total score	75	11.01	77	11.42	152	11.21	0.81
Language status	18	2.64	13	1.93	31	2.28	0.38
Gender	17	2.50	13	1.93	30	2.21	0.48
Race/ethnicity	17	2.50	13	1.93	30	2.21	0.48
Free or reduced price lunch	18	2.64	13	1.93	31	2.28	0.38
Special education	17	2.64	13	1.93	30	2.21	0.48

Source: Authors' analysis of student GRADE scores and study records.

A procedure similar to that used in the attrition analysis was used to run a series of two-level hierarchical generalized linear models (HLM) to test whether the study condition predicted missing GRADE scores at posttest. Model M1 included the *CSR* indicator. (See appendix G for details of the models used in these analyses.) Model M2 examined whether missingness was related to student demographics. Model M3, conducted as a sensitivity analysis, included the *CSR* indicator, the demographic variables, and their interactions. Additional sensitivity analyses were conducted by analyzing the same model using logistic regressions.

The missing data analysis including only the *CSR* indicator variable (model M1) showed that study condition was not significantly associated with missing posttest GRADE scores (estimate = –0.12, standard error = 0.22, $p = 0.58$). The results of model M2 indicated that the study condition remained insignificant in the model (estimate = –0.09, standard error = 0.23, $p = 0.68$). However, student pretest GRADE scores were negatively associated with the likelihood of having missing data on the posttest GRADE variable (estimate = –0.03, standard error = 0.01, $p < 0.01$), whereas special education status and ELL status were both positively associated with being more likely to have missing data on the posttest GRADE variable (special education status: estimate = 1.09, standard error = 0.26, $p < 0.01$; ELL status: estimate = 0.74, standard error = 0.25, $p < 0.01$). These results suggest that students with special education status, ELL status, or lower pretest GRADE scores were more likely to have missing posttest GRADE scores.

Results from models M1 and M2 consistently showed that study condition was not related to missing data on the posttest GRADE. The results of model M3, which included *CSR*demographics* interactions, suggest that special education students (estimate = 1.27, standard error = 0.35, $p < 0.01$); students who did not participate in a free or reduced-price lunch program (estimate = –0.75, standard error = 0.32, $p = 0.19$); and students with lower pretest GRADE scores (estimate = –0.05, standard error = 0.01, $p < 0.01$) were more likely to report missing data on the *Posttest* GRADE variable (table K-2). Students in CSR classrooms were also less likely to have missing data on the *Posttest* variable. However, the likelihood that students in the CSR condition had missing data on the *Posttest* variable increased if they also participated in

a free or reduced-price lunch program (estimate = 1.21, standard error = 0.49, $p = 0.01$). The results of logistic regression models coincided with most of the results of the HLM analyses, with one exception: the likelihood for students in the CSR condition to have missing posttest GRADE data was lower if they were also receiving free or reduced-price lunch.

Table K-2. Analysis of missing posttest GRADE scores using two-level hierarchical linear model

Model/variable	*Coefficient*	p-*value*
Model M1:Two-level HLM		
Study condition	–0.12 (0.22)	0.58
Model M2: Two-level HLM with student covariates		
Study condition	–0.09 (0.23)	0.68
Pretest GRADE score	–0.03 (0.01)	< 0.01
Special education status	1.09 (0.26)	< 0.01
Hispanic	–0.42 (0.25)	0.09
Free or reduced-price lunch status	–0.20 (0.24)	0.40
Gender	0.15 (0.20)	0.46
Language status	0.74 (0.25)	< .01
Model M3: Two-level HLM with student covariates and interaction terms		
Study condition	–4.06 (1.82)	0.03
Pretest GRADE score	–0.05 (0.01)	< 0.01
Special education status	1.27 (0.35)	< 0.01
Hispanic	–0.58 (0.34)	0.09
Free or reduced-price lunch status	–0.75 (0.32)	0.02
Gender	–0.07 (0.28)	0.79
Language status	0.41 (0.35)	0.24
Pretest GRADE score*Study condition	0.02 (0.02)	0.17
Special education status*Study condition	–0.38 (0.52)	0.47
Hispanic*Study condition	0.35 (0.49)	0.47
Free or reduced-price lunch status*Study condition	1.21 (0.49)	0.01
Gender*Study condition	0.45 (0.40)	0.26
Language status*Study condition	0.72 (0.51)	0.16

Note: Figures in parentheses are standard errors.
Source: Authors' analysis of student GRADE scores and study records.

Similar analyses were conducted to explore whether study condition or demographic information (race/ethnicity, gender, language status, free or reduced-price lunch status, special education status) were related to missing pretest GRADE scores. According to models M4 and M5, the study condition is not significantly related to missing pretest GRADE scores. Model M5 also

shows that none of the student characteristics is significantly related to missing pretest GRADE scores. (A model corresponding to M3 that would have included the *CSR* indicator and its interaction terms with demographic variables would not converge and was therefore omitted from table K-3 and the analysis exploring missing patterns for pretest GRADE scores.)

Table K-3. Analysis of missing pretest GRADE scores using two-level hierarchical generalized linear model

Model/variable	*Coefficient*	p-*value*
Model M4:Two-level HLM		
Study condition	–0.44 (0.49)	0.36
Model M5: Two-level HLM with student covariates		
Study condition	–0.48 (0.49)	0.33
Special education status	–0.17 (0.76)	0.82
Hispanic	–0.23 (0.54)	0.68
Free or reduced-price lunch status	–0.65 (0.47)	0.17
Gender	–0.27 (0.42)	0.52
Language status	–0.50 (0.50)	0.32

Note: Figures in parentheses are standard errors.

Source: Authors' analysis of student GRADE scores and study records.

Imputation model

The IVEware program uses multivariate stochastic sequential regression-based multiple imputation. It yields imputed values for each individual in the dataset, conditional on all the values observed for that individual. The basic strategy is to create imputations through a sequence of multiple regressions, varying the type of regression model by the type of variable (continuous, binary, categorical, counts) being imputed. Covariates include all other variables observed or imputed for that individual (that is, all other variables included in the dataset used for imputation). The imputations are defined as draws from the posterior predictive distribution specified by the regression model with a flat or noninformative prior distribution for the parameters in the regression model.

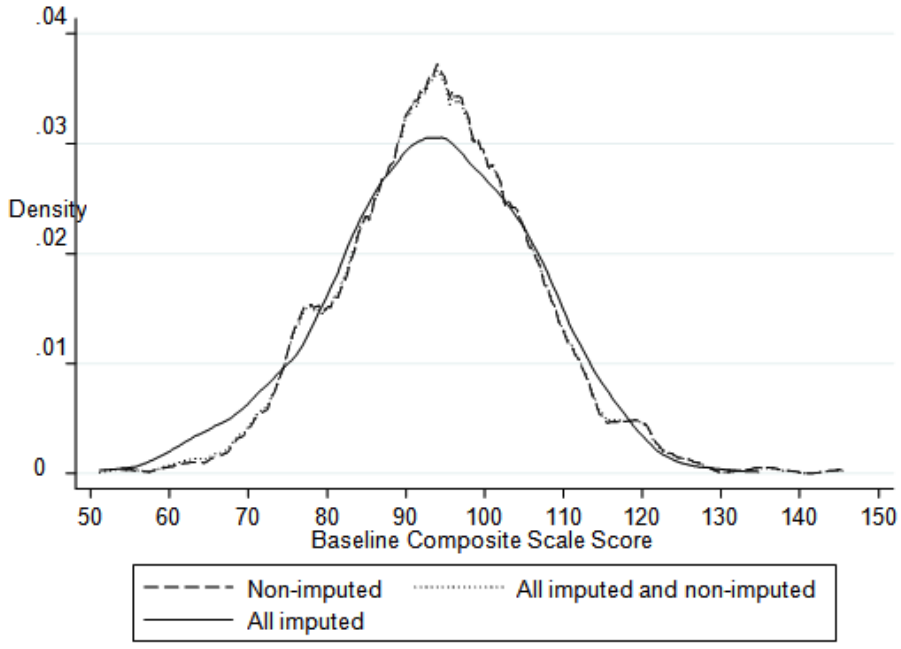
Multiple imputation was conducted separately for the CSR and control groups. Ten multiply imputed datasets were created and then combined into a final analytic dataset. The imputation model included variables for the following: pretest and posttest GRADE scores, race/ethnicity, gender, free or reduced-price lunch status, special education status, language status, teachers' Spanish fluency, class size, teaching experience (total number of years), and dummy indicator variables to acknowledge teacher and school clustering. The model included more variables than the analytic models themselves (as often recommended) and is based on findings from the attrition analyses (Allison 2002; Schafer 1997; Schafer and Graham 2002).

Diagnostics

Graphic diagnostics compare the distribution of observed and imputed values through kernel density plot estimations. *Imputed values* include only the values from each of the 10 datasets that were imputed; *nonimputed values* refer to the original, observed data values. Graphic diagnostics could flag a potential misspecification in the multiple imputation model if, for example, the distribution of imputed posttest skewed greatly to the right: because low-performing students are more likely to attrite, they would not be expected to have higher imputed posttest values than students without the missing posttest.

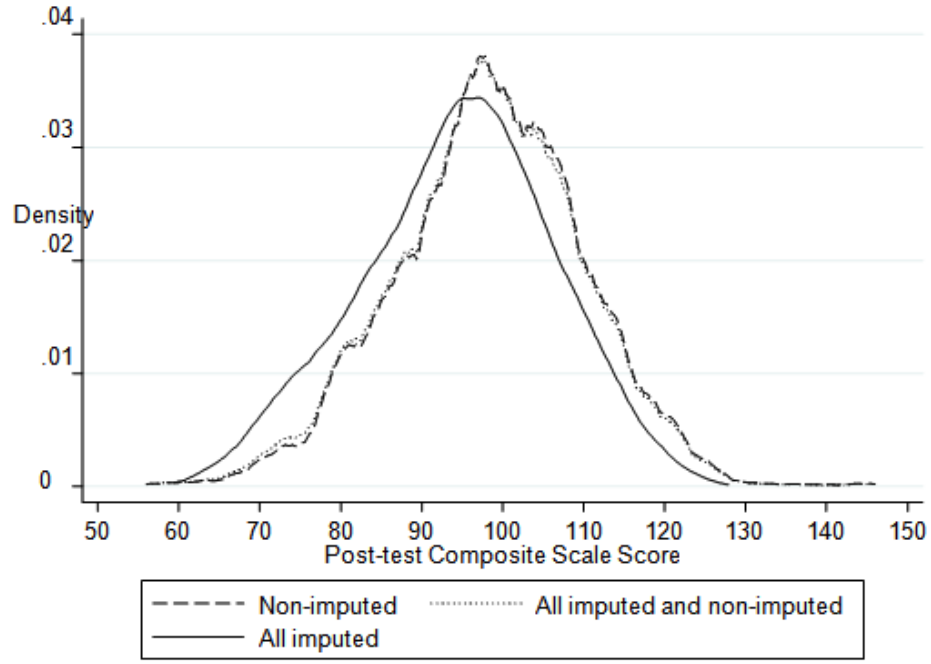
Figures K-1 and K-2 show the distributions of imputed scores, nonimputed scores, and a combination of imputed and nonimputed scores when all 10 simulations are combined. The distribution of the combined scores follows very closely the distribution of the nonimputed scores (as expected, given that 11.2 percent of the posttest data and 5.5 percent of the pretest data were missing and the correlation between observed pretest and posttest scores is 0.84). The distribution of the imputed scores is somewhat different from the distribution of the nonimputed scores. As suggested by attrition analyses, the imputed posttest GRADE scores had lower means than the nonimputed scores.

Figure K-1. Pretest imputed datasets overlaid onto nonimputed dataset



Note: Figure is based on the complete sample.
Source: Authors' analysis of student GRADE scores and study records.

Figure K-2. Posttest imputed datasets overlaid onto nonimputed dataset



Note: Figure is based on the complete sample.

Source: Authors' analysis of student GRADE scores and study records.

Numeric diagnostics analyze the mean and standard deviation of observed values, imputed values, and combined observed and imputed values in search of anomalies. Table K-4 presents the means and standard deviations for pretest and posttest scores for the sample. Conclusions parallel the graphic diagnostics, with similar means and standard deviations of imputed and observed posttest scores. (This result is not surprising, given that 11.2 percent of posttest and 5.5 percent of pretest were missing.) Table K-4 also presents the ratio of the difference between the mean of imputed and observed values to the standard deviation of observed values. For this ratio, Stuart et al. (2009) suggest that an absolute value greater than 1 may indicate that the variable should be flagged for further investigation. The numeric diagnostics do not suggest potential misspecification of the multiple imputation model, as none of the ratios from the imputed datasets approaches this threshold.

Table K-4. Pretest and posttest GRADE scores

Score	n	*Mean*	*Minimum*	*Maximum*	*Mean/standard deviation ratio*[a]	*Standard deviation ratio*[b]
Pretest						
Nonimputed	12,800	94.30 (12.09)	54.00	146.00		
Imputed	750	95.18 (12.81)	60.12	122.96	0.07	1.06
Combined	13,550	94.35 (12.05)	54.00	146.00		
Posttest						
Nonimputed	12,030	98.75 (11.52)	56.00	146.00		
Imputed	1,520	94.64 (13.68)	50.63	128.64	0.36	1.19
Combined	13,550	98.29 (11.60)	56.00	146.00		

Note: Figures in parentheses are standard deviations. GRADE scores include values from all 10 imputed datasets; the reported sample sizes for the observed values are 10 times greater than the actual observed values, because they are duplicated in each of the 10 complete datasets.
a. Ratio of difference between mean of imputed and observed values to standard deviation of observed values.
b. Ratio of standard deviation of imputed values to standard deviation of observed values.
Source: Authors' analysis of student GRADE scores and study records.

Appendix L. Assigning students to cooperative learning groups

As described in chapter 3, the initial two-day training session provided to Collaborative Strategic Reading (CSR) teachers consisted of four segments. During the third segment (implementation practices and logistics), teachers were taught a four-step procedure for assigning students to the small tutoring groups that would be used throughout the school year. (The study team did not conduct observations to determine whether teachers followed the prescribed procedure.) The four-step procedure consisted of the following:

- *Step 1:* Rank students by achievement. List all students in the class, starting with the highest achiever and ending with the lowest achiever, based on recent test scores, grades, and knowledge of students' reading levels. This ranking can be done fairly quickly, as the purpose is simply to ensure that all of the strong readers or weak readers are not assigned to any one group.
- *Step 2:* Identify leaders. Put a mark next to the names of students who can lead a group. Each CSR group should have at least one leader.
- *Step 3:* Select the first group. Choose the top, bottom, and middle student from the ranked class list. Assign these students to group one, unless they are all the same gender, do not reflect the ethnic composition of the class, or are worst enemies or best friends. If any of these are the case, readjust the group by moving up or down one student on the list. Choose a leader for the group.
- *Step 4:* Select the remaining groups. To select the second group, repeat step 3. Continue until all students have been assigned to a group. If there are any remaining students, assign them to one of the existing groups. After using the procedure above, review the composition of each group. Think about whether each group can work together as a team. If potential problems are evident, make adjustments. Make sure that all groups include students at different levels of English language development.

Appendix M. Critical procedural behaviors for Collaborative Strategic Reading strategies

This appendix lists the cooperative learning discussion topics and the key critical procedural behaviors pertaining to the modeling and teacher-assisted phases of the Collaborative Strategic Reading (CSR) study.

Cooperative learning discussion topics covered

- A four-step procedure for assigning students to mixed-ability CSR cooperative learning groups.
- Descriptions of the CSR roles students are expected to assume in small groups (leader, timekeeper, scorekeeper, clunk expert, gist expert, announcer, and encourager).
- Classroom activities that teach students the roles they should assume in their small groups.
- Behavioral management guidance for students working in small groups.
- Instructions for students in proper use of CSR materials, including student learning logs, role cue sheets, and small group score sheets.
- Follow-up instructional activities designed to support aspects of strategy use that students tend to find particularly difficult.
- Accommodations for at-risk students and individuals with special needs.

Modeling phase

Preview

1. The teacher explains the strategy.
2. The teacher explains when the strategy is used.
3. The teacher explains why the strategy is important.
4. The teacher models the strategy.
    a. The teacher models how to preview using text features.
    b. The teacher lists at least two brainstorms.
    c. Using a think-aloud, the teacher tells the students what information he or she used to generate the brainstorms.
    d. The teacher lists at least two predictions.
    e. Using a think-aloud, the teacher tells students what information was used to generate the predictions.

Click and Clunk

1. The teacher explains the strategy.
2. The teacher explains when the strategy is used.
3. The teacher explains why the strategy is important.
4. The teacher models the strategy.
    a. Before reading the first paragraph, the teacher tells the students that he or she is going to make a mental note of words that are not understood.
    b. After reading the paragraph, the teacher identifies "clunks" and explains that there are four fix-up strategies that can help to figure out the meaning of the clunk.
    c. Using the clunk fix-up strategies transparency, the teacher demonstrates by thinking aloud while using each fix-up strategy until one helps to figure out the meaning of the word.
    d. The teacher writes the definition on the learning log transparency.
    e. The teacher generates definitions for additional words using the fix-up strategies and think-aloud procedures.

Get the Gist

1. The teacher explains the strategy.
2. The teacher explains when the strategy is used.
3. The teacher explains why the strategy is important.
4. The teacher models the strategy.
    a. Using the *Get the Gist* transparency, the teacher names the three steps for getting the gist.
    b. The teacher uses the learning log transparency to show students where to write the gist.
    c. The teacher tells the students that they will be shown how to get the gist for the paragraph they just read.
    d. After reading a paragraph, the teacher names the most important *who* or *what* in that paragraph.
    e. Using a think-aloud, the teacher explains how he or she determined the *who* (person) or *what* (thing).
    f. The teacher tells students the most important information about the *who* or *what* in the paragraph. The information is listed on the learning log transparency. The teacher thinks aloud while synthesizing information into a general idea or concept that becomes the basis for the gist.
    g. The teacher tells students that they are going to write a gist of 10 words or less, leaving out the details. Students are reminded that the main *who* or *what* counts as

one word; gists must be a complete sentence and contain information that will help to remember the important details in the paragraph.

h. The teacher writes the gist on the transparency as they think aloud explaining how the gist helps to remember the important information about the *who* or *what*.

i. The teacher models generating gists for additional paragraphs using the procedures described above.

Wrap up (questioning)

1. The teacher explains the strategy.
2. The teacher explains when the strategy is used.
3. The teacher explains why the strategy is important.
4. The teacher models the strategy.
    a. The teacher reminds students of the two parts of *Wrap up:* generating questions and writing a summary.
    b. The teacher thinks aloud while modeling how to generate at least two questions, locate answers, and label the question type (right there, think and search, author and you, on my own). They explain the reason for the label.
    c. The teacher writes the questions on the learning log transparency.

Wrap up (review)

1. The teacher explains the strategy.
2. The teacher explains when the strategy is used.
3. The teacher explains why the strategy is important.
4. The teacher models the strategy.
    a. The teacher chooses the first gist from a student's learning log.
    b. The teacher makes a list of the important information that will be included in the summary.
    c. The teacher writes the summary using the gist as the topic sentence.
    d. The teacher models writing a summary for at least one additional paragraph.

Teacher-assisted phase

Strategy: Preview

1. The teacher reviews the *Preview* strategy by asking questions. (What is the strategy? When is it used? Why is it used? How is it performed?)
2. The teacher introduces the day's topic.
3. While previewing the text with students (using text features), the teacher reminds students to think about what they already know about this topic.

4. The teacher calls on students to share their brainstorms.
5. The teacher asks students to tell how they came up with the brainstorm.
6. The teacher records students' brainstorming ideas in the Brainstorm section of the learning log transparency.
7. The teacher asks students to write brainstorms (what they know about the topic) in their learning logs.
8. The teacher asks students to predict what they might learn.
9. The teacher calls on students to share their predictions.
10. The teacher asks students to explain how they came up with their predictions.
11. The teacher records students' predictions in the Predict section of the learning log transparency.
12. The teacher asks students to record their predictions in their learning logs.

Strategy: Click and Clunk

1. The teacher reviews the *Click and Clunk* strategy by asking questions (What is the strategy? When is it used? Why is it used? How is it performed?).
2. The teacher reads or asks a student to read the first paragraph (or section) of the passage aloud.
3. The teacher reminds students to make a mental note of words they do not understand.
4. The teacher asks students to write their clunks in their learning logs.
5. The teacher asks a volunteer to share a clunk.
6. The teacher tells students to locate the clunk in the passage and write it in their learning logs.
7. The teacher uses the clunk fix-up strategies transparency and asks the student who has the clunk to state the first fix-up strategy.
8. The teacher asks all students to go back to the paragraph and apply the first fix-up strategy, by rereading the sentence with the clunk in it and thinking about what would make sense.
9. The teacher asks students if the first fix-up strategy helped.
10. If the first fix-up strategy helped, the teacher asks the students to explain how they came up with the meaning. If the strategy did not help, the teacher goes to the second fix-up strategy and follows the same procedure. The teacher continues this process until the meaning of the clunk is determined. The teacher asks all students to write the definition in their learning logs.
11. The teacher uses the learning log transparency to show students where to write their definition.

12. The teacher reads or asks a student to read the next paragraph aloud, using the same procedure. He or she continues the procedure of reading and figuring out clunks for several paragraphs.

Strategy: Get the Gist

1. The teacher reviews the *Get the Gist* strategy by asking questions (What is the strategy? When is it used? Why is it used? How is it performed?).
2. The teacher reads or asks a student to read the first paragraph (or section) of the passage aloud.
3. The teacher asks students to state whether the passage was mostly about a *who* or a *what*. The teacher calls on several students to determine whether there is agreement about whether the passage was mostly about a *who* or a *what*.
4. After establishing whether the passage was about a *who* or *what*, the teacher asks students to identify the *who* or the *what* (topic), writing the answer on the learning log transparency.
5. After students have determined the *who* or *what* for the main idea, the teacher leads the class in a discussion to determine the most important information about the *who* or *what*. The teacher should emphasize that students should be looking for the most important information, not details. The teacher writes the important information on the learning log transparency.
6. The teacher asks students to synthesize the important information into a general idea or concept that becomes the basis for the gist.
7. The teacher asks students to write a gist of 10 words or less, either independently or with a partner.
8. The teacher selects students to share their gists. The teacher critiques the gist by explaining how it either meets or does not meet the criterion of helping students remember the important information in the paragraph. If students' gists contain specific details, the teacher reminds them that the gist should be a general statement that helps them remember the important information in the paragraph. The teacher calls on additional students until a student provides a good gist. If there is a gist that requires only minor revisions, the teacher and students collaboratively work to improve it. The teacher may need to review main idea statements generated during the modeling phase to reinforce this concept.
9. The teacher considers eliciting more than one appropriate gist for each paragraph.

Strategy: Wrap up (questioning)

1. The teacher reviews the *Wrap up* (questioning) strategy by asking questions (When you think of questions, what do you think about? With which words do questions begin?).
2. The teacher leads the class in generating questions using a selection the class has recently read. One or more of the question types can be used.
3. The teacher writes the questions on the learning log transparency.

4. After each question is generated, the teacher calls on a student to answer it, label the question type, and explain the reason for the label. As students become proficient in generating questions of a particular type, the teacher may skip steps 2–4 and proceed to step 5.
5. The teacher uses the learning log transparency to show students where to write the questions.
6. The teacher gives students a few minutes to write questions in their learning logs.
7. After questions are generated, the teacher calls on students to ask one of their questions.
8. After each question is asked, the teacher calls on a student to answer it, label the question type, and explain the reason for the label.

Strategy: Wrap up (review)

1. The teacher reviews the *Wrap up* (review) strategy by asking questions (What is the strategy? Why is the strategy important?).
2. The teacher selects a student and chooses the first gist from the student's learning log.
3. The teacher leads students in generating a list of important information that will be included in the summary.
4. The teacher and students write a class summary, using the gist as the topic sentence.
5. The teacher and students write class summaries for at least one additional paragraph.
6. As students become more proficient, the teacher may ask students to write a summary, on which the teacher provides feedback.

Appendix N. Observer training for the subscale Expository Reading Comprehension observation instrument and interrater reliability

Observations using the Expository Reading Comprehension (ERC) observation instrument (ERC) were conducted by study members who had extensive experience with the protocol as a result of participating as observers in the Institute of Education Sciences–funded study for which the ERC was developed (James-Burdumy et al. 2009). For that study, each observer conducted about 25 observations over a two-year data collection period. They participated in a one-day refresher training, conducted by Dr. Joseph Dimino, as part of the current study.

Several items related to Collaborative Strategic Reading (CSR) practices were used in the current study, under the assumption that they offer reasonable indicators that CSR practices were in use. These appended items were completed only in classrooms in which instructional formats may have allowed CSR practices to be administered (for example, classrooms in which students worked in groups or pairs as opposed to whole class instruction).

ERC subscale interrater reliability was assessed by including a second observer in 20 percent (15 of 74) of the observations. Interrater reliability was calculated for a combined measure, including classroom format, materials used for instruction, and classroom management.

Interrater reliability calculations excluded all strategies for an interval in which no observations were made to avoid inflating agreement estimates. (That is, if both observers agreed that a given behavior did not happen, it was not included in the calculations.) These agreement rates are, in essence, tallies. Therefore, if one observer marked six observations and another observer marked the same six, the rate of agreement would be 100 percent. Categorical variables with yes/no responses yielded simple agreement and disagreement rates.

Average interrater reliability for classroom management (as measured by the percentage agreement between observers for the ERC protocol) was 97 percent (standard deviation = 4.12), ranging from 88 to 100 percent.

Appendix O. Descriptive statistics on Group Reading Assessment and Diagnostic Evaluation scores

This appendix presents Group Reading Assessment and Diagnostic Evaluation (GRADE) pretest and posttest descriptive statistics for the nonimputed and multiply imputed datasets.

Table O-1. Pretest and posttest GRADE scores: all study students, nonimputed data

Statistic	n	*Mean*	*Minimum*	*Maximum*	*Pre-post difference*
CSR					
Pretest score	659	93.88 (12.34)	54.00	146.00	
Posttest score	606	98.75 (12.01)	63.00	146.00	4.87
Control					
Pretest score	621	94.75 (11.82)	55.00	134.00	
Posttest score	597	98.75 (11.02)	56.00	138.00	4.00

Note: Figures in parentheses are standard deviations.
Source: Authors' analysis of student GRADE scores and study records.

Table O-2. Pretest and posttest GRADE scores: former and current English language learner students, nonimputed data

Statistic	n	*Mean*	*Minimum*	*Maximum*	*Pre-post difference*
CSR					
Pretest score	308	90.59 (10.73)	61.00	120.00	
Posttest score	301	95.20 (10.57)	63.00	120.00	4.61
Control					
Pretest score	238	92.57 (11.60)	55.00	121.00	
Posttest score	244	96.62 (10.63)	72.00	120.00	4.05

Note: Figures in parentheses are standard deviations.
Source: Authors' analysis of student GRADE scores and study records.

Table O-3. Pretest and posttest GRADE scores: non–English language learner students, nonimputed data

Statistic	n	*Mean*	*Minimum*	*Maximum*	*Pre-post difference*
CSR					
Pretest score	333	96.90 (13.03)	54.00	146.00	
Posttest score	301	102.21 (12.39)	68.00	146.00	5.31
Control					
Pretest score	370	96.18 (11.80)	61.00	134.00	
Posttest score	352	100.24 (11.06)	56.00	138.00	4.06

Note: Figures in parentheses are standard deviations.
Source: Authors' analysis of student GRADE scores and study records.

Table O-4. Pretest and posttest GRADE scores: all study students, multiply imputed data

Statistic	n	*Mean*	*Minimum*	*Maximum*	*Pre-post difference*
CSR					
Pretest score	681	93.93 (12.38)	54.00	146.00	
Posttest score	681	98.47 (12.06)	63.00	146.00	4.54
Control					
Pretest score	674	94.77 (11.95)	55.00	134.00	
Posttest score	674	98.10 (11.71)	50.63	138.00	3.33

Note: Figures in parentheses are standard deviations.
Source: Authors' analysis of student GRADE scores and study records.

Table O-5. Pretest and posttest GRADE scores: former and current English language learner students, multiply imputed data

Statistic	n	*Mean*	*Minimum*	*Maximum*	*Pre-post difference*
CSR					
Pretest score	316	90.61 (10.79)	61.00	121.23	
Posttest score	316	95.26 (10.63)	63.00	120.00	4.65
Control					
Pretest score	263	92.71 (11.87)	55.00	121.75	
Posttest score	263	96.12 (11.04)	63.36	120.00	3.41

Note: Figures in parentheses are standard deviations. Data include only students with known language status.
Source: Authors' analysis of student GRADE scores and study records.

Table O-6. Pretest and posttest GRADE scores: Non–English language learner students, multiply imputed data

Statistic	n	*Mean*	*Minimum*	*Maximum*	*Pre-post difference*
CSR					
Pretest score	349	96.74 (13.08)	54.00	146.00	
Posttest score	349	101.30 (12.52)	64.29	146.00	4.56
Control					
Pretest score	398	96.17 (11.89)	61.00	134.00	
Posttest score	398	99.44 (11.78)	50.63	138.00	3.27

Note: Figures in parentheses are standard deviations. Data include only students with known language status.
Source: Authors' analysis of student GRADE scores and study records.

Table O-7. Difference in pretest and posttest GRADE scores by student language status (former and current English language learner and non–English language learner students), multiply imputed data)

Statistic	n	*Mean*	*Estimate*	*Standard error*	*P-value*
Pretest					
FC-ELL	535	91.67 (11.44)	–2.47	0.70	0.00
Non-ELL	664	96.35 (12.55)			
Posttest					
FC-ELL	535	95.93 (10.88)	–2.52	0.68	0.00
Non-ELL	664	100.21 (12.23)			

Note: Standard deviation in parenthesis. Data include only students with known language status.

Source: Authors' analysis of student GRADE scores and study records.

Appendix P. Baseline equivalence results for multiply imputed analytic dataset

Table P-1. Mean sample characteristics for multiply imputed sample

Student characteristic	*CSR*	*Control*	*Total*	p-*value*
Gender				
Male	47.94 (51.90)	46.96 (50.46)	47.45 (51.43)	0.82
Female	52.06 (51.90)	53.04 (50.46)	52.55 (51.43)	
Race/ethnicity				
American Indian	1.91 (13.70)	1.62 (20.89)	1.76 (17.64)	0.77
Asian	7.59 (29.03)	6.41 (29.28)	7.00 (27.09)	0.89
Black	18.27 (38.72)	23.37 (42.44)	20.80 (40.66)	0.53
Hispanic	62.16 (49.63)	59.99 (52.22)	61.08 (49.93)	0.75
White	10.07 (30.19)	8.62 (28.12)	9.35 (29.16)	0.96
Free or reduced-price lunch				
Yes	72.88 (44.78)	71.82 (45.44)	72.35 (45.15)	0.68
No	27.12 (44.78)	28.18 (45.44)	27.65 (45.15)	
Special education				
Yes	9.72 (29.87)	9.53 (30.52)	9.62 (30.05)	0.78
No	88.94 (29.87)	90.47 (30.52)	90.38 (30.05)	
Language status				
FC–ELL student	47.74 (50.39)	39.14 (48.98)	43.46 (49.80)	0.14
Non–ELL student	52.26 (50.39)	60.86 (48.98)	56.54 (49.80)	
Baseline reading proficiency				
Pretest GRADE score	93.93 (12.38)	94.77 (11.95)	94.35 (12.14)	0.43

Note: Figures in parentheses are standard deviations. A three-level hierarchical linear model (HLM) with students nested in classrooms and classrooms nested in schools was used in the estimation process. The HLM included only the treatment indicator; no multiple-comparison correction was applied to yield more conservative *p*-values. The HLM varied based on the dependent variable. A logistic HLM was used for binomial outcomes such as *race/ethnicity* (categories dummy coded), *free or reduced-price lunch*, and *special education status*. The means and standard deviations were calculated from the available data across schools. Results are based on 10 multiply imputed datasets. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,355 students (681 CSR, 674 control).

Source: Authors' analysis of student GRADE scores and study records.

Appendix Q. Full analytic output tables

Confirmatory impact analysis

Table Q-1. Model 1: confirmatory impact analysis

Variable	*Estimate*	t-*value*	Pr > |*t*|
Fixed effects			
School 1	96.42 (1.21)	80.00	< .001
School 2	97.73 (1.82)	53.69	< .001
School 3	99.02 (1.23)	80.77	< .001
School 4	100.39 (2.35)	42.78	< .001
School 5	100.59 (1.58)	63.48	< .001
School 6	96.87 (1.34)	72.23	< .001
School 7	98.37 (1.63)	60.40	< .001
School 8	102.96 (1.69)	60.89	< .001
School 9	97.53 (1.72)	56.68	< .001
School 10	97.37 (1.23)	79.17	< .001
School 11	95.48 (1.30)	73.42	< .001
School 12	98.77 (1.62)	60.89	< .001
School 13	98.77 (1.02)	97.05	< .001
School 14	98.42 (1.16)	85.05	< .001
School 15	100.39 (1.85)	54.16	< .001
School 16	97.95 (1.61)	60.75	< .001
School 17	95.71 (1.40)	68.12	< .001
School 18	96.55 (1.22)	79.40	< .001
School 19	98.16 (2.30)	42.65	< .001
School 20	99.04 (1.79)	55.38	< .001
School 21	97.58 (1.27)	77.11	< .001
School 22	98.66 (1.80)	54.70	< .001
School 23	97.33 (1.04)	94.00	< .001
School 24	95.63 (1.75)	54.75	< .001
School 25	96.07 (1.50)	63.84	< .001
School 26	96.70 (1.32)	73.07	< .001
School 1*CSR	4.15 (1.65)	2.51	0.013
School 2*CSR	–2.68 (2.80)	–0.96	0.344
School 3*CSR	–2.12 (1.70)	–1.24	0.214

Variable	*Estimate*	t-*value*	Pr > |*t*|
School 4*CSR	–1.38 (2.70)	–0.51	0.609
School 5*CSR	–1.51 (1.94)	–0.78	0.436
School 6*CSR	–2.72 (2.82)	–0.96	0.342
School 7*CSR	5.00 (2.48)	2.02	0.048
School 8*CSR	–3.80 (1.96)	–1.93	0.054
School 9*CSR	4.95 (2.27)	2.18	0.029
School 10*CSR	2.67 (1.77)	1.51	0.132
School 11*CSR	4.00 (1.81)	2.21	0.027
School 12*CSR	–3.48 (2.33)	–1.50	0.135
School 13*CSR	–0.31 (1.64)	–0.19	0.851
School 14*CSR	0.79 (1.74)	0.45	0.651
School 15*CSR	–1.62 (2.25)	–0.72	0.473
School 16*CSR	0.03 (1.97)	0.01	0.989
School 17*CSR	5.07 (2.12)	2.39	0.017
School 18*CSR	–0.07 (2.12)	–0.03	0.975
School 19*CSR	–0.89 (3.06)	–0.29	0.772
School 20*CSR	–2.38 (2.23)	–1.07	0.285
School 21*CSR	1.00 (2.52)	0.40	0.691
School 22*CSR	1.97 (2.60)	0.76	0.448
School 23*CSR	1.79 (1.37)	1.31	0.189
School 24*CSR	0.69 (2.47)	0.28	0.781
School 25*CSR	1.78 (2.13)	0.84	0.404
School 26*CSR	3.46 (1.99)	1.73	0.083
Pretest GRADE score	0.78 (0.02)	40.64	< .001
Language status	–0.70 (0.42)	–1.68	0.094
Little Spanish	–1.64 (0.81)	–2.03	0.043
Spanish as a second language	–0.72 (0.95)	–0.76	0.450
Spanish as a first language	–0.45 (1.05)	–0.43	0.670

Random effects	*Variance*	*Chi–square*	p–*value*
Intercept (U_0)	0.07 (0.26)	15.99	>.50
Level 1 R	38.73 (6.22)		

Note: Figures in parentheses are standard errors for fixed effects and standard deviations for random effects. Results are based on 10 multiply imputed datasets. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,355 students (681 CSR, 674 control). The estimate of the effect of CSR is the weighted combined estimate of all *School*CSR* interaction terms (see appendix E).

Source: Authors' analysis of student GRADE scores, study records, and teacher surveys.

Table Q-2. Results of F-test for variation in school-level impacts

Outcome	*F-value*	*P-value of F-test*
GRADE scores	1.68	0.02

Note: A composite F-test was used to test whether the school-by-school variation in impacts is statistically significant. Results are based on 10 multiply imputed data sets.

Source: Authors' analysis of student GRADE scores, school records, and teacher survey.

Sensitivity analyses

Table Q-3. Sensitivity analysis 1: equal weighting

Variable	*Estimate*		t-*value*	Pr > |*t*|
Fixed effects				
School 1	96.42	(1.21)	80.00	< .001
School 2	97.73	(1.82)	53.69	< .001
School 3	99.02	(1.23)	80.77	< .001
School 4	100.39	(2.35)	42.78	< .001
School 5	100.59	(1.58)	63.48	< .001
School 6	96.87	(1.34)	72.23	< .001
School 7	98.37	(1.63)	60.4	< .001
School 8	102.96	(1.69)	60.89	< .001
School 9	97.53	(1.72)	56.68	< .001
School 10	97.37	(1.23)	79.17	< .001
School 11	95.48	(1.30)	73.42	< .001
School 12	98.77	(1.62)	60.89	< .001
School 13	98.77	(1.02)	97.05	< .001
School 14	98.42	(1.16)	85.05	< .001
School 15	100.39	(1.85)	54.16	< .001
School 16	97.95	(1.61)	60.75	< .001
School 17	95.71	(1.40)	68.12	< .001
School 18	96.55	(1.22)	79.4	< .001
School 19	98.16	(2.30)	42.65	< .001
School 20	99.04	(1.79)	55.38	< .001
School 21	97.58	(1.27)	77.11	< .001
School 22	98.66	(1.80)	54.70	< .001

Variable	*Estimate*		t-*value*	Pr > |*t*|
School 23	97.33	(1.04)	94.00	< .001
School 24	95.63	(1.75)	54.75	< .001
School 25	96.07	(1.50)	63.84	< .001
School 26	96.70	(1.32)	73.07	< .001
School 1*CSR	4.15	(1.65)	2.51	0.013
School 2*CSR	–2.68	(2.80)	–0.96	0.344
School 3*CSR	–2.12	(1.70)	–1.24	0.214
School 4*CSR	–1.38	(2.70)	–0.51	0.609
School 5*CSR	–1.51	(1.94)	–0.78	0.436
School 6*CSR	–2.72	(2.82)	–0.96	0.342
School 7*CSR	5.00	(2.48)	2.02	0.048
School 8*CSR	–3.80	(1.96)	–1.93	0.054
School 9*CSR	4.95	(2.27)	2.18	0.029
School 10*CSR	2.67	(1.77)	1.51	0.132
School 11*CSR	4.00	(1.81)	2.21	0.027
School 12*CSR	–3.48	(2.33)	–1.50	0.135
School 13*CSR	–0.31	(1.64)	–0.19	0.851
School 14*CSR	0.79	(1.74)	0.45	0.651
School 15*CSR	–1.62	(2.25)	–0.72	0.473
School 16*CSR	0.03	(1.97)	0.01	0.989
School 17*CSR	5.07	(2.12)	2.39	0.017
School 18*CSR	–0.07	(2.12)	–0.03	0.975
School 19*CSR	–0.89	(3.06)	–0.29	0.772
School 20*CSR	–2.38	(2.23)	–1.07	0.285
School 21*CSR	1.00	(2.52)	0.40	0.691
School 22*CSR	1.97	(2.60)	0.76	0.448
School 23*CSR	1.79	(1.37)	1.31	0.189
School 24*CSR	0.69	(2.47)	0.28	0.781
School 25*CSR	1.78	(2.13)	0.84	0.404
School 26*CSR	3.46	(1.99)	1.73	0.083
Pretest GRADE score	0.78	(0.02)	40.64	< .001
Language status	–0.70	(0.42)	–1.68	0.094
Little Spanish	–1.64	(0.81)	–2.03	0.043

Variable	*Estimate*		t-*value*	Pr > |*t*|
Spanish as a second language	–0.72	(0.95)	–0.76	0.450
Spanish as a first language	–0.45	(1.05)	–0.43	0.670
Random effects	*Variance*		*Chi–square*	p–*value*
Intercept (U0)	0.07	(0.26)	15.99	>.50
Level 1 R	38.73	(6.22)		

Note: Figures in parentheses are standard errors for fixed effects and standard deviations for random effects. Results are based on 10 multiply imputed datasets. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,355 students (681 CSR, 674 control). The estimate of the effect of CSR is the unweighted combined estimate of all *School*CSR* interaction terms.
Source: Authors' analysis of student GRADE scores, study records, and teacher surveys.

Table Q-4. Sensitivity analysis 2: random effects

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
Fixed effects				
Intercept	98.27	(0.20)	491.05	< .001
District 1	–0.29	(0.43)	–0.68	0.505
District 2	1.28	(0.40)	3.17	0.005
District 3	0.46	(0.44)	1.12	0.277
District 4	–0.60	(0.44)	–1.37	0.184
CSR	0.80	(0.46)	1.74	0.097
District 1*CSR	–1.35	(0.87)	–1.55	0.137
District 2*CSR	1.27	(0.97)	1.31	0.206
District 3*CSR	–1.26	(0.96)	–1.31	0.205
District 4*CSR	–0.26	(0.99)	–0.27	0.794
Pretest GRADE score	0.78	(0.02)	42.46	< .001
Language status	–0.50	(0.39)	–1.28	0.202
Little Spanish	–0.45	(0.53)	–0.85	0.399
Spanish as a second language	0.18	(0.69)	0.26	0.792
Spanish as a first language	–1.02	(0.68)	–1.49	0.141
Random effects	*Variance*		*Chi–square*	p–*value*
Intercept1 (R_0)	0.04	(0.20)	20.32	0.375
Level 1 E	38.66	(6.22)		
Level 3				
Intercept1/Intercept 2, U_{00}	0.16	(0.40)	32.09	0.06
Intercept1/CSR, U_{04}	1.44	(1.20)	36.33	0.02

Note: Figures in parentheses are standard errors for fixed effects and standard deviations for random effects. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,355 students (681 CSR, 674 control).
Source: Authors' analysis of student GRADE scores, study records, and teacher surveys.

Table Q-5. Sensitivity analysis 3: list-wise deletion

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
Fixed effects				
School 1	96.96	(1.16)	83.36	< .001
School 2	98.40	(1.65)	59.57	< .001
School 3	99.35	(1.23)	80.86	< .001
School 4	100.81	(2.40)	42.03	< .001
School 5	101.13	(1.59)	63.42	< .001
School 6	97.17	(1.27)	76.54	< .001
School 7	99.81	(1.83)	54.66	< .001
School 8	102.99	(1.74)	59.16	< .001
School 9	97.83	(1.63)	60.08	< .001
School 10	97.75	(1.21)	80.58	< .001
School 11	95.89	(1.29)	74.10	< .001
School 12	99.15	(1.65)	60.28	< .001
School 13	99.38	(1.02)	97.71	< .001
School 14	99.04	(1.07)	92.64	< .001
School 15	100.56	(1.82)	55.15	< .001
School 16	98.20	(1.70)	57.94	< .001
School 17	96.27	(1.43)	67.21	< .001
School 18	97.48	(1.25)	77.79	< .001
School 19	98.50	(2.39)	41.14	< .001
School 20	99.17	(1.77)	55.89	< .001
School 21	97.87	(1.22)	80.47	< .001
School 22	99.49	(1.81)	55.00	< .001
School 23	97.81	(1.03)	95.11	< .001
School 24	97.92	(2.13)	45.91	< .001
School 25	97.02	(1.56)	62.29	< .001
School 26	97.71	(1.45)	67.56	< .001
School 1*CSR	3.93	(1.62)	2.43	0.025
School 2*CSR	–2.99	(2.46)	-1.21	0.240
School 3*CSR	–2.42	(1.72)	-1.40	0.177
School 4*CSR	–1.26	(2.75)	-0.46	0.653
School 5*CSR	–1.63	(1.91)	-0.85	0.405

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
School 6*CSR	–2.45	(2.41)	-1.02	0.322
School 7*CSR	2.84	(2.61)	1.09	0.290
School 8*CSR	–3.72	(2.03)	-1.83	0.082
School 9*CSR	4.74	(2.21)	2.14	0.045
School 10*CSR	2.65	(1.73)	1.53	0.143
School 11*CSR	4.06	(1.77)	2.29	0.033
School 12*CSR	–3.43	(2.33)	-1.47	0.158
School 13*CSR	–0.16	(1.67)	-0.10	0.925
School 14*CSR	0.88	(1.51)	0.58	0.569
School 15*CSR	–1.65	(2.20)	-0.75	0.461
School 16*CSR	0.14	(2.04)	0.07	0.947
School 17*CSR	4.72	(2.15)	2.19	0.041
School 18*CSR	–0.66	(2.19)	-0.30	0.766
School 19*CSR	–1.38	(3.20)	-0.43	0.671
School 20*CSR	–2.13	(2.22)	-0.96	0.351
School 21*CSR	1.35	(2.49)	0.54	0.594
School 22*CSR	1.92	(2.45)	0.78	0.443
School 23*CSR	1.57	(1.43)	1.10	0.287
School 24*CSR	–0.66	(2.83)	-0.23	0.817
School 25*CSR	1.24	(2.11)	0.58	0.566
School 26*CSR	3.59	(2.14)	1.68	0.110
Pretest GRADE score	0.78	(0.02)	41.61	< .001
Language status	–0.82	(0.45)	–1.82	0.069
Little Spanish	–1.48	(0.84)	–1.77	0.093
Spanish as a second language	–0.31	(0.96)	–0.32	0.749
Spanish as a first language	–0.29	(1.00)	–0.29	0.774
Random effects	*Variance*		*Chi–square*	p–*value*
Intercept (U_0)	0.025	(0.159)	11.61	> 0.500
Level 1 R	38.006	(6.165)		

Note: Figures in parentheses are standard errors for fixed effects and standard deviations for random effects. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,123 students (580 CSR, 543 control). The estimate of the effect of CSR is the weighted combined estimate of all *School*CSR* interaction terms (see appendix E).
Source: Authors' analysis of student GRADE scores, study records, and teacher surveys.

Table Q-6. Sensitivity analysis 4: exclusion of crossover students

Variable	*Estimate*		t-*value*	*Pr > |t|*
Fixed effects				
School 1	96.42	(1.21)	79.99	< .001
School 2	97.76	(1.68)	58.19	< .001
School 3	99.01	(1.23)	80.78	< .001
School 4	100.39	(2.35)	42.74	< .001
School 5	100.60	(1.58)	63.59	< .001
School 6	96.94	(1.34)	72.28	< .001
School 7	98.29	(1.63)	60.21	< .001
School 8	103.22	(1.75)	59.14	< .001
School 9	97.60	(1.72)	56.70	< .001
School 10	97.37	(1.23)	79.20	< .001
School 11	95.39	(1.30)	73.32	< .001
School 12	98.80	(1.62)	61.03	< .001
School 13	98.74	(1.02)	96.86	< .001
School 14	98.68	(1.18)	83.91	< .001
School 15	100.32	(1.85)	54.11	< .001
School 16	97.89	(1.61)	60.71	< .001
School 17	95.94	(1.44)	66.80	< .001
School 18	96.45	(1.22)	79.37	< .001
School 19	98.19	(2.30)	42.64	< .001
School 20	99.05	(1.79)	55.31	< .001
School 21	97.46	(1.27)	77.01	< .001
School 22	98.53	(1.87)	52.75	< .001
School 23	97.35	(1.04)	93.94	< .001
School 24	95.68	(1.75)	54.81	< .001
School 25	96.11	(1.51)	63.78	< .001
School 26	96.48	(1.35)	71.23	< .001
School 1*CSR	4.32	(1.67)	2.59	0.010
School 2*CSR	–2.62	(2.64)	–0.99	0.323
School 3*CSR	–2.10	(1.70)	–1.23	0.218
School 4*CSR	–1.45	(2.70)	–0.54	0.591

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
School 5*CSR	–1.54	(1.94)	–0.79	0.427
School 6*CSR	–2.49	(2.94)	–0.85	0.402
School 7*CSR	5.10	(2.49)	2.05	0.045
School 8*CSR	–4.12	(2.01)	–2.05	0.041
School 9*CSR	4.85	(2.27)	2.14	0.033
School 10*CSR	2.59	(1.77)	1.47	0.143
School 11*CSR	3.93	(1.81)	2.18	0.030
School 12*CSR	–3.62	(2.33)	–1.55	0.121
School 13*CSR	–0.36	(1.65)	–0.22	0.826
School 14*CSR	0.45	(1.76)	0.25	0.800
School 15*CSR	–1.48	(2.26)	–0.66	0.512
School 16*CSR	–0.12	(1.98)	–0.06	0.950
School 17*CSR	4.71	(2.18)	2.16	0.031
School 18*CSR	0.17	(2.12)	0.08	0.935
School 19*CSR	–0.91	(3.06)	–0.30	0.767
School 20*CSR	–2.42	(2.23)	–1.09	0.277
School 21*CSR	1.07	(2.51)	0.42	0.671
School 22*CSR	2.06	(2.65)	0.78	0.438
School 23*CSR	2.09	(1.37)	1.52	0.128
School 24*CSR	0.73	(2.47)	0.30	0.767
School 25*CSR	1.79	(2.16)	0.83	0.408
School 26*CSR	4.27	(2.05)	2.09	0.037
Pretest GRADE score	0.78	(0.02)	40.54	< .001
Language status	–0.76	(0.42)	–1.81	0.071
Little Spanish	–1.75	(0.81)	–2.14	0.033
Spanish as a second language	–0.75	(0.95)	–0.79	0.430
Spanish as a first language	–0.60	(1.05)	–0.57	0.570
Random effects	*Variance*		*Chi–square*	p–*value*
Intercept (U_0)	0.08	(0.28)	16.19	>.500
Level 1 R	38.58	(6.21)		

Note: Figures in parentheses are standards for fixed effects and standard deviations for random effects. Results are based on 10 multiply imputed datasets. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,341 students (673 CSR, 668 control). The estimate of the effect of CSR is the weighted combined estimate of all *School*CSR* interaction terms (see appendix E).
Source: Authors' analysis of student GRADE scores, study records, and teacher surveys.

Table Q-7. Sensitivity analysis 5: exclusion of schools with increased risk of contamination

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
Fixed effects				
School 1	96.74	(1.25)	77.61	< .001
School 2	98.04	(1.83)	53.55	< .001
School 3	99.30	(1.28)	77.83	< .001
School 4	100.68	(2.37)	42.54	< .001
School 5	100.87	(1.60)	62.89	< .001
School 6	97.19	(1.36)	71.7	< .001
School 7	98.66	(1.64)	60.17	< .001
School 8	103.24	(1.71)	60.43	< .001
School 9	97.86	(1.80)	54.36	< .001
School 10	97.69	(1.25)	77.88	< .001
School 11	95.77	(1.31)	72.86	< .001
School 12	99.08	(1.64)	60.35	< .001
School 13	99.06	(1.02)	96.76	< .001
School 14	98.74	(1.18)	83.7	< .001
School 16	98.27	(1.63)	60.31	< .001
School 17	96.00	(1.42)	67.6	< .001
School 18	96.85	(1.23)	79.02	< .001
School 20	99.29	(1.81)	54.72	< .001
School 21	97.86	(1.28)	76.71	< .001
School 22	98.98	(1.83)	54.22	< .001
School 23	97.64	(1.05)	93.06	< .001
School 24	95.92	(1.77)	54.26	< .001
School 25	96.38	(1.52)	63.58	< .001
School 26	97.01	(1.34)	72.55	< .001
School 1*CSR	4.12	(1.69)	2.43	0.016
School 2*CSR	–2.67	(2.86)	–0.93	0.355
School 3*CSR	–2.11	(1.72)	–1.22	0.222
School 4*CSR	–1.39	(2.73)	–0.51	0.611
School 5*CSR	–1.52	(1.97)	–0.77	0.441
School 6*CSR	–2.72	(2.85)	–0.96	0.344
School 7*CSR	4.99	(2.50)	1.99	0.051

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
School 8*CSR	–3.81	(1.99)	–1.92	0.056
School 9*CSR	4.89	(2.42)	2.02	0.044
School 10*CSR	2.64	(1.81)	1.46	0.145
School 11*CSR	3.95	(1.83)	2.16	0.031
School 12*CSR	–3.49	(2.36)	–1.48	0.139
School 13*CSR	–0.28	(1.66)	–0.17	0.865
School 14*CSR	0.78	(1.76)	0.45	0.657
School 16*CSR	0.04	(1.99)	0.02	0.985
School 17*CSR	5.09	(2.15)	2.37	0.018
School 18*CSR	–0.02	(2.17)	–0.01	0.994
School 20*CSR	–2.35	(2.26)	–1.04	0.299
School 21*CSR	1.03	(2.54)	0.41	0.685
School 22*CSR	1.96	(2.62)	0.75	0.455
School 23*CSR	1.80	(1.38)	1.30	0.194
School 24*CSR	0.72	(2.56)	0.28	0.778
School 25*CSR	1.79	(2.16)	0.83	0.409
School 26*CSR	3.45	(2.02)	1.71	0.088
Pretest GRADE score	0.78	(0.02)	39.76	< .001
Language status	–0.69	(0.44)	–1.59	0.113
Little Spanish	–1.62	(0.82)	–1.98	0.048
Spanish as a second language	–0.68	(1.00)	–0.68	0.496
Spanish as a first language	–0.46	(1.13)	–0.41	0.682
Random effects	*Variance*		*Chi–square*	p–*value*
Intercept (U_0)	0.09	(0.29)	15.53	>.500
Level 1 R	39.59	(6.29)		

Note: Figures in parentheses are standard errors for fixed effects and standard deviations for random effects. Results are based on 10 multiply imputed datasets. Analysis included 24 schools, 69 teachers (34 CSR, 35 control), and 1,282 students (631 CSR, 651 control). The estimate of the effect of CSR is the weighted combined estimate of all *School*CSR* interaction terms (see appendix E).
Source: Authors' analysis of student GRADE scores, study records, and teacher surveys.

Table Q-8. Sensitivity analysis 6: unconditional treatment effect

Variable	*Estimate*		t-*value*	Pr > |*t*|
Fixed effects				
School 1	97.39	(3.11)	31.30	< .001
School 2	97.26	(4.42)	22.01	< .001
School 3	97.53	(3.16)	30.86	< .001
School 4	101.61	(5.02)	20.24	< .001
School 5	101.81	(4.41)	23.09	< .001
School 6	93.21	(3.25)	28.64	< .001
School 7	104.29	(4.24)	24.62	< .001
School 8	104.09	(4.38)	23.76	< .001
School 9	98.49	(4.35)	22.66	< .001
School 10	96.11	(3.22)	29.84	< .001
School 11	102.11	(4.20)	24.32	< .001
School 12	93.86	(4.43)	21.21	< .001
School 13	100.12	(2.99)	33.53	< .001
School 14	97.29	(3.05)	31.91	< .001
School 15	101.55	(4.63)	21.94	< .001
School 16	96.42	(4.45)	21.66	< .001
School 17	101.07	(4.29)	23.59	< .001
School 18	100.48	(3.16)	31.82	< .001
School 19	94.13	(5.05)	18.64	< .001
School 20	107.66	(4.55)	23.69	< .001
School 21	105.12	(3.22)	32.67	< .001
School 22	89.82	(4.52)	19.87	< .001
School 23	94.15	(2.59)	36.36	< .001
School 24	93.65	(4.44)	21.14	< .001
School 25	89.20	(4.26)	20.92	< .001
School 26	92.52	(3.33)	27.82	< .001
School 1*CSR	4.34	(4.42)	0.98	0.326
School 2*CSR	–2.60	(6.23)	–0.42	0.677
School 3*CSR	–5.52	(4.55)	–1.21	0.226
School 4*CSR	0.21	(5.97)	0.04	0.972
School 5*CSR	–2.40	(5.38)	–0.45	0.656
School 6*CSR	0.78	(5.84)	0.13	0.893
School 7*CSR	2.44	(6.00)	0.41	0.685
School 8*CSR	4.01	(5.38)	0.75	0.456
School 9*CSR	4.46	(6.10)	0.73	0.465
School 10*CSR	8.03	(4.57)	1.76	0.079
School 11*CSR	17.21	(5.95)	2.89	0.004
School 12*CSR	4.20	(6.30)	0.67	0.505

Variable	*Estimate*		t-*value*	Pr > |*t*|
School 13*CSR	–3.31	(5.13)	–0.64	0.520
School 14*CSR	1.84	(4.35)	0.42	0.672
School 15*CSR	–9.76	(5.56)	–1.76	0.079
School 16*CSR	–7.48	(5.47)	–1.37	0.172
School 17*CSR	–0.60	(6.20)	–0.10	0.923
School 18*CSR	–10.66	(5.44)	–1.96	0.051
School 19*CSR	–5.58	(6.96)	–0.80	0.423
School 20*CSR	–12.64	(5.59)	–2.26	0.024
School 21*CSR	–8.06	(5.94)	–1.36	0.175
School 22*CSR	10.61	(6.35)	1.67	0.095
School 23*CSR	–0.90	(4.00)	–0.22	0.822
School 24*CSR	1.96	(6.20)	0.32	0.752
School 25*CSR	7.53	(5.97)	1.26	0.208
School 26*CSR	3.43	(5.52)	0.62	0.534

Random effects	*Variance*		*Degrees of freedom*	*Chi–square*	p-*value*
Intercept (U_0)	13.53	(3.68)	22	71.77	0.00
Level 1 R	103.75	(10.19)			

Note: Figures in parentheses are standard errors for fixed effects and standard deviations for random effects. Results are based on 10 multiply imputed datasets. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,355 students (681 CSR, 674 control). The estimate of the effect of CSR is the combined weighted estimate of all *School*CSR* interaction terms.
Source: Authors' analysis of student GRADE scores.

Table Q-9. Sensitivity analysis 7: heterogeneity of slopes (using 26 schools)

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
Fixed effects				
Intercept	98.199	(0.340)	288.61	< .001
School 1	–1.882	(1.290)	–1.46	0.146
School 2	–0.715	(2.032)	–0.35	0.727
School 3	0.759	(1.260)	0.60	0.547
School 4	3.228	(2.433)	1.33	0.187
School 5	2.236	(1.628)	1.37	0.170
School 6	–1.337	(1.381)	–0.97	0.333
School 7	0.736	(1.874)	0.39	0.695
School 8	5.017	(1.836)	2.73	0.007
School 9	–0.759	(1.817)	–0.42	0.677
School 10	–1.003	(1.307)	–0.77	0.443
School 11	–2.147	(1.568)	–1.37	0.171
School 12	1.586	(1.835)	0.86	0.388
School 13	0.563	(1.076)	0.52	0.601
School 14	0.003	(1.248)	0.00	0.998
School 15	2.113	(1.870)	1.13	0.259
School 16	–0.402	(1.681)	–0.24	0.811
School 17	–2.635	(1.580)	–1.67	0.096
School 18	–1.812	(1.405)	–1.29	0.200
School 19	–0.833	(2.318)	–0.36	0.720
School 20	1.157	(2.577)	0.45	0.655
School 21	–0.221	(1.805)	–0.12	0.903
School 22	0.561	(2.623)	0.21	0.831
School 23	–0.840	(1.069)	–0.79	0.433
School 24	–2.335	(1.768)	–1.32	0.188
School 25	–0.279	(1.792)	–0.16	0.877
School 1*CSR	4.119	(1.705)	2.42	0.017
School 2*CSR	–2.647	(3.009)	–0.88	0.384
School 3*CSR	–1.774	(1.904)	–0.93	0.352
School 4*CSR	–3.217	(2.846)	–1.13	0.261
School 5*CSR	–1.496	(1.990)	–0.75	0.453

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
Fixed effects				
School 6*CSR	–2.716	(2.817)	–0.96	0.340
School 7*CSR	4.961	(2.829)	1.75	0.083
School 8*CSR	–3.942	(2.351)	–1.68	0.095
School 9*CSR	5.007	(2.337)	2.14	0.033
School 10*CSR	2.200	(1.888)	1.17	0.245
School 11*CSR	4.126	(3.192)	1.29	0.196
School 12*CSR	–4.793	(2.509)	–1.91	0.056
School 13*CSR	–0.642	(1.703)	–0.38	0.707
School 14*CSR	0.692	(1.845)	0.38	0.709
School 15*CSR	–1.807	(2.489)	–0.73	0.468
School 16*CSR	–1.717	(2.199)	–0.78	0.435
School 17*CSR	4.689	(2.223)	2.11	0.035
School 18*CSR	–3.050	(2.618)	–1.17	0.244
School 19*CSR	–1.582	(3.457)	–0.46	0.647
School 20*CSR	–3.022	(2.905)	–1.04	0.301
School 21*CSR	0.068	(2.904)	0.02	0.981
School 22*CSR	1.772	(3.239)	0.55	0.585
School 23*CSR	1.361	(1.587)	0.86	0.391
School 24*CSR	0.278	(2.502)	0.11	0.912
School 25*CSR	0.005	(2.479)	0.00	0.998
CSR	0.155	(0.503)	0.31	0.758
CSR*Pretest GRADE score	0.014	(0.047)	0.3	0.765
Pretest GRADE score	0.762	(0.035)	21.71	< .001
Language status	–0.717	(0.428)	–1.67	0.094
Little Spanish	–1.863	(0.900)	–2.07	0.039
Spanish as a second language	–0.806	(0.983)	–0.82	0.413
Spanish as a first language	–0.388	(1.130)	–0.34	0.732
Little Spanish*Pretest GRADE score	–0.076	(0.072)	–1.05	0.293
Spanish as a second language*Pretest GRADE score	–0.144	(0.098)	–1.47	0.142
Spanish as a first language*Pretest GRADE score	–0.032	(0.104)	–0.31	0.759
School 1*Pretest GRADE score	0.043	(0.130)	0.33	0.744

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
Fixed effects				
School 2*Pretest GRADE score	–0.120	(0.236)	–0.51	0.614
School 3*Pretest GRADE score	0.012	(0.151)	0.08	0.938
School 4*Pretest GRADE score	–0.317	(0.232)	–1.37	0.172
School 5*Pretest GRADE score	0.157	(0.158)	1.00	0.320
School 6*Pretest GRADE score	0.027	(0.118)	0.23	0.821
School 7*Pretest GRADE score	–0.120	(0.151)	–0.79	0.429
School 8*Pretest GRADE score	0.006	(0.202)	0.03	0.975
School 9*Pretest GRADE score	–0.042	(0.182)	–0.23	0.817
School 10*Pretest GRADE score	–0.004	(0.110)	–0.04	0.972
School 11*Pretest GRADE score	–0.114	(0.116)	–0.98	0.329
School 12*Pretest GRADE score	0.205	(0.175)	1.17	0.244
School 13*Pretest GRADE score	0.052	(0.092)	0.57	0.571
School 14*Pretest GRADE score	–0.022	(0.114)	–0.20	0.844
School 15*Pretest GRADE score	–0.022	(0.206)	–0.11	0.915
School 16*Pretest GRADE score	–0.073	(0.171)	–0.43	0.671
School 17*Pretest GRADE score	–0.019	(0.114)	–0.17	0.868
School 18*Pretest GRADE score	–0.017	(0.094)	–0.19	0.853
School 19*Pretest GRADE score	–0.182	(0.213)	–0.85	0.394
School 20*Pretest GRADE score	0.033	(0.140)	0.23	0.816
School 21*Pretest GRADE score	–0.080	(0.132)	–0.61	0.543
School 22*Pretest GRADE score	0.039	(0.189)	0.21	0.835
School 23*Pretest GRADE score	0.036	(0.083)	0.43	0.668
School 24*Pretest GRADE score	0.104	(0.138)	0.75	0.451
School 25*Pretest GRADE score	0.266	(0.154)	1.72	0.094
School 1*CSR*Pretest GRADE score	–0.035	(0.166)	–0.21	0.832
School 2*CSR*Pretest GRADE score	0.228	(0.282)	0.81	0.420
School 3*CSR*Pretest GRADE score	0.115	(0.192)	0.60	0.549
School 4*CSR*Pretest GRADE score	0.501	(0.284)	1.76	0.079
School 5*CSR*Pretest GRADE score	–0.030	(0.227)	–0.13	0.895
School 6*CSR*Pretest GRADE score	0.128	(0.258)	0.49	0.622
School 7*CSR*Pretest GRADE score	0.053	(0.224)	0.24	0.813
School 8*CSR*Pretest GRADE score	–0.021	(0.219)	–0.09	0.925

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
Fixed effects				
School 9*CSR*Pretest GRADE score	0.028	(0.230)	0.12	0.902
School 10*CSR*Pretest GRADE score	0.156	(0.175)	0.89	0.372
School 11*CSR*Pretest GRADE score	0.031	(0.154)	0.20	0.839
School 12*CSR*Pretest GRADE score	–0.222	(0.244)	–0.91	0.362
School 13*CSR*Pretest GRADE score	–0.129	(0.140)	–0.92	0.357
School 14*CSR*Pretest GRADE score	–0.110	(0.161)	–0.69	0.492
School 15*CSR*Pretest GRADE score	0.058	(0.248)	0.23	0.815
School 16*CSR*Pretest GRADE score	–0.066	(0.211)	–0.31	0.755
School 17*CSR*Pretest GRADE score	–0.216	(0.223)	–0.97	0.333
School 18*CSR*Pretest GRADE score	–0.341	(0.212)	–1.61	0.109
School 19*CSR*Pretest GRADE score	0.157	(0.275)	0.57	0.568
School 20*CSR*Pretest GRADE score	0.006	(0.210)	0.03	0.978
School 21*CSR*Pretest GRADE score	–0.086	(0.298)	–0.29	0.773
School 22*CSR*Pretest GRADE score	0.088	(0.246)	0.36	0.722
School 23*CSR*Pretest GRADE score	–0.070	(0.126)	–0.56	0.576
School 24*CSR*Pretest GRADE score	–0.016	(0.209)	–0.08	0.938
School 25*CSR*Pretest GRADE score	–0.189	(0.213)	–0.89	0.379
Random effects	*Variance*		*Chi–square*	p–*value*
Intercept (U_0)	0.08	(0.29)	14.73	>.500
Level 1 R	38.93	(6.24)		

Note: Figures in parentheses are standard errors for fixed effects and standard deviations for random effects. Analysis included 26 schools, 74 teachers (37 CSR, 37 control), and 1,355 students (681 FC–ELL CSR, 674 in FC–ELL control).
Source: Authors' analysis of student GRADE scores, study records, and teacher surveys.

Table Q-10. Sensitivity analysis 8: heterogeneity of slopes (using 23 schools)

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
Fixed effects				
Intercept	98.08	(0.36)	275.77	< .001
School 1	–1.76	(1.26)	–1.40	0.163
School 3	0.83	(1.25)	0.66	0.508
School 5	2.25	(1.60)	1.40	0.161
School 6	–1.37	(1.35)	–1.01	0.312
School 7	0.79	(1.87)	0.42	0.673
School 8	5.02	(1.79)	2.81	0.006
School 9	–0.77	(1.76)	–0.44	0.661
School 10	–0.96	(1.29)	–0.75	0.455
School 11	–2.07	(1.55)	–1.33	0.183
School 12	1.62	(1.80)	0.90	0.368
School 13	0.63	(1.08)	0.58	0.561
School 14	0.04	(1.22)	0.03	0.977
School 15	2.18	(1.87)	1.16	0.246
School 16	–0.34	(1.65)	–0.20	0.838
School 17	–2.55	(1.57)	–1.62	0.105
School 19	–0.79	(2.29)	–0.34	0.731
School 20	1.11	(2.54)	0.44	0.664
School 21	–0.11	(1.82)	–0.06	0.951
School 22	0.52	(2.59)	0.20	0.840
School 23	–0.84	(1.04)	–0.81	0.420
School 24	–2.36	(1.75)	–1.35	0.180
School 25	–0.28	(1.75)	–0.16	0.874
School 1*CSR	3.59	(1.73)	2.07	0.039
School 3*CSR	–2.21	(1.89)	–1.17	0.244
School 5*CSR	–1.77	(1.96)	–0.90	0.367
School 6*CSR	–3.02	(2.55)	–1.19	0.238
School 7*CSR	4.58	(2.85)	1.61	0.112
School 8*CSR	–4.32	(2.35)	–1.84	0.067
School 9*CSR	4.49	(2.28)	1.97	0.049
School 10*CSR	1.78	(1.84)	0.97	0.334

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
School 11*CSR	3.76	(3.17)	1.19	0.236
School 12*CSR	–5.07	(2.46)	–2.06	0.040
School 13*CSR	–0.95	(1.66)	–0.57	0.568
School 14*CSR	0.39	(1.77)	0.22	0.826
School 15*CSR	–2.40	(2.52)	–0.95	0.341
School 16*CSR	–2.09	(2.16)	–0.96	0.335
School 17*CSR	4.37	(2.18)	2.00	0.046
School 19*CSR	–2.15	(3.44)	–0.62	0.533
School 20*CSR	–3.40	(2.85)	–1.19	0.236
School 21*CSR	–0.34	(2.85)	–0.12	0.904
School 22*CSR	1.46	(3.20)	0.46	0.649
School 23*CSR	0.99	(1.56)	0.63	0.526
School 24*CSR	–0.11	(2.52)	–0.05	0.964
School 25*CSR	–0.25	(2.40)	–0.11	0.916
CSR	0.51	(0.53)	0.97	0.333
CSR*Pretest GRADE scores	0.00	(0.05)	0.05	0.958
Pretest GRADE scores	0.78	(0.03)	22.88	< .001
Language status	–0.59	(0.45)	–1.32	0.187
Little Spanish	–1.79	(0.90)	–1.99	0.048
Spanish as a second language	–0.58	(1.07)	–0.54	0.587
Spanish as a first language	–0.29	(1.13)	–0.25	0.799
Little Spanish*Pretest GRADE score	–0.07	(0.07)	–0.94	0.349
Spanish as a second language*Pretest GRADE score	–0.14	(0.11)	–1.28	0.202
Spanish as a first language*Pretest GRADE score	–0.03	(0.11)	–0.25	0.801
School 1*Pretest GRADE score	0.02	(0.14)	0.15	0.878
School 3*Pretest GRADE score	–0.07	(0.14)	–0.48	0.635
School 5*Pretest GRADE score	0.13	(0.16)	0.85	0.393
School 6*Pretest GRADE score	0.01	(0.12)	0.06	0.949
School 7*Pretest GRADE score	–0.13	(0.15)	–0.88	0.382
School 8*Pretest GRADE score	–0.01	(0.20)	–0.07	0.942
School 9*Pretest GRADE score	–0.06	(0.19)	–0.32	0.751
School 10*Pretest GRADE score	–0.01	(0.11)	–0.12	0.905

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
School 11*Pretest GRADE score	–0.13	(0.12)	–1.09	0.277
School 12*Pretest GRADE score	0.18	(0.17)	1.05	0.294
School 13*Pretest GRADE score	0.03	(0.09)	0.36	0.717
School 14*Pretest GRADE score	–0.03	(0.11)	–0.28	0.777
School 15*Pretest GRADE score	–0.03	(0.21)	–0.16	0.873
School 16*Pretest GRADE score	–0.09	(0.17)	–0.50	0.622
School 17*Pretest GRADE score	–0.03	(0.11)	–0.28	0.778
School 19*Pretest GRADE score	–0.20	(0.21)	–0.96	0.340
School 20*Pretest GRADE score	0.01	(0.14)	0.09	0.930
School 21*Pretest GRADE score	–0.09	(0.13)	–0.71	0.480
School 22*Pretest GRADE score	0.02	(0.19)	0.09	0.932
School 23*Pretest GRADE score	0.02	(0.08)	0.23	0.817
School 24*Pretest GRADE score	0.08	(0.14)	0.62	0.537
School 25*Pretest GRADE score	0.24	(0.15)	1.62	0.113
School 1*CSR*Pretest GRADE score	–0.04	(0.17)	–0.2	0.838
School 3*CSR*Pretest GRADE score	0.19	(0.19)	1.0	0.317
School 5*CSR*Pretest GRADE score	–0.02	(0.23)	–0.10	0.924
School 6*CSR*Pretest GRADE score	0.15	0.25)	0.60	0.552
School 7*CSR*Pretest GRADE score	0.05	(0.23)	0.24	0.810
School 8*CSR*Pretest GRADE score	0.00	(0.22)	–0.02	0.984
School 9*CSR*Pretest GRADE score	0.04	(0.23)	0.16	0.873
School 10*CSR*Pretest GRADE score	0.16	(0.17)	0.92	0.358
School 11*CSR*Pretest GRADE score	0.04	(0.15)	0.29	0.775
School 12*CSR*Pretest GRADE score	–0.20	(0.24)	–0.84	0.401
School 13*CSR*Pretest GRADE score	–0.11	(0.14)	–0.80	0.423
School 14*CSR*Pretest GRADE score	–0.10	(0.16)	–0.63	0.530
School 15*CSR*Pretest GRADE score	0.06	(0.25)	0.24	0.813
School 16*CSR*Pretest GRADE score	–0.06	(0.21)	–0.27	0.789
School 17*CSR*Pretest GRADE score	–0.21	(0.22)	–0.93	0.355
School 19*CSR*Pretest GRADE score	0.17	(0.28)	0.61	0.540
School 20*CSR*Pretest GRADE score	0.02	(0.21)	0.11	0.909
School 21*CSR*Pretest GRADE score	–0.07	(0.30)	–0.24	0.809
School 22*CSR*Pretest GRADE score	0.11	(0.25)	0.44	0.661

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
School 23*CSR*Pretest GRADE score	–0.06	(0.13)	–0.51	0.614
School 24*CSR*Pretest GRADE score	0.00	(0.21)	0.00	0.999
School 25*CSR*Pretest GRADE score	–0.17	(0.21)	–0.81	0.424
Random effects	*Variance*		*Chi–square*	p–*value*
Intercept (U_0)	0.07	(0.27)	12.523	>.500
Level 1 R	39.18	(6.26)		

Note: Figures in parentheses are standard errors for fixed effects and standard deviations for random effects. Analysis included 23 schools, 66 teachers (33 in CSR and 33 in control condition), and 1,199 students (279 FC–ELL CSR, 256 FC–ELL control; 322 non–ELL CSR, 342 non–ELL control).
Source: Authors' analysis of student GRADE scores, study records, and teacher surveys.

Table Q-11. Sensitivity analysis 9: cohort effect

Variable	*Estimate*		t-*value*	Pr > |*t*|
Fixed effects				
Intercept	98.36	(0.51)	194.21	< .001
Cohort 1	–1.49	(1.72)	–0.87	0.387
Cohort 1*CSR	–4.25	(3.60)	–1.18	0.245
CSR	1.53	(0.92)	1.67	0.100
Pretest GRADE score	0.78	(0.02)	40.64	< .001
Language status	–0.70	(0.42)	–1.68	0.094
Little Spanish	–1.64	(0.81)	–2.03	0.043
Spanish as a second language	–0.72	(0.95)	–0.76	0.450
Spanish as a first language	–0.45	(1.05)	–0.43	0.670
School 1	–0.45	(1.74)	–0.26	0.796
School 2	0.85	(2.23)	0.38	0.703
School 3	2.15	(1.74)	1.23	0.219
School 4	3.52	(2.71)	1.30	0.196
School 5	3.72	(2.05)	1.81	0.071
School 7	0.01	(1.59)	0.01	0.995
School 8	4.60	(1.67)	2.75	0.007
School 9	–0.83	(1.82)	–0.46	0.649
School 10	–0.99	(1.29)	–0.77	0.443
School 11	–2.88	(1.31)	–2.19	0.029
School 12	0.41	(1.59)	0.26	0.796
School 13	0.41	(1.10)	0.37	0.712
School 14	0.06	(1.30)	0.04	0.966
School 15	2.03	(1.81)	1.12	0.263
School 16	–0.41	(1.60)	–0.25	0.799
School 17	–2.65	(1.40)	–1.89	0.059
School 18	–1.81	(1.22)	–1.48	0.139
School 19	–0.20	(2.27)	–0.09	0.928
School 20	0.68	(1.78)	0.38	0.703
School 21	–0.78	(1.29)	–0.61	0.544
School 22	0.30	(1.75)	0.17	0.863
School 23	–1.03	(1.12)	–0.92	0.358

Variable	*Estimate*		t-*value*	Pr > |*t*|
School 24	–2.73	(1.72)	–1.58	0.114
School 25	–2.29	(1.49)	–1.54	0.125
School 1*CSR	6.86	(3.37)	2.04	0.047
School 2*CSR	0.04	(4.24)	0.01	0.993
School 3*CSR	0.60	(3.35)	0.18	0.860
School 4*CSR	1.33	(3.74)	0.36	0.723
School 5*CSR	1.20	(3.44)	0.35	0.728
School 7*CSR	3.47	(2.47)	1.40	0.164
School 8*CSR	–5.33	(2.08)	–2.57	0.011
School 9*CSR	3.41	(2.43)	1.40	0.162
School 10*CSR	1.14	(1.93)	0.59	0.556
School 11*CSR	2.46	(1.90)	1.30	0.196
School 12*CSR	–5.02	(2.39)	–2.10	0.036
School 13*CSR	–1.84	(1.92)	–0.96	0.339
School 14*CSR	–0.74	(2.02)	–0.37	0.715
School 15*CSR	–3.15	(2.17)	–1.45	0.148
School 16*CSR	–1.51	(2.05)	–0.73	0.463
School 17*CSR	3.53	(2.16)	1.63	0.103
School 18*CSR	–1.60	(2.06)	–0.78	0.438
School 19*CSR	–2.42	(3.03)	–0.80	0.425
School 20*CSR	–3.92	(2.23)	–1.76	0.080
School 21*CSR	–0.53	(2.62)	–0.20	0.839
School 22*CSR	0.44	(2.54)	0.17	0.863
School 23*CSR	0.26	(1.67)	0.16	0.876
School 24*CSR	–0.85	(2.54)	–0.33	0.740
School 25*CSR	0.25	(2.29)	0.11	0.914
Random effects	*Variance*		*Chi–square*	p–*value*
Intercept (U_0)	0.07	(0.26)	15.99	>.50
Level 1 R	38.73	(6.22)		

Note: Figures in parentheses are standard errors for fixed effects and standard deviations for random effects. Analysis included 26 schools, 74 teachers, and 1,355 students (681 in CSR, 674 control).
Source: Authors' analysis of student GRADE scores, study records, and teacher surveys.

Exploratory analysis

Table Q-12. Exploratory analysis: former and current English language learner and non–English language learner subgroup results

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
Fixed effects				
School 1	96.17	(1.45)	66.26	< .001
School 3	98.73	(1.27)	77.74	< .001
School 5	101.50	(1.67)	60.61	< .001
School 6	97.00	(1.39)	69.79	< .001
School 7	98.50	(1.76)	56.04	< .001
School 8	102.69	(1.69)	60.79	< .001
School 9	97.79	(1.82)	53.69	< .001
School 10	97.83	(1.31)	74.69	< .001
School 11	95.64	(1.32)	72.36	< .001
School 12	97.80	(4.86)	20.13	< .001
School 13	98.54	(1.53)	64.34	< .001
School 14	98.84	(1.20)	82.35	< .001
School 15	100.57	(1.90)	52.80	< .001
School 16	98.15	(1.64)	59.76	< .001
School 17	96.56	(1.46)	66.04	< .001
School 19	98.72	(2.49)	39.60	< .001
School 20	100.03	(1.92)	52.03	< .001
School 21	97.86	(1.45)	67.55	< .001
School 22	97.99	(1.90)	51.49	< .001
School 23	97.34	(1.26)	77.14	< .001
School 24	95.39	(1.82)	52.39	< .001
School 25	95.74	(1.59)	60.07	< .001
School 26	94.69	(1.62)	58.32	< .001
School 1*CSR	3.82	(1.99)	1.91	0.06
School 3*CSR	–1.94	(1.79)	–1.08	0.28
School 5*CSR	–2.42	(2.03)	–1.19	0.24
School 6*CSR	–2.51	(2.55)	–0.99	0.33
School 7*CSR	4.59	(2.65)	1.73	0.09

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
School 8*CSR	–3.59	(2.08)	–1.72	0.09
School 9*CSR	4.14	(2.49)	1.66	0.10
School 10*CSR	2.12	(1.85)	1.15	0.25
School 11*CSR	3.50	(2.04)	1.72	0.09
School 12*CSR	–3.25	(5.69)	–0.57	0.57
School 13*CSR	1.61	(2.70)	0.60	0.55
School 14*CSR	0.58	(1.73)	0.34	0.74
School 15*CSR	–0.27	(2.93)	–0.09	0.93
School 16*CSR	–0.95	(2.30)	–0.41	0.68
School 17*CSR	5.20	(2.59)	2.01	0.05
School 19*CSR	–3.75	(3.71)	–1.01	0.31
School 20*CSR	–3.70	(2.62)	–1.41	0.16
School 21*CSR	1.58	(3.32)	0.48	0.64
School 22*CSR	0.23	(3.00)	0.08	0.94
School 23*CSR	2.16	(1.57)	1.38	0.17
School 24*CSR	1.24	(2.74)	0.45	0.65
School 25*CSR	2.14	(2.60)	0.82	0.41
School 26*CSR	5.35	(2.44)	2.20	0.03
Pretest GRADE score	0.77	(0.02)	39.17	< .001
Little Spanish	–1.25	(0.89)	–1.41	0.160
Spanish as a second language	0.10	(1.14)	0.09	0.929
Spanish as a first language	–0.31	(1.22)	–0.25	0.799
Language status*Little Spanish	–0.69	(1.84)	–0.37	0.709
Language status*Spanish as a second language	–3.66	(2.41)	–1.52	0.131
Language status*Spanish as a first language	–2.11	(2.40)	–0.88	0.382
School 1*Language status	–1.95	(3.16)	–0.62	0.537
School 3*Language status	–1.43	(2.75)	–0.52	0.604
School 5*Language status	5.83	(3.64)	1.60	0.110
School 6*Language status	1.00	(2.88)	0.35	0.729
School 7*Language status	–1.37	(3.38)	–0.41	0.686
School 8*Language status	–2.58	(3.46)	–0.75	0.456
School 9*Language status	3.02	(3.85)	0.79	0.432

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
School 10*Language status	–1.54	(2.66)	–0.58	0.561
School 11*Language status	–2.91	(2.92)	–1.00	0.320
School 12*Language status	–2.71	(11.16)	–0.24	0.811
School 13*Language status	–1.68	(3.21)	–0.52	0.602
School 14*Language status	–1.25	(2.43)	–0.52	0.607
School 15*Language status	–2.35	(3.88)	–0.60	0.546
School 16*Language status	–2.52	(3.48)	–0.73	0.468
School 17*Language status	3.69	(3.22)	1.15	0.252
School 19*Language status	2.80	(5.35)	0.52	0.601
School 20*Language status	–6.23	(3.77)	–1.65	0.099
School 21*Language status	–1.30	(3.15)	–0.41	0.679
School 22*Language status	1.62	(3.53)	0.46	0.647
School 23*Language status	–1.68	(2.32)	–0.73	0.469
School 24*Language status	–1.10	(4.01)	–0.27	0.784
School 25*Language status	–1.11	(3.55)	–0.31	0.755
School 26*Language status	5.31	(3.33)	1.59	0.114
School 1*CSR*Language status	4.45	(4.15)	1.07	0.284
School 3*CSR*Language status	1.88	(3.92)	0.48	0.634
School 5*CSR*Language status	–3.80	(4.37)	–0.87	0.385
School 6*CSR*Language status	0.38	(5.33)	0.07	0.944
School 7*CSR*Language status	–0.05	(4.66)	–0.01	0.992
School 8*CSR*Language status	1.29	(4.43)	0.29	0.772
School 9*CSR*Language status	–1.62	(5.28)	–0.31	0.758
School 10*CSR*Language status	2.37	(3.85)	0.62	0.539
School 11*CSR*Language status	–0.94	(4.54)	–0.21	0.837
School 12*CSR*Language status	–1.23	(12.97)	–0.10	0.925
School 13*CSR*Language status	4.21	(5.79)	0.73	0.467
School 14*CSR*Language status	0.81	(3.58)	0.23	0.821
School 15*CSR*Language status	–0.35	(5.77)	–0.06	0.952
School 16*CSR*Language status	2.96	(4.63)	0.64	0.523
School 17*CSR*Language status	–2.64	(5.76)	–0.46	0.647
School 19*CSR*Language status	3.38	(7.65)	0.44	0.659
School 20*CSR*Language status	7.89	(5.11)	1.54	0.123

Variable	*Estimate*		t-*value*	*Pr* > |*t*|
School 21*CSR*Language status	–3.27	(6.37)	–0.51	0.609
School 22*CSR*Language status	3.33	(5.58)	0.60	0.551
School 23*CSR*Language status	–4.28	(3.23)	–1.33	0.186
School 24*CSR*Language status	–0.25	(5.60)	–0.05	0.964
School 25*CSR*Language status	–0.84	(5.29)	–0.16	0.874
School 26*CSR*Language status	–7.80	(5.16)	–1.51	0.134
Random effects	*Variance*		*Chi–square*	p–*value*
Intercept (U_0)	0.07	(0.27)	11.77	>.50
Language slope (U_1)	1.55	(1.24)	16.20	0.44
Level 1 R	38.69	(6.22)		

Note: Figures in parentheses are standard errors for fixed effects and standard deviations for random effects. Analysis included 23 schools, 66 teachers, and 1,199 students (279 FC–ELL CSR, 256 FC–ELL control; 322 non–ELL CSR , 324 non–ELL control).
Source: Authors' analysis of student GRADE scores, study records, and teacher surveys.

References

Allison, P. D. (2002). *Missing data.* Thousand Oaks, CA: Sage Publications, Inc.

American Guidance Service, Inc. (2001). *Group Reading Assessment and Diagnostic Evaluation.* San Antonio, TX: Pearson Education.

August, D., and Shanahan, T. (Eds.). (2006). *Developing literacy in second language learners: Report of the National Literacy Panel on Language Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum.

Biancarosa, C., and Snow, C. E. (2006). *Reading next—a vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education.

Billingsley, B. S., and Wildman, T. M. (1990). Facilitating reading comprehension in learning disabled students. *Remedial and Special Education, 11*(2), 18–31.

Bloom, H. S. (2005). Randomizing groups to evaluate place-based programs. In H. S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches* (pp. 115–172). New York: Russell Sage Foundation.

Boruch, R. (1997). *Randomized experiments for planning and evaluation: A practical guide*. Thousand Oaks, CA: Sage Publications, Inc.

Calderón, M., Hertz-Lazarowitz, R., and Slavin, R. (1998). Effects of bilingual cooperative integrated reading and composition on students making the transition from Spanish to English reading. *Elementary School Journal, 99,* 153–165.

Carlisle, J. F., and Rice, M. S. (2002). *Improving reading comprehension research-based principles and practices*. Timonium, MD: York Press, Inc.

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., et al. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly, 39,* 188–215.

Chiappe, P., Siegel, L. S., and Wade-Woolley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading, 6,* 369–400.

Cohen, E. G. (1994a). *Designing groupwork*. New York: Teachers College Press.

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. San Diego, CA: College Hill.

Duffy, G. G. (2002). The case for direct explanation of strategies. In C. C. Block and M. Pressley (Eds.), *Comprehension and instruction: research-based best practices* (pp. 28–41). New York: Guilford Press.

Durkin, D. (1978). What classroom observations reveal about reading comprehension. *Reading Research Quarterly, 14*(4), 481–533.

Gajria, M., and Salvia, J. (1992). The effects of summarization instructions on text comprehension of students with learning disabilities. *Exceptional Children, 58,* 508–516.

Garcia, E. (1994). *Understanding and meeting the challenge of student cultural diversity*. Boston: Houghton Mifflin.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, et al. (2008). *The impact of two professional development interventions on early reading instruction and achievement* (NCEE 2008-4030). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Gersten, R. (1996). Literacy instruction for language-minority students: The transition years. *Elementary School Journal, 96,* 227–244.

Gersten, R., and Baker, S. (2000). What we know about effective instructional practices for English language learners. *Exceptional Children, 66*(4), 454–470.

Glazerman, S., Levy, D., and Myers, D. (2002). *Nonexperimental replications of social experiments: A systematic review.* Washington, DC: Mathematica Policy Research, Inc.

Glazerman, S., Levy, D., and Myers, D. (2003). Nonexperimental versus experimental estimates of earnings impacts. *The ANNALS of the American Academy of Political and Social Science, 589,* 63–93.

Hakuta, K. (1990). *Bilingualism and bilingual education: A research perspective*. Washington, DC: George Washington University, Center for Applied Linguistics.

Hudelson, S. (1987). The role of native language literacy in the education of language minority children. *Language Arts, 64,* 827–841.

Jacob, E., Rottenberg, L., Patrick, S., and Wheeler, E. (1996). Cooperative learning: Context and opportunities for acquiring academic English. *TESOL Quarterly, 30,* 253–280.

James-Burdumy, S., Mansfield, W., Deke, J., Carey, N., Lugo-Gil, J., Hershey, A., et al. (2009). *Effectiveness of selected supplemental reading comprehension interventions: Impacts on a first cohort of fifth-grade students* (NCEE 2009-4032). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Jiménez, R. T., García, G. E., and Pearson, P. D. (1995). Three children, two languages, and strategic reading: Case studies of bilingual and monolingual readers. *American Educational Research Journal, 32*(1), 67–97.

Klingner, J. K., and Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities who use English as a second language. *The Elementary School Journal, 96,* 275–293.

Klingner, J. K., and Vaughn, S. (1998). Using collaborative strategic reading. *Exceptional Children, 30,* 32–37.

Klingner, J. K., and Vaughn, S. (1999). Promoting reading comprehension, content learning, and English acquisition through collaborative strategic reading (CSR). *The Reading Teacher, 52*, 738–747.

Klingner, J. K., and Vaughn, S. (2000). The helping behaviors of fifth graders while using collaborative strategic reading during ESL content classes. *TESOL Quarterly, 34*, 69–98.

Klingner, J. K., Vaughn, S., Dimino, J. A., Bryant, D. P., and Schumm, J. S. (2001). *Collaborative strategic reading: Strategies for improving comprehension.* Longmont, CO: Sopris West.

Klingner, J. K., Vaughn, S., and Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth grade classrooms. *The Elementary School Journal, 99*, 1–22.

Lipsey, M. W., and Cordray, D. S. (2000). Evaluation methods for social intervention. *Annual Review of Psychology, 51*, 345–375.

Little, R. J. A., and Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Logan-Fain, L., Sharma, N., Grant, M., Ribalta, G., and Wilkins, C. (2008). *What's in a need? Identifying customers' needs to help improve student outcomes*. Interim Research Bulletin. San Antonio, TX: Regional Educational Laboratory Southwest and Washington, DC: U.S. Department of Education.

Long, M., Adams, L., McLean, M., and Castanos, F. (1976). Doing things with words-verbal interaction in lockstep and small group classroom situations. In J. Fanselow and R. Crymes (Eds.), *On TESOL '76* (pp. 137–153). Washington, DC: TESOL.

Long, M., and Porter, P. (1985). Group work, interlanguage talk, and second language acquisition. *TESOL Quarterly, 19*, 207–228.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., and Hughes, K. E. (2006). *Gates-MacGinitie Reading Tests® (GMRT®) Fourth Edition Forms S and T* - Paper-Pencil*. Chicago: Riverside Publishing Company.

Market Data Retrieval. (2010). *Education universe 2009–2010.* Shelton, CT: Market Data Retrieval. Retrieved July 20, 2010, from http://www.schooldata.com

McGavock, M. (2009). *Bilingual and English language learner/limited English proficient technical assistance guide: Guide to meeting the needs of language learners*. Oklahoma City, OK: Oklahoma State Department of Education. Retrieved July 9, 2010, from http://sde.state.ok.us/Curriculum/Bilingual/pdf/TechAssist.pdf

Moss, M., and Puma, M. (1995). *Prospects: The congressionally mandated study of educational growth and opportunity*. Cambridge, MA: Abt Associates, Inc.

National Center for Education Statistics. (n.d.). *Common Core of Data, 2007–2008*. Washington, DC: U.S. Department of Education, Institute for Education Sciences. Retrieved June 22, 2010, from http://nces.ed.gov/ccd/

National Center for Education Statistics. (2004). *2003–2004 School Staffing Survey.* Retrieved July 16, 2010, from http://nces.ed.gov/surveys/sass/tables_list.asp

National Clearinghouse for English Language Acquisition. (2008). *The growing numbers of limited English proficient students: 1995/96–2005/06.* Washington, DC: Office of English Language Acquisition, Language Enhancement and Academic Achievement for Limited English Proficient Students.

National Institute for Literacy. (2008). *Explicit comprehension strategy instruction.* Retrieved July 10, 2010, from http://www.adlit.org/article/27896

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Education Research, 78*(1), 33–84.

Palincsar, A. S. (1986). Reciprocal teaching. In Palincsar, A., Ogle, D., Jones, B.F., and Carr, E.G. (Eds.)*Teaching reading as thinking*. Oak Brook, IL: North Central Regional Laboratory.

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

Pearson, P. D., and Dole, J. A. (1987). Explicit comprehension instruction: A review of research and a new conceptualization of instruction. *Elementary School* Journal, 88, 151–165.

Pica, T., Lincoln-Porter, F., Paninos, D., and Linnell, J. (1996). Language learners' interaction: How does it address the input, output, and feedback needs of L2 learners? *TESOL Quarterly, 30,* 59–84.

Pressley, M., Wood, E., Woloshyn, V., Martin, V., King, A., and Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist, 27,* 91–109.

Ragan, A., and Lesaux, N. K. (2006). Federal, state, and district level English language learner program entry and exit requirements: Effects on the education of language minority learners. *Education Policy Analysis Archives, 14*(20), 1–29.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., and Solenberger, P. (2001). A multivariate technique for multiple imputing missing values using a sequence of regression models. *Survey Methodology,* 27(1), 85–95.

Raghunathan, T. E., Solenberger, P. W., and Van Hoewyk, J. (2002). IVEware: Imputation and Variance Estimation Software [Computer software]. Ann Arbor, MI: University of Michigan. Retrieved July 20, 2010, from http://www.isr.umich.edu/src/smp/ive/

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

Rosenshine, B., and Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64,* 479–530.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Ruiz-de-Velasco, J., and Fix, M. (2000). *Overlooked and underserved: Immigrant students in U.S. secondary schools.* Washington, DC: The Urban Institute.

Saenz, L. M., Fuchs, L. S., and Fuchs, D. (2005). Peer-assisted learning strategies for English language learners with learning disabilities. *Exceptional Children, 71*, 231–247.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Schafer, J. L., and Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Shadish, W., Cook, T., and Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Short, D., and Fitzsimmons, S. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners: A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.

Stuart, E. A., Azur, M., Frangakis, C. E., and Leaf, P. (2009). Multiple imputation with large datasets: A case study of the Children's Mental Health Initiative. *American Journal of Epidemiology, 169*(9), 1133–1139.

Texas Education Agency. (2008). *TELPAS reading: Information booklet.* Austin, TX: Texas Education Agency.

Texas Education Agency. (2008–09a). *Limited English proficient decision chart.* Retrieved July 15, 2010, from http://ritter.tea.state.tx.us/curriculum/biling/

Texas Education Agency. (2008–09b). *2008–2009 English proficiency exit criteria chart.* Retrieved July 15, 2010, from http://ritter.tea.state.tx.us/curriculum/biling/

Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Cardenas-Hagan, E., et al. (2006). Effectiveness of a Spanish intervention and an English intervention for English language learners at risk for reading problems. *American Educational Research Journal, 43*(3), 449–487.

Vaughn, S., Hughes M. T., Schumm, J. S., and Klingner, J. K. (1998). A collaborative effort to enhance reading and writing instruction in inclusion classrooms. *Learning Disability Quarterly, 21*, 57–74.

Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., et al. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *Elementary School Journal, 107*(2), 153–180.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Willis, A. (2000). *Addressing literacy needs in culturally and linguistically diverse classrooms* [Critical Issue]. North Central Regional Education Laboratory. Retrieved June 15, 2010,from http://www.ncrel.org/sdrs/areas/issues/content/cntareas/reading/li400.htm


DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA