Routledge
Taylor & Francis Group

Journal of Research on Educational Effectiveness

ISSN: 1934-5747 (Print) 1934-5739 (Online) Journal homepage: https://www.tandfonline.com/loi/uree20

Collaborative Strategic Reading: Findings From Experienced Implementers

Sharon Vaughn , Greg Roberts , Janette K. Klingner , Elizabeth A. Swanson , Alison Boardman , Stephanie J. Stillman-Spisak , Sarojani S. Mohammed & Audrey J. Leroux

To cite this article: Sharon Vaughn , Greg Roberts , Janette K. Klingner , Elizabeth A. Swanson , Alison Boardman , Stephanie J. Stillman-Spisak , Sarojani S. Mohammed & Audrey J. Leroux (2013) Collaborative Strategic Reading: Findings From Experienced Implementers, Journal of Research on Educational Effectiveness, 6:2, 137-163, DOI: 10.1080/19345747.2012.741661

To link to this article: https://doi.org/10.1080/19345747.2012.741661

Journal of Research on Educational Effectiveness, 6: 137–163, 2013

ISSN: 1934-5747 print / 1934-5739 online
DOI: 10.1080/19345747.2012.741661

Routledge
Taylor & Francis Group

Collaborative Strategic Reading: Findings From Experienced Implementers

Sharon Vaughn and Greg Roberts
The University of Texas at Austin, Austin, Texas, USA

Janette K. Klingner
University of Colorado, Boulder, Colorado, USA

Elizabeth A. Swanson
The University of Texas at Austin, Austin, Texas, USA

Alison Boardman
University of Colorado, Boulder, Colorado, USA

Stephanie J. Stillman-Spisak, Sarojani S. Mohammed, and Audrey J. Leroux
The University of Texas at Austin, Austin, Texas, USA

Abstract: This study examined the effects and fidelity of collaborative strategic reading (CSR) implemented by experienced CSR teachers (participated in previous study; Vaughn et al., 2011) on the reading comprehension outcomes of students in English/Language Arts (ELA) or Reading classes. Eligible teachers (12 of 17; others reassigned to teach grades/subjects not eligible for inclusion) in middle schools in Texas and Colorado who participated in the previous year in a study examining the effects of CSR on the reading comprehension outcomes of their students participated in a 2nd-year, new cohort of students. Teachers taught multiple sections of ELA or reading; sections were randomly assigned to a treatment or comparison condition, and any extra classes were assigned to the treatment condition. There were 26 CSR and 22 comparison classes. Teachers were asked to implement CSR in their treatment classes only for approximately two 50-min sessions per week for 18 weeks. Examining the role of fidelity revealed that CSR was more prevalent in treatment classes than the comparison classes and that ELA teachers had significantly less treatment spillover than the Reading teachers. Findings indicate that CSR was associated with a greater effect when implemented in ELA classrooms compared to Reading classrooms.

Keywords: Reading, comprehension, fidelity, cooperative, collaboration

Researchers have referred to reading comprehension as the "sine qua non of reading" (Beck & McKeown, 1998), "acquiring meaning from written text" (Block, Gambrell, & Pressley, 2002), "the process of extracting and constructing meaning through interaction and involvement with written language" (Sweet & Snow, 2003), and "thinking guided by print" (Perfetti, 1985). Reading comprehension is essential as students advance through the grades and transition from learning to read to reading to learn. Students' difficulties with comprehension can lead to their becoming disaffected learners and dropping out of school,

Address correspondence to Sharon Vaughn, The University of Texas at Austin, The Meadows Center for Preventing Educational Risk, 1912 Speedway, D4900, Austin, TX 78712-1284, USA. E-mail: srvaughn@austin.utexas.edu

restriction of their access to postsecondary learning, and reduction of their employment opportunities (Daniel et al., 2006; Scanlon & Mellard, 2002).

Although reading comprehension is central to reading and necessary for success in school, studies report that it is not frequently taught well. Observation studies in general education and special education classrooms over the last 30 years suggest that teachers infrequently provide instruction in how to read for understanding, including how to use comprehension strategies (Durkin, 1978–1979; Gelzheiser & Myers, 1991; Klingner, Urbach, Golos, Brownell, & Menon, 2010; E. A. Swanson, 2008; E. A. Swanson & Vaughn, 2010; Vaughn, Moody, & Schumm, 1998). This problem is exacerbated only at the middle school level, where it is assumed that students read for understanding and teachers often perceive that reading instruction is not their role (Hall, 2005). Secondary teachers perceive that students can understand and learn from text, and so they do not focus enough on supporting their students in learning comprehension strategies (Biancarosa & Snow, 2006).

Although there have been numerous calls over the years for an increased focus on reading comprehension (e.g., Snow, 2002; or, at the secondary level, Biancarosa & Snow, 2006), it appears that adolescents' reading comprehension is not improving. Less than 32% of eighth graders across the United States comprehended what they read at or above proficient levels on the National Assessment of Educational Progress (National Center for Educational Statistics, 2009). This percentage is similar to what it was in 1992, and scores have shown little fluctuation in the years since. For students in urban schools and students from lower income homes, the percentages of students at or above proficient levels are even lower (Lapp, Grigg, & Tay-Lim, 2002).

In this article, we report findings from the 2nd year of a 2-year study of the efficacy of Collaborative Strategic Reading (CSR), a multicomponent approach to teaching reading comprehension strategies, with seventh- and eighth-grade English/Language Arts or Reading teachers. We previously reported findings from the 1st-year experimental study where teachers, who were novice implementers of CSR, had the sections of their classes randomized to either implementing CSR two to three times per week for approximately 18 weeks or teaching typical English/Language Arts or Reading instruction (Vaughn et al., 2011). Findings from the 1st year of the study indicated that students who were assigned to sections of CSR outperformed those who did not receive CSR on a standardized reading comprehension measure ($g = 0.12$).

Many teachers find it challenging to learn to teach comprehension strategies and also to implement and effectively use collaborative groups (an essential feature of CSR implementation; e.g., Klingner, Vaughn, Argüelles, Hughes, & Ahwee, 2004; Scanlon, Deshler, & Schumaker, 1996). In response, we invited the teachers who participated in the first study (Year 1; Vaughn et al., 2011) to continue participating in the study a 2nd year. The goal was to determine whether additional professional development and practice would improve implementation fidelity and be associated with improved outcomes for participating students (new cohort of students). Our primary research question was to determine the efficacy of CSR with experienced CSR teachers on middle school students' reading comprehension, and second we were interested in the role of implementation fidelity on students' outcomes.

THEORETICAL FRAMEWORK

CSR has roots in cognitive psychology (Flavell, 1979; Palincsar & Brown, 1984) and sociocultural theory (Perez, 1998; Vygotsky, 1978). CSR was originally adapted from reciprocal

teaching (Palincsar & Brown, 1984) and shares many features with its predecessor. As in reciprocal teaching, CSR teaches students to apply metacognitive and cognitive strategies for improving comprehension. The metacognitive processes that the reader learns are those involved in monitoring understanding, selecting what to remember, and regulating strategy use (Flavell, 1979). The cognitive strategies students use include previewing text, clarifying unknown words, determining main ideas, and generating questions and summarizing about what they have read. Guided by their teacher, students work collaboratively with their peers to read a text and use comprehension strategies. CSR also systematically builds students' background knowledge, motivation, and engagement, as well as assists teachers in identifying appropriate levels of text difficulty for their students.

Reading comprehension is an active process that may be thought of as a way to promote student engagement with text by developing mental models that represent what they are reading (Kintsch, 1998, 2004). These mental models stay activated as long as they seem to fit the text, or, in other words, make sense. Readers are taught to continually fine-tune these mental models through an integrative process where they resolve inconsistencies. When readers struggle to read the words in a text, they may overrely on background knowledge and guess about what they are reading. On the other hand, when readers lack background knowledge, they have a difficult time thoroughly understanding the text.

Many factors affect students' reading comprehension (Biancarosa & Snow, 2006; Carlisle & Rice, 2002; Kintsch, 2004; Snow, 2002), including the characteristics of the reader, the text, the purpose of the task, and the learning environment. Kintsch (1998) emphasized students' "zones of learnability" and asked that educators consider their background knowledge, motivation, and interest, as well as their decoding skills. Similarly, Edmonds et al. (2009) and Guthrie, Wigfield, and VonSecker (2000) reported on the centrality of engagement, motivation, self-efficacy, and interests. Text features also influence understanding, including the readability of the text and variations in text structure and organization. Ideal texts are neither too easy nor too difficult and require students to engage with the text in cognitively demanding ways (Mastropieri, Scruggs, & Graetz, 2003).

Research Support for Reading Comprehension Instruction

Reading comprehension strategy instruction is based on research showing that good readers apply many different strategies to facilitate their comprehension of text and that when poor comprehenders learn these same strategies, their comprehension can improve (Kamil et al., 2008; Paris, Wasik, & Turner, 1991). Several researchers have conducted research syntheses or meta-analyses on reading comprehension strategy instruction and have reported positive outcomes (Edmonds et al., 2009; Gajria, Jitendra, Sood, & Sacks, 2007; Gersten, Fuchs, Williams, & Baker, 2001; Mastropieri, Scruggs, Bakken, & Whedon, 1996; Mastropieri et al., 2003; Scammacca et al., 2007; H. L. Swanson, Hoskyn, & Lee, 1999; Talbott, Lloyd, & Tankersley, 1994; Vaughn, Gersten, & Chard, 2000). Edmonds et al. (2009) synthesized research examining the benefits of reading comprehension strategies, specifically for adolescent struggling learners, and found that adolescents with reading difficulties or learning disabilities (LD) improved in comprehension when provided with a targeted reading intervention in comprehension or multiple reading components. They noted that it is important to engage students in thinking about text, learning from text, and discussing what they know.

Although with fifth graders rather than adolescents, McKeown, Beck, and Blake (2009) conducted a study of two reading comprehension approaches, "content" and "strategies,"

that is relevant to our research. In the content instruction condition, teachers focused students' attention on the content of the text they were reading through open, meaning-based questions. In the strategies instruction, students learned specific procedures to guide their access to text. All outcome measures were researcher developed and included a Sentence Verification Technique, Oral Recall, Lesson Discourse, Comprehension Monitoring, and a Strategies Task. McKeown et al. (2009) found that content students outperformed strategies students on all measures except for the Sentence Verification Technique. They concluded that methods focused on content are superior to approaches centered on strategies. Yet they did not look specifically at students with LD or struggling readers, who may benefit from explicit strategy instruction (McKeown et al., 2009). We conclude that the most effective approaches for struggling readers probably include elements focused on both content and strategies. CSR addresses content learning through teacher- and student-lead conversations as well as the enhancement of strategy use within information texts.

Previous CSR Research

CSR incorporates many features associated with effective instruction, particularly for struggling readers, such as (a) explicit instruction, (b) modeling, (c) guided practice, (d) procedural strategies to facilitate learning, (e) collaborative partner or group work, and (f) opportunities for interactive dialogue among students as well as between teachers and students (Biancarosa & Snow, 2006; Gajria et al., 2007; Gersten et al., 2001; H. L. Swanson et al., 1999; Vaughn, Chard, et al., 2000).

The precursor to CSR was a modified version of Palincsar and Brown's (1984) reciprocal teaching (Klingner & Vaughn, 1996). Twenty-six eighth-grade English Language Learners with LD learned reading comprehension strategies and then either applied them in cooperative learning groups or taught them to younger peers. Even students who were poor decoders made significant improvements in reading comprehension on the Gates–MacGinitie Reading Comprehension Test. Next, in a quasi-experimental study, Klingner, Vaughn, and Schumm (1998) taught fourth-grade students in diverse, heterogeneous classes to implement CSR while reading social studies texts. Students in a comparison condition participated in typical teacher-directed instruction in the same content. CSR students made greater gains than comparison students on the Gates–MacGinitie Reading Comprehension Test ($p = .001$; Hedges's $g = 0.44$) and equivalent scores on a social studies unit test from the textbook.

As a means of understanding the influence of CSR on students who were English Language Learners, Klingner and Vaughn (2000) tape-recorded and transcribed fifth-grade bilingual students' discourse while they used CSR in small groups with their science textbook. Students spent more than 99% of their time engaged in academic-related strategic discussion and almost no time off task. Students helped one another to relate what they were learning to previous knowledge, understand word meanings, get main ideas, and ask and answer questions. CSR was also examined in 10 inclusive fourth-grade classrooms with more than 50% English language learners. Students in five CSR classes and five comparison classes participated. CSR students improved more on the Gates–MacGinitie Reading Comprehension Test than comparison students ($p = .01$; effect sizes = 0.25 for high- and average-achieving students, 0.51 for low-achieving students, and 0.38 for students with LDs).

Related specifically to the study reported in this manuscript is a previous randomized control trial conducted in seventh- and eighth-grade English/Language Arts and Reading

classes in six diverse, low- to mid-income middle schools (Vaughn et al., 2011). Participating teachers taught CSR in some of their classes and provided typical instruction in their other classes. Students in CSR classes made greater gains on the Gates–MacGinitie Reading Comprehension Test than did their peers in comparison classes. Results for a sample of low-achieving students were similar to those for the total sample. In previous CSR studies (Klingner et al., 2004), as well as the study preceding this one (Vaughn et al, 2011), the research team has recognized that teachers have difficulty effectively implementing cooperative grouping structures and working with these groups after they are organized. The present study extends previous research by including the same sample of teachers and a new cohort of students to determine the relative effects of CSR with experienced teachers on the reading comprehension of middle school students.

Research on Measuring Fidelity

The importance of implementation fidelity is acknowledged across a number of sources (e.g., Hulleman & Cordray, 2009; E. A. Swanson et al., 2011). However, the vast majority of studies to date treat fidelity data descriptively, primarily as a means of estimating the extent to which treatment units implement the normative program model. Formal measurement models are generally not employed. Traditionally, observations of fidelity have been used to illustrate the extent to which intervention strategies are implemented in treatment classes only. However, advances in the field have shown that fidelity measures can provide more useful quantitative information. When applied to both treatment and comparison conditions, fidelity measures can be used to investigate not only the extent to which implementation of the treatment has occurred but also the extent to which this treatment is different from the condition to which it is being compared, sometimes called "achieved relative intervention strength" (Hulleman & Cordray, 2009, p. 88). Thus, fidelity can be incorporated into analyses as a latent variable, providing clearer evidence for the connection between implementation of the treatment (as distinct from instructional strategies being utilized in comparison conditions) and any subsequent main effects found. Building on this growing body of research, this study incorporates fidelity into multilevel analyses to more fully explore and document the relationship between implementation of CSR strategies by experienced teachers and middle school students' reading comprehension.

METHOD

Participant Characteristics and Sampling Procedures

This study was conducted in six middle schools located in three school districts (two near-urban and one urban) in two states. Each school served a diverse student population.

Teachers

During Year 2 of the study, 12 teachers (3 male, 9 female) who taught English/Language Arts or Reading to seventh and eighth graders and who participated in the CSR study the previous year (Vaughn et al., 2011) were included. Five teachers did not continue after Year 1 due to changes in teaching assignment that made them no longer eligible to continue. All teachers earned a bachelor's degree, and eight possessed a master's degree. Teaching

Table 1. Teacher demographics

Characteristic	Year 1[a]	Year 2[b]
Gender		
Male	3	3
Female	14	9
Ethnicity		
White	16	11
Asian/Pacific Islander	1	1
Average education	16.9 years	17.3 years
Degrees earned		
Bachelor's	9	4
Master's	8	8
Experience	10.1 years	12.0 years
Specializations		
ELA	15	10
Secondary	7	4
Elementary	7	6
Reading	6	5
ESL	2	1
Special education	2	1
Speech	1	1

Note. ELA = English/Language Arts; ESL = English as a Second Language.
[a]$n = 17$.[b]$n = 12$.

experience ranged from 2 to 36 years ($M = 12$, $Mdn = 11$). Ten teachers were certified to teach English/Language Arts in addition to other subject areas or special populations (e.g., special education, English as a second language). Teacher demographic information is provided in Table 1.

Students

Prior to the school year, each teacher identified the class periods during which regular English/Language Arts or Reading curriculum would be delivered (i.e., not classes identified as sheltered learning for newcomers because students did not speak English and spoke multiple languages). Schools assigned students to these 48 identified classes. These 48 classes were then randomly assigned within teacher to either CSR treatment or Typical Practice (TP) comparison condition. There were 26 CSR and 22 TP classes. For teachers with an odd number of classes, the additional class was assigned to the treatment condition.

A total of 528 students were randomized prior to treatment onset; 472 students completed the study, as indicated by a posttest score on the Gates–MacGinitie. Students' average age in the CSR group was 13.52 years and in the TP group was 13.50 years. Most students were in seventh grade ($n = 451$), with fewer in eighth grade ($n = 77$). Twelve percent of all students were identified as experiencing reading difficulty. These students failed the previous year's state-developed reading assessment and scored at least 1 standard deviation below the standardized mean on the TOWRE (Scanlon & Mellard, 2002). Demographic information for students in the full sample and the reading difficulty subsample are displayed in Table 2. A preliminary screening of the student data indicated a systematically irregular pattern of scores on the Gates–MacGinitie. In the

Table 2. Student demographics

	Year 1				Year 2			
	CSR		TP		CSR		TP	
Characteristic	*n*	%	*n*	%	*n*	%	*n*	%
Gender								
Female	199	43.4	188	46.1	123	46.6	117	44.3
Male	170	37.1	166	40.7	133	50.4	138	52.3
Ethnicity								
African American	13	2.8	13	3.2	5	1.9	7	2.7
American Indian	4	0.9	3	0.7	0	0.0	1	0.4
Asian	2	0.4	9	2.2	5	1.9	4	1.5
Caucasian	172	37.6	136	33.3	145	54.9	124	47.0
Hispanic	178	38.9	193	47.3	101	38.3	119	45.1
LEP	10	2.2	11	2.7	18	6.8	14	5.3
Special education	56	12.2	28	6.9	17	6.4	18	6.8

Note. CSR = Collaborative Strategic Reading; TP = Typical Practice; LEP = Limited English Proficiency.

comparison condition for one teacher, students made an average gain of approximately 12.5 standard score points, with the gains evenly distributed across the 22 cases. In the same teacher's treatment condition ($n = 24$), the average gain was about –0.8. The gain in the comparison group appeared to be due to non-completion of the Gates–MacGinitie assessment at pretest (no student in this class completed more than half of the items at pretest), resulting in artificially deflated pretest scores. This pattern was not evident in any other teacher's CSR or TP classes, and it indicates a threat to the validity of this teacher's data. Accordingly, we elected to remove this teacher from the analysis of the Gates–MacGinitie data, reducing the available Gates-specific sample to 482 students across 45 classes and 11 teachers. The full sample ($n = 528$) was used to evaluate other student outcomes, to model implementation variables, and to evaluate attrition.

Attrition. Overall attrition can bias a study's external validity to the extent that the initial sample represents a population of interest and the students who attrite differ from those who remain in the study (Graham & Donaldson, 1993; Miller & Wright, 1995; Valentine & McHugh, 2007). Differential attrition is evident when treatment and control groups differ (a) in the proportion of cases leaving prior to the study's terminal event, or (b) in the average attributes of (i.e., demographic characteristics, etc.) of noncontinuing participants. Although differential attrition can bias effect estimates and threaten the design's *internal* validity, its effect can be particularly profound when teamed with elevated levels of overall attrition (Miller & Hollist, 2007). A two-step method first described by Cook and Campbell was used to evaluate the effects of overall and differential attrition. Treatment conditions, completer status (i.e., did or did not attrite), and the interaction of condition and completer status were regressed on the Gates–MacGinitie pretest measures and key demographic variables. Significant main effects for the group of completers indicated significant overall attrition and a potential threat to the result's external validity. The Condition × Completer Status interaction provided an indicator of differential attrition. A significant interaction in this context signifies systematic group differences in the characteristics of dropped cases.

Assumptions regarding missing data and causality were informed by Rubin's framework (e.g., Little & Rubin, 1987; Rubin, 1976).

Demographic data were available for the full sample of 511 of the 528 initially randomized cases. There were no significant main or interaction effects for the binomial logistic regression of gender on condition, completer status, or for the interaction term. For the binomial logistic regression with limited English proficiency, there were no main effects or Condition × Completer Status interaction effects. A multinomial logistic regression with ethnicity as the independent variable indicated no differences for condition, or the interaction term. The same pattern of no differences was evident for free or reduced lunch, special education, and English Language Learner status. When Gates–MacGinitie scores were regressed on completer status and the interaction of completer status and condition, there were no significant main effects for completer status or the interaction. In sum, overall attrition was relatively minimal (11%) and there was no evidence of differential attrition.

Procedures

Professional Development and Teacher Support

All teachers had participated in professional development for CSR during the previous year that included an initial training in CSR over 3 days (6 hr per day) on implementing the treatment practices. This professional development focused on (a) an overview of the study, (b) a careful description of an experimental study including the importance of adhering to "business as usual" in comparison classes, and implementing instructional practices in treatment classes, (c) critical features of the intervention practices and how to teach them to students, and (d) how to use collaborative groups within the CSR model. All necessary materials to implement the treatment were provided, including sample lessons, examples of reading materials, and overheads for classroom instruction. We also provided focused professional development on three occasions for about 90 min spread throughout the implementation (approximately 20 weeks including holiday breaks) to enhance and refine implementation and use of CSR in their classrooms.

During the 2nd year of the CSR implementation, we provided a 6-hr full day of professional development that focused on (a) an overview of the Year 2 study, (b) a reminder of the importance adhering to "business as usual" in the control classes, (c) a refresher on the components of the intervention including a discussion of lessons learned from Year 1 on how best to teach students these components, and (d) how to improve the effectiveness of collaborative groups. Teachers were shown preliminary results showing the effectiveness of CSR for improving student outcomes in Year 1, and the professional development included a discussion of areas in which refinement of implementation could lead to even greater effectiveness of the intervention (e.g., gist and clunks). Similar to Year 1, teachers were provided all necessary materials to complete the implement the treatment, including sample lessons and reading materials. Also similar to Year 1, we provided focused professional development on three occasions for approximately 90 min spread throughout the implementation. In Year 2, these follow-up sessions focused more on refining the teachers' practices to maximize the effectiveness of the intervention rather than on modeling the intervention components, which was the focus of these professional development follow-up sessions in Year 1. Professional development was designed to enhance teachers' use and instruction of the comprehension strategies, organization and use of collaborative groups, and purposeful instruction before and after group work.

In addition, the research team provided in-class support and coaching. One research support person was assigned to each participating teacher and was initially present one to two times per month during a treatment class, and then less so over the course of the intervention. Research support person activities included modeling how to teach CSR strategies and demonstrating "think-aloud" practices, side-by-side teaching, and teacher observation with feedback.

Description of the Treatment Intervention

Students in the treatment classes received the intervention during their regularly scheduled seventh-grade or eighth-grade English/Language Arts or Reading class. Teachers were asked to implement the intervention for 50 min a day, 2 days a week for approximately 18 weeks. Teachers reported that the number of sessions they implemented ranged from 18 to 61 sessions. Typical instruction was provided for students whose English/Language Arts or Reading classes were randomly assigned to comparison classes. One of the critical elements of the design was that students in both the treatment and comparison conditions covered the same curriculum material over the same period using the same English/Language Arts or Reading content and standards. Thus, all students had ample opportunities to improve reading comprehension from within the given curriculum, regardless of randomized assignment to treatment or comparison classes. The precise timing of the implementation of each of the strategies could not be controlled by the researchers because teachers needed to teach the strategies at the rate in which their students could learn and apply them. All of the teachers introduced all of the strategies to all of their students in the first two to three lessons. Teachers provided further explanation and application of each of the strategies to adjust to students' responses.

Collaborative Strategic Reading. The CSR intervention comprises four comprehension strategies that students learn to use before reading (i.e., preview), during reading (i.e., click and clunk, and get the gist), and after reading (i.e., wrap up). Furthermore, students were taught to use the strategies independently and within a leadership role in collaborative groups where each student rotates roles, and eventually assumes responsibility for guiding others in their group in the use of each of the comprehension strategies. During the treatment implementation of CSR, teachers were asked to use expository text about half of the time. In addition, we asked teachers to model each strategy to students through an extensive use of think-alouds with time for guided practice of each strategy and multiple opportunities for providing feedback to students. If students struggle with mastering the use of a particular strategy, teachers were instructed to implement short lessons to provide practice and review.

Students were taught to use previewing before reading a passage. The preview practice encompasses four activities: (a) the teacher (or preview leader when students were working in cooperative groups) introduced the passage topic and preteaches any proper nouns or specialized vocabulary that may be difficult, (b) students discussed what they already knew about the topic, (c) students were taught to use text features such as headings and graphics to learn as much about the text quickly before reading, and (d) students predicted what they thought they would learn from the passage. Students recorded their brainstormed ideas and predictions in their learning log.

Students were taught to use two strategies during reading: (a) click and clunk or word/idea monitoring, and (b) get the gist (or main idea strategy) during reading. The click and clunk strategy was designed to help students identify breakdowns in understanding and then resolve the misunderstandings using a series of "fix up" strategies. As students

read the first section of text, they were instructed to identify "clunks," or breakdowns in understanding, and record them in their learning logs. After reading the section, students returned to the clunks and used the following "fix-up" strategies to understand the meaning: (a) reread the sentence without the word—think about what word meaning would make sense; (b) reread the sentences before and after the clunk, looking for clues to determine the word meaning; (c) identify key elements in the word (e.g., prefixes, suffixes, a known word part); and (d) identify word parts that may aid in understanding. Also during reading, students were instructed to use a practice called "Get the Gist," which is similar to writing the main idea. Students were taught to restate in their own words the most important information from a section of reading as a way of making sure they understood the most important ideas of what they just read.

After reading, students engaged in the final review strategy that encompasses question generation (Raphael, 1982) and summative statement writing. The goal of question generation is to improve students' knowledge, understanding, and memory of the passage read. Students were taught to write three levels of questions. "Right there" questions are those with answers that can be found in one sentence in the passage. These questions help students remember facts and focus on the most important information. "Think and search" questions were more difficult to write and require students to remember several events or facts from different sections of the passage in order to answer the question. These questions help students synthesize information from the passage. "Author and you" questions require inference on the students' part. Students were taught to use facts from the passage to make inferential conclusions. Students generated and answered each type of question in their learning logs. Finally, students were taught to write in their learning logs a summative statement that includes the most important ideas from the passage. Students were also asked to use the text to justify why these were identified as the most important ideas to remember.

Collaborative Grouping. Students first learn to use the strategies with modeling and feedback from the teacher. As they increase in proficiency with using the strategies (4–6 weeks), the teacher assigns them to cooperative learning groups of four to five students. Students were assigned roles as strategy leaders (e.g., gist expert) and served as guides in ensuring that all students are using the strategy effectively. Each student assumes one of the following four roles: (a) leader, (b) clunk expert, (c) gist expert, and (d) question expert. Other roles (encourager and timekeeper) were assigned as needed. During a 40- to 50-min session students worked in their small groups, with teacher feedback, utilizing all of their CSR strategies to assist all group members in raising the level of comprehension of the assigned text. In a typical CSR session, the teacher might begin by reviewing an instructional practice (e.g., take 5 min to remind students how to use feedback from other members of the group to develop a "gist" or main idea). Students then work in their groups reading the text, and each student exercises their leadership role to assure the strategies are implemented by the group. Students used their learning logs to record previews, gists, clunks, questions, and summaries. The teacher moved from group to group, guiding instruction, listening and providing feedback, and asking questions to check for understanding and to enhance student learning of the text content.

Measures

Both fidelity and student achievement measures were administered to participants in this study. Next, fidelity measures are described first, followed by student measures, which focused on achievement.

Internal Validity Checklist (IVC; Vaughn et al., 2011)

A three-part observation rubric was developed for recording fidelity. The *first section* focused on procedural fidelity, with possible scores ranging from 0 (*not observed*) to 4 (*highly aligned to CSR practices*). In this section, teachers were rated on each component of CSR, including brainstorm, predict, identify clunks, use fix-up strategies, get the gist, ask and answer questions, and review. Scores were assigned according to quality and adherence to the intervention, ranging from 1 (*inconsistent with CSR procedures*) to 4 (*highly aligned with CSR procedures*). In Section 2, three global quality items, (a) quality of teacher's instruction, (b) quality of teacher's classroom management, and (c) quality of teacher's CSR implementation, were rated on a scale from 1 (*low*) to 7 (*high*). For Section 3, teacher and student behaviors (e.g., teacher monitors quality of strategy use; teacher provides effective feedback; students help each other in groups; students use assigned CSR roles) that are essential to high-quality implementation of CSR received scores from 1 (*not observed*) to 4 (*observed*). Scores of 2 and 3 in this section indicated partly observed and mostly observed respectively. Detailed field notes were also recorded. All lessons were audio recorded to review if a score for any of the aforementioned ratings was questionable. No discourse analysis was completed of these recordings for the purposes of this article.

Classroom observations were conducted to monitor implementation fidelity four times during the school year. Observations were evenly spaced throughout the year and coordinated in advance with participating teachers. All observations were in-person, and the lesson as a whole was considered when completing the Internal Validity Checklist (i.e., an interval rating system was not used). Three researchers at each site conducted the observations. Observers were trained on use of the rubric and interrater reliability of 90% or more was established before observations began. Training on the observation instrument consisted of using a scoring manual to address implementation fidelity of in-person observed CSR lessons conducted by teachers who were not study participants. Ratings were discussed until consensus was reached. Discrepant scores were addressed and the process was repeated until interrater agreement of .80 (Cohen's kappa) was reached consistently among all four raters. Interrater agreement was established before each of the four observations.

Evidence of Strategy Use in Typical Comparison Classrooms

We were interested in determining the extent to which reading strategies were implemented in typical comparison classrooms. Because there was an emphasis in all of our districts on improving reading comprehension outcomes for students, we wanted to document the extent to which the strategies that we emphasized in CSR were also being implemented in typical comparison classes. For example, as part of typical practice, some teachers previewed text prior to reading, asked students to write main ideas statements, and taught students questioning procedures that were the same as the CSR question-and-answer component. This is not surprising, because these instructional practices have been documented as being associated with improved outcomes in reading for middle school students and are likely part of most teachers' preparation (Kamil et al., 2008). We used the same observation measure in typical classrooms to determine whether there was spillover between treatment and typical classes (i.e., whether teachers were implementing CSR in typical classrooms).

To monitor the extent to which reading strategies might be taught in typical classrooms, we relied on information provided by the teachers in reflections, coaching sessions, and interviews along with information from four observations of the typical classes of each teacher over the course of the study. For example, if teachers provided instruction on a main

idea strategy in a typical classroom, they were assigned a score that measured the alignment with the CSR strategy. If a strategy was not observed, a teacher was given a "not observed" score. Overall, teachers in typical classrooms received a procedural fidelity average rating of 2.34, indicating that instruction was different from CSR classroom instruction (3.20).

In the 2nd year of the study, issues of spillover were infrequently identified or observed. Teachers reported and we observed that they understood clearly the instructional components related to CSR and the need to contain CSR strategies to treatment classrooms. Spillover of CSR strategy instruction was not observed. However, we suspect that there may have been an unanticipated overall teacher effect of participating in the 2 years of CSR professional development that influenced the quality of instruction in the typical classes. Teachers were provided ongoing professional development in which they worked together to develop their professional understanding of strategy instruction, shared implementation ideas and problem solved with one another, had multiple opportunities to receive in and out of class-support and feedback from researchers, and were provided with materials and teaching tools to support implementation, all features aligned with principles of effective professional development (e.g., Desimone, 2009).

Implementation Logs

At the close of each CSR class period, teachers reported the number of minutes spent engaged in CSR strategy instruction and/or practice as well as recorded notes about the day's instruction. Research support personnel collected logs from all teachers at the end of each week of the intervention implementation. These data, along with classroom observations, were used to determine the frequency of use of CSR.

Student Measures

All student measures were administered by trained research personnel who were unaware of students' condition (treatment or typical). All student measures were administered prior to treatment and immediately following treatment, with the exception of the Test of Word Reading Efficiency, which was administered only prior to treatment in order to help identify struggling readers.

Gates–MacGinitie Reading Test (4th ed.; MacGinitie, MacGinitie, Maria, & Dreyer, 2000)

The Gates–MacGinitie Reading Test is a timed, group-administered assessment of reading comprehension. The assessment consists of narrative and expository passages ranging in length from three to 15 sentences. Students read each passage silently and then answer three to six multiple-choice questions related to the most recently read passage. Items increase in difficulty as the student progresses through the assessment during the 35-min time limit. Internal consistency reliability ranges from .91 to .93 and alternate form reliability is reported as .80 to 87.

Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner, Torgesen, Rashotte, & Pearson, 2010)

This timed, group-administered assessment measures a student's silent reading efficiency (i.e., speed and accuracy) and comprehension during a 3-min silent reading task. Students receive a list of sentences and are asked to verify the truthfulness of as many sentences

as possible. Sentence length ranges from four to 10 words and sentences increase in difficulty as the student progresses through the assessment. The total number of correct items comprises the score. TOSREC has an alternate-form reliability ranging from .74 to .83 in Grades 6 through 8.

Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999)

The TOWRE measures students' ability to read words out of context. It includes two separate, individually administered, timed assessments. The Sight Word Efficiency subtest measures students' ability to recognize common words quickly and accurately, whereas the Phonemic Decoding Efficiency subtest measures a student's ability to sound out nonwords quickly and accurately. Depending upon the subtest, students are provided a list of either 104 words or 66 nonwords that are listed in order of increasing difficulty. The raw score is determined by the number of words or nonwords read correctly in 45 s. Internal consistency exceeds .95 for both subtests.

Data Analysis Procedures

Treatment fidelity and spillover effects were estimated using confirmatory factor models. Program effects were estimated within an intent-to-treat framework using multilevel models. Multigroups models were used to estimate the mediating effect of fidelity. The data met distributional assumptions associated with multivariate inferential analyses.

Fidelity and Spillover

As previously described, classroom observational protocols were based on the CSR program model (*intended model*), which represents the key teacher activities and the causal mechanisms that hypothetically link instruction to anticipated student outcomes. The intended model for CSR is the program described in earlier sections of this article. The implemented (or enacted) model represents the varieties of CSR presented to students in classrooms. In the present study, enacted programs were assumed to exist across conditions (and across classes) within teachers. Ideally, teachers would enact a version of CSR in their treatment classes and enact something other than CSR in their typical classes; however, recognizing that elements of CSR could spill over into typical classes and threaten the study's internal validity because of the within-teacher design, teacher behaviors in both treatment and typical classes were characterized in terms of elements of CSR for purposes of developing and administering observation protocols. Collecting implementation data against a common benchmark (the CSR intended model, in this case) provided a vehicle for explicitly indexing the extent of fidelity and spillover. Blocking on teacher controlled the effect of "teacher quality."

Protocols were administered on eight occasions within each teacher, four times in treatment classes and four times in typical classes. Correspondence between the intended and enacted program in treatment classes was the basis for conceptualizing *fidelity*, addressing the question, "To what extent did the teacher implement the program as intended?" Within a given teacher, the degree of alignment between the intended program and the enacted program in the typical classes represented treatment *spillover* ("To what extent did the teacher implement the program in the typical condition?"), which was a primary threat

to the design's internal validity. These data were examined descriptively and also used as covariates in models estimating treatment effects.

Confirmatory factor analysis with categorical indicators was used to estimate factors related to fidelity and spillover. Six latent variables were specified, each corresponding to a key CSR-aligned teacher practice: (a) Brainstorming, (b) Preview, (c) Click and Clunk, (d) Fix-Up Strategies, (e) Questioning, and (f) Get the Gist. A higher-order factor representing overall alignment of intended and enacted models was also specified. Because implementation was "phased in" over time, the likelihood of observing a given element varied by occasion. For example, brainstorming was an early-implemented program element and observers in treatment classes had a high expectation of witnessing its use (while observers in typical classes would expect not to see brainstorming). Click and Clunk was implemented later and observer expectations were calibrated accordingly. Score data from appropriate time points were used to estimate latent factors, such that each latent variable represents the degree to which a given teacher implemented a particular CSR component (or CSR generally in the case of the higher order factor) within a given class across the treatment frame according to the implementation schedule.

Fidelity-related parameters were estimated as continuous, with means of 0 and freely estimated variances. Class-level data ($n = 48$) were modeled as clustered within teachers. Robust weighted least squares estimator was used (in MPLUS 6) with path loadings estimated as probit regressions of the observed indicators on the latent variable. Factor scores for fidelity/spillover were estimated in a preliminary model and imported to the student-level data file for subsequent analysis of treatment effects.

Treatment Effects

Treatment effect was conceptualized in terms of several influences, including the intended program model (CSR) and how the model was implemented, if at all, in the treatment and typical conditions. The stratified and nested character of the data was addressed using multilevel models, with students nested in classes and classes randomized to treatment within teachers. Parameter values were estimated for the within- and between-groups' model components using MPLUS 6. An *intent-to-treat* effect was modeled in relation to students' pretest scores and assigned treatment condition. Pretest scores were used as cluster-level covariates to maintain power. Within groups, posttest scores were regressed on pretest scores. Between-group (class) differences were evaluated by regressing posttest intercepts (i.e., intercept is the mean of the posttest scores when the effect of pretest is removed or when the pretest score is 0) on the dummy-coded indicator for treatment condition (preliminary analyses confirmed no differences in groups' covariance structures), which varied within teachers and between clusters. The coefficient for the between-groups' regression of outcome on condition provided an index of significance and effect. Values differing significantly from 0 represented treatment effects. Intent-to-treat effect sizes were reported in raw and standardized units. Effects were also evaluated separately for English/Language Arts teachers and Reading teachers.

Intent-to-treat analyses use initial group assignment as the indicator of treatment, regardless of the type or dose of the actual treatment received. Cross-classified structures are generally ignored. Fidelity of implementation and potential spillover effects are similarly disregarded, at least for purposes of estimating or adjusting main effects. However, because low fidelity and high spillover typically bias effects by diminishing sample-estimated values, measuring implementation and modeling English/Language Arts nested data, as outlined in earlier sections of this article, may minimize bias and yield more accurate effect

estimates. The within-teacher aspect of the study's design was motivating in this respect, particularly in terms of potential spillover. Latent factor scores from the confirmatory analyses described earlier were modeled as mediating the effect of treatment on reading outcomes. Hypothetically, to the extent that fidelity was high and spillover was low, the latent scores indicating implementation would *mediate* the effect of treatment. That is, the effect of treatment assignment, under the mediating condition, would influence fidelity (or minimize spillover), which in turn would influence outcomes.

Student data were analyzed using multilevel, multigroup models with sampling- and assignment-related clustering adjusted according to the TYPE = COMPLEX and TYPE = TWOLEVEL defaults in MPLUS 6 (i.e., teacher was specified as a *stratification* variable and class was the *cluster* variable). Data were assumed to be missing at random and models were parameterized using a full information maximum likelihood estimator.

RESULTS

Descriptive summaries for student outcomes are reported in Tables 3 and 4. Pretest and posttest mean values are indicated by treatment condition and teacher type for the Gates–MacGinitie and TOSREC. Estimates of between- and within-groups variance are reported, as well. Results of the inferential analyses are presented in three parts: (a) the confirmatory factor model of implementation fidelity, (b) of intent-to-treat, and (c) modified as-treated effect estimates.

Model of Implementation Fidelity

An unconditional factor model fit the data extremely well, $\chi^2(118) = 123.28$, $p = .35$; comparative fit index = .99, Tucker–Lewis index = .99, root mean square error of approximation = .031. As indicated earlier, we modeled observed data according to the phased-in plan for implementing elements of CSR; Brainstorming, Preview, Click and Clunk observed scores from the first three of four observations and Time 2 through Time 4 Internal Validity Checklist values for Fix-Up Strategies, Get the Gist, and Questioning were used to estimate latent variables. Also, note that indicators in the baseline measurement model include observations from both treatment and typical conditions, corresponding to the within-teacher nature of the fidelity data. Thus, although the overall mean for a latent factor score is 0 by default, fidelity scores vary across conditions according to the alignment of intended and enacted models, with higher levels of observed implementation associated with higher scores on a given latent factor, whether data were collected in a treatment or a typical class. A high score in a typical class indicates spillover, whereas a low score in a treatment class suggests low fidelity. Residual variances for the first-order factors did not differ significantly from 0 and were constrained as equal across groups in the final model for purposes of model identification. The higher order latent factor, CSR, was fit using the six skill-specific latent variables.

Figure 1 presents parameter estimates. Not surprisingly, model elements were highly correlated at both the measurement and structural levels (i.e., the higher order factor). Note also that inclusion of the time three data point for GS resulted in a nonidentified model, due presumably to its perfect correlation with other indicators. It was estimated from the final model. Average factor scores for each latent variable are presented in Table 5. The mean score in the treatment condition for overall CSR fidelity (i.e., the higher order factor, CRS)

Table 3. Descriptive summary of means and standard deviations by program year, measurement occasion, and condition

	Year 1				Year 2			
	Pretest		Posttest		Pretest		Posttest	
Measure	TP	CSR	TP	CSR	TP	CSR	TP	CSR
Gates[a]	95.68(13.4)	96.35(13.7)	95.48(13.4)	97.13(13.6)	98.20(14.5)	98.45(13.9)	99.19(14.3)	100.71(14.2)
TOSREC	23.49(7.4)	24.67(7.2)	29.03(7.7)	29.22(7.7)	24.67(7.0)	24.92(5.7)	29.42(7.8)	29.81(7.3)
TOWRE–SW	90.17(7.4)	90.59(8.4)			89.68(8.9)	89.36(8.8)		
TOWRE–PD	91.53(11.3)	91.20(10.9)			89.35(12.0)	89.28(11.4)		

Note. Standard deviations are in parentheses. TP = Typical Practice; CSR = Collaborative.
Strategic Reading; Gates = Gates–MacGinitie Reading Test; TOSREC = Test of Silent Reading Efficiency and Comprehension; TOWRE–SW = Test of Word Reading Efficiency Sight Word Efficiency subtest; TOWRE–PD = Test of Word Reading Efficiency Phonemic Decoding Efficiency subtest.
[a]The values for Gates–MacGinitie reflect the 11 teachers/45 classes sample described in the narrative. Standard scores are presented for the Gates and the TOWRE. We report raw scores for the TOSREC.

Table 4. Descriptive summary of means and standard deviations by program year, measurement occasion, and condition for the struggling readers sample

	Year 1				Year 2			
	Pretest		Posttest		Pretest		Posttest	
Measure	TP	CSR	TP	CSR	TP	CSR	TP	CSR
Gates[a]	84.37(10.0)	85.62(11.2)	84.25(9.1)	87.74(10.0)	85.29(10.2)	81.15(9.9)	87.20(12.5)	87.17(8.3)
TOSREC	18.52(4.6)	20.30(5.9)	23.87(6.3)	24.83(5.8)	18.27(4.9)	19.22(4.9)	21.70(5.5)	23.12(5.4)
TOWRE-SW	84.48(5.5)	83.76(4.6)			82.52(5.5)	82.13(7.3)		
TOWRE-PD	81.57(5.5)	82.55(5.6)			78.00(6.1)	78.90(7.2)		

Note. Standard deviations are in parentheses. TP = Typical Practice; CSR = Collaborative Strategic Reading; Gates = Gates-MacGinitie Reading Test; TOSREC = Test of Silent Reading Efficiency and Comprehension; TOWRE-SW = Test of Word Reading Efficiency Sight Word Efficiency subtest; TOWRE-PD = Test of Word Reading Efficiency Phonemic Decoding Efficiency subtest.

[a]The values for Gates-MacGinitie reflect the 11 teachers/45 classes sample described in the narrative. Standard scores are presented for the Gates and the TOWRE. We report raw scores for the TOSREC.

Table 5. Implementation values based on confirmatory factor model by condition and teacher type

	Treatment Condition					Comparison Condition				
	BR	PR	CL	FX	CSR	BR	PR	CL	FX	CSR
ELA	.651 (.338)	.650 (.303)	.690 (.321)	.722 (.348)	.661 (.318)	–.649 (.436)	–.655 (.392)	–.722 (.399)	–.739 (.437)	–.674 (.404)
Rdg.	.575 (.520)	.545 (.511)	.642 (.499)	.682 (.509)	.597 (.496)	–.371 (.032)	–.291 (.055)	–312 (.045)	–.377 (.045)	–.331 (.032)
Total	.636 (.378)	.625 (.355)	.679 (.361)	.707 (.380)	.656 (.359)	–.574 (.391)	–.559 (.370)	–.613 (.382)	–.645 (.405)	–.583 (.374)

Note. Values in parentheses are standard deviation units for the mean value (above each standard deviation) within each condition and teacher type and across all teacher types (in the Total section). BR = Brainstorming; PR = Preview; CL = Click and Clunk; FX = Fix-Up Strategies; CSR = Collaborative Strategic Reading; ELA = English/Language Arts; Rdg. = Reading.

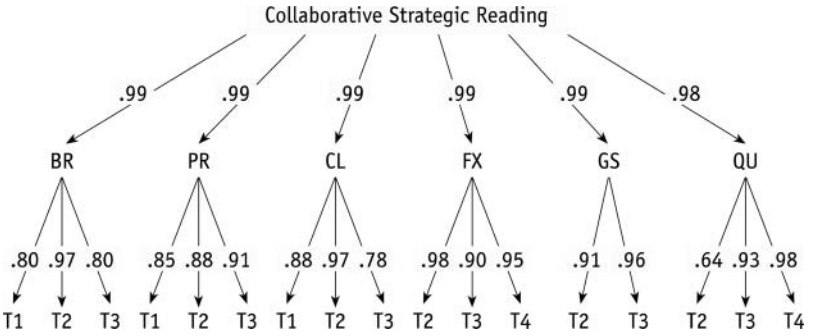

Figure 1. Confirmatory model for implementation fidelity. *Note.* Coefficients are standardized. All values are significant at $p < .001$. PR = Preview; CL = Click and Clunk; FX = Fix-Up Strategies; GS = Get the Gist; QU = Questioning.

was 0.591 (Table 5). The standard deviation was 0.285 (given the high intercorrelation among the six first-order factors, we focus primarily on the CSR variable). In the typical classes, the mean average was –0.5896 and the standard deviation was 0.407. The difference in levels of fidelity across the two conditions was statistically significant ($\Delta\chi^2 = 38.51$, $\Delta df = 1, p < .001$), suggesting that CSR was more prevalent, on average, in the treatment classes. The effect size in z-score units was 1.19 (calculated as the sum of the absolute values for the mean CSR scores in the treatment and comparison groups), suggesting more than 1 standard deviation of separation between the two groups. Fidelity/spillover estimates were calculated within teacher type (English/Language Arts and Reading teachers) and by condition (Table 5). The means for English/Language Arts and Reading teachers within the treatment condition differed significantly ($\Delta\chi^2 = 4.57$, $\Delta df = 1, p < .05$).

Intent-to-Treat Effects

Intent-to-treat effects were estimated according to the nested and stratified structure of the data. We anticipated site-related differences in treatment effects given dissimilarities in sampled teachers and the students served. In one state, English/Language Arts teachers (with one exception) were targeted for the study, meaning that student participants were likely to be more variable and closer to the schoolwide mean in terms of reading achievement to the extent that all students participated in English/Language Arts classes. In the other state, the sample comprised Reading teachers (with one exception), who were tasked with supporting the group of lower performing students. To evaluate the extent to which CSR's effect may have varied by students' general reading level, we considered the interaction of treatment condition and teacher type by fitting specific models for each of the four indicated groups (i.e., Reading teachers' typical condition, Reading teachers' treatment condition, English/Language Arts teachers' typical condition, and English/Language arts teachers' treatment condition). Main treatment effects were estimated as a multigroup model with clustering controlled using the TYPE = COMPLEX defaults (i.e., clustering was not explicitly modeled). Multigroup, multilevel models were not possible because of the confounding of cluster and treatment condition (due to the within-teacher design). Pretest scores were centered on the grand mean. The model-estimated posttest means on the Gates–MacGinitie were 99.5 and 100.95 for the typical and treatment groups,

respectively. The mean difference (about 1.5 standard score points) represents the main effect for treatment across the 11 teachers and 45 classes with Gates–MacGinitie data. Although not statistically significant ($\Delta\chi^2 = 1.35$, $\Delta df = 1$, $p = .25$), the effect size is virtually identical to that reported for Year 1 of this same study (Vaughn et al., 2011). For dichotomous independent variables that are standardized using the variances of continuous latent variables and outcome variables ($\beta_{\text{StdY}} = b/SD(y)$, β_{StdY} is the change in y standard deviation units when x changes from zero to 1, which represents a traditional effect size; note also that 1.5 is 10% of the normative standard deviation for the Gates. There were no statistically significant group differences on the TOSREC standard score (unstandardized $b = 0.84$, $p = .44$) for the full sample (i.e., 48 classes and 12 teachers), and there were no statistically significant group differences in the group of low-performing readers (51 students across 27 classes within 11 teachers) on the Gates–MacGinitie ($b = 0.071$), or TOSREC ($b = 1.58$, $p = .536$).

As anticipated, average student performance as indicated by teacher type differed. The between-clusters unstandardized coefficient for the posttest intercept on teacher type was 12.37 ($p < .001$) for the Gates–MacGinitie, meaning an estimated average difference of more than 12 standard score points at posttest between students in English/Language Arts and Reading classes across treatment condition (this indicates general differences, not differences in effect). The *interaction* of condition by teacher type on Gates–MacGinitie scores was evaluated using multigroup, multilevel models with group-specific models fit and compared across the four possible conditions (TOSREC scores were not evaluated). In the general multilevel model, posttest scores were modeled in terms of pretest values both within- and between-classes. Condition-specific posttest means (or intercepts) were estimated separately on the between-clusters level of the model.

As an omnibus test of potential differences, the aforementioned model with freely estimated intercepts for the four groups was compared to the model with the four posttest means constrained as equal. The freely estimated model fit the data significantly better than the constrained model ($\Delta\chi^2 = 39.99$, $\Delta df = 3$, $p < .001$), suggesting the presence of significant differences across the four posttest Gates's scores. Further, in a multilevel model with posttest Gates scores regressed on treatment condition, teacher type, and the interaction of condition and type (rather than modeling the within-group covariance via the multigroup approach), the interaction term was the only statistically significant predictor ($b = 5.027$, $p = .05$), suggesting that group differences may be due to more than teacher type only. To isolate their source(s), a final multilevel, multigroup model was fit with treatment condition regressed on posttest scores for each of the two teacher types. The constrained model, with the effect of condition (i.e., $b_{y/x}$) fixed as equal across teacher type, fit the data significantly less well than the model with effects allowed to vary across groups ($\Delta\chi^2 = 5.172$, $\Delta df = 1$, $p = .02$). This suggests that CSR may have a greater effect when implemented in English/Language Arts classrooms compared to Reading classrooms. It is not clear whether this is due to differences in the students who typically attend these classrooms, to teachers who typically teach them, or to other factors. The significantly greater levels of spillover in the Reading teacher typical classes compared to English/Language Arts typical conditions may be a contributing factor, as well. Table 6 summarizes parameters for this final model.

Model of Implementation Fidelity

Treating the latent fidelity/spillover scores as mediators of the effect of treatment provided a mechanism for evaluating the extent to which assignment corresponds to actual use of the

Table 6. Summary of model-estimated posttest sample means and estimates of between- and within-group residual variance

Condition	Teacher Type	Measure	Intercept	Between-Group Variance	Within-Group Variance
CSR	ELA	Gates[a]	103.9	13.0	6.9
		TOSREC	30.7	6.9	2.7
	Reading	Gates[a]	88.8		
		TOSREC	24.8		
TP	ELA	Gates[a]	101.7	13.1	5.2
		TOSREC	29.5	7.1	3.8
	Reading	Gates[a]	91.8		
		TOSREC	25.3		

Note. Intercept represents model-estimated posttest group means. CSR = Collaborative Strategic Reading; TP = Typical Practice; ELA = English/Language Arts; Gates = Gates–MacGinitie Reading Test; TOSREC = Test of Silent Reading Efficiency and Comprehension.

[a]The values for Gates–MacGinitie reflect the 11 teachers/45 classes sample described in the narrative. Standard scores are presented for the Gates. We report raw scores for the TOSREC.

intervention (i.e., the extent to which the intervention was implemented in treatment conditions and not implemented in the typical conditions). A model with outcomes regressed on the fidelity/spillover scores (and on pretest scores) was fit in a multigroup framework. Teacher type (English/Language Arts or Reading) was included as a covariate given the earlier findings. Fitting the treatment-group-specific models' situated condition as the independent variable, treatment-group differences in the relationship of fidelity/spillover and student outcomes represented a mediating role for fidelity/spillover.

As anticipated, the results for the mediated models do not differ substantively from those of the intent-to-treat analyses. There were no statistically significant differences in estimated Gates–MacGinitie ($\Delta\chi^2 = 0.8459$, $\Delta df = 1$, $p \geq .05$) posttest scores. Similarly, when the regression of student outcomes on fidelity was constrained as equal across treatment groups (a constraint that *hypothetically* eliminates any difference due to assignment), the expected pattern emerged. Differences on the Gates–MacGinitie ($\Delta\chi^2 = 4.97$, $\Delta df = 1$, $p = .03$) were statistically significant, suggesting that implementation mediated the effect of assignment on outcomes, that fidelity was reasonably well established, and that spillover was relatively minimized.

DISCUSSION

In this study, we examined the relative effects of CSR implementation by experienced CSR teacher-implementers (Year 2) who had participated in the treatment the previous year (Year 1). In the Year 1 study, students in the treatment significantly outperformed typical students on reading comprehension (Vaughn et al., 2011). We hypothesized similar findings for the Year 2 study; however, this hypothesis was not confirmed with findings revealing no statistically significant impact for treatment students over typical students. Although the effect size was similar across Years 1 and 2, there were fewer available units (i.e., classes with teachers) for randomization in Year 2 and thus less power. With respect to our primary research question, whether teachers who were experienced CSR implementers would yield

significantly higher outcomes in reading comprehension for their target students, there is no evidence from this study to support advantaged outcomes for students from experienced teachers. There are several possible explanations for this lack of significant effect for Year 2.

Perhaps the most relevant influences on the Year 2 study are the complications of implementing a 2-year study in ever-changing schools. These historical influences affected our study by reducing the number of teachers from Study 1 to Study 2 (some teachers were reassigned to roles that were not appropriate for continued participation in the study), changing the instructional roles of participating teachers from Study 1 to Study 2 (some teachers who were English/Language Arts teachers in Study 1 were assigned to be Reading Intervention teachers in Study 2), and the changing student composition resulting from teachers who were previously English/Language Arts teachers assigned to teach Reading Intervention resulting in same-ability, low readers in these classes rather than mixed-ability readers. In particular, we think the shift in several teachers' roles from English/Language Arts to Reading Intervention teachers resulting in classes of same-ability struggling readers was a barrier to CSR effectiveness.

There is a substantial research base on the effects of same-ability grouping on student achievement. Students in same-ability low-achieving groups are recipients of slower paced instruction that is perceived as being of lower quality (e.g., Gamoran, 1989; Oakes, 1992) and have fewer models for how to think about and interact with text (e.g., Rosenbaum, 1980). The lack of adequate models for thinking and discussing text seems to be particularly problematic in CSR because a significant component of implementation is the use of cooperative learning groups in which students are expected to assist one another (Klingner & Vaughn, 2000). In previous CSR research studies (e.g., Klingner et al., 2004; Klingner et al., 1998), students worked together in heterogeneous rather than homogeneous groups. Klingner et al. (2004) looked at the relative gains of students at different achievement levels and found that students considered to be low achieving had the highest effect sizes when compared with similar peers in typical classes (effect sizes = 0.25 for high- and average-achieving students, 0.51 for low-achieving students, and 0.38 for students with LDs). We do not know how well the low-achieving students would have fared without the support of high- and average-achieving students and can only speculate that they would not have done this well. The students in the Klingner et al. (2004) study were fourth graders, however—much younger than the middle school students in the current study.

We also examined the role of fidelity empirically. We conceptualized fidelity as the extent to which the treatment provided by teachers conformed to the specified treatment elements. Because teachers' classes were assigned to treatment and typical, we were also interested in the extent to which these treatment elements were part of their instruction in the typical classes. This study, like previous studies we have conducted using within teacher random-assignment at the class level, determined that there was very little overlap of the CSR elements as specified in the treatment (e.g., teachers were not previously using cooperative learning with student roles and they did not use them in typical groups after the study was launched; Vaughn et al., 2009). However, this finding does not mean that many of the critical comprehension practices that are recommended in CSR were not used in both the treatment and typical classes (e.g., main idea). In other words, teachers were asked to keep their typical classes the same as they usually were—make no changes. So if they taught main idea, they should continue teaching it the way they usually did in their typical classes. Because CSR is based on research that has been available for decades (Palincsar & Brown, 1984), it is likely that many of the reading strategies that are part of CSR were also part of the typical classes. The extent to which these CSR strategies (e.g., summarizing)

were used in typical classes would give us a true estimate of the distinction between the treatment and typical conditions with respect to implementation.

Reading teachers were more likely than English/Language Arts' teachers to use elements of CSR in their typical instruction. We do not necessarily consider this an example of lack of adherence to the design of the study (i.e. use of CSR only in treatment classes) but consider it a likely outcome associated with implementation from teachers whose roles were to support the reading success of students with reading difficulties. These teachers were informed about effective reading interventions and continued to use those practices in their typical classes (e.g., previewing, summarizing). The instructional practices are not "unique" to CSR but are represented in a range of reading intervention practices. The unique element is how the practices are taught and the role of students who take on increasingly more responsibility for guiding the thinking and discussing of text in their groups.

Students of English/Language Arts teachers implementing CSR were more likely to outperform their peers in typical classes in the English/Language Arts classes than were treatment students in reading teachers' CSR classes who performed about the same as their typical reading comparison. We believe that there are two possible explanations for this finding. We think this finding relates to two important points we briefly presented earlier. First, we know that English/Language Arts teachers were less likely to use CSR-related practices in their typical classes. Thus, there was considerably less overlap in CSR instructional elements in English/Language Arts treatment classes than English/Language Arts typical classes providing a clearer and more distinct test of the treatment (CSR) resulting in students' reading comprehension scores in CSR classes that were associated with improved outcomes. In contrast, Reading teachers' intervention and typical classrooms were more alike with respect to elements of CSR being less distinct and demonstrated more overlap, providing a less successful test of the treatment (CSR) resulting in students' scores in the treatment classes that were not different from students' scores in the typical classes.

Another explanation for the difference in performance of students in the English/Language Arts classes and the Reading classes can be derived from the research on same- and mixed-ability grouping. This literature is relevant because the English/Language Arts classes represented a mixed-ability grouping format and the Reading classes represented a same-ability grouping format. This grouping format is particularly salient when implementing CSR because fundamental to the success of CSR (at least as it is hypothesized to work), is that students of mixed ability are placed in small, cooperative groups with specific roles and responsibilities that, while they rotate, are maintained long enough to demonstrate effective group functioning. These roles require considerable knowledge and expertise (e.g., "gist" or main idea expert) and require students to question, model, and support the learning of others in the group. In the Reading classes, which comprised similar-ability low readers, students were also asked to work in cooperative groups and provide the same modeling and learning support to others in the group. However, without the range of reading ability in the group, we suspect that the overall group functioning was less successful with students providing fewer of the "teacher" behaviors that make improvements in students' reading comprehension within cooperative groups possible. This is an empirical question for future research to delve into more deeply.

Limitations and Implications

This study was designed as the 2nd year in a 2-year study in which teachers who participated during Year 1 would continue in the study as experienced CSR implementers for Year 2.

Due to several important historical effects, there was a combination of teacher attrition and teacher reassignment that made testing the primary question challenging. Longitudinal studies relying on teacher participation, in which teachers need to stay in the same role for the study to be successful, are difficult to implement because of the ongoing and considerable changes made in school personnel on a yearly basis. Estimation of treatment fidelity and implementation is an ongoing priority within the Institute of Education Sciences, National Institutes of Health, and other funding entities. In this study, we used data-reduction techniques to generate indicators of teachers' use of CSR strategies. The rigor of confirmatory models and the fit indices available in confirmatory analyses provide a strong basis for making assumptions about the validity and meaning of identified factors. Further, factor scores in confirmatory models are expressed along a standard metric, providing an element of precision that is not possible when treating data descriptively. These represent significant advantages over current methods for indexing fidelity, and though confirmatory models are sometimes difficult to fit to teacher observation data, this only highlights their value. Although an absence of systematic patterns in a set of teacher observations (as indicated by poorly fitting models) does not render the data useless, it does limit their use, certainly in comparison to data that can be reduced to meaningful factors that fit a theoretical model.

ACKNOWLEDGMENTS

This research was supported by grant R305AO80608 from the Institute of Education Sciences, US Department of Education. The content is solely the responsibility of the authors and does not neccessarily represent the official views of the Institute of Education Sciences or the US Department of Education.

REFERENCES

Beck, I. L., & McKeown, M. G. (1998). Comprehension: The sine qua non of reading. In S. Patton, & M. Holmes (Eds.), *The keys to literacy* (pp. 40–52). Washington, DC: Council for Basic Education.

Biancarosa, G., & Snow, C. E. (2006). *Reading next–a vision for action and research in middle and high school literacy: A report from Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education.

Block, C. C., Gambrell, L. B., & Presley, M. (2002). *Improving comprehension instruction: Rethinking research, theory, and classroom practice*. New York, NY: Jossey-Bass.

Carlisle, J. F., & Rice, M. S. (2002). *Improving reading comprehension, research-based principles and practices*. Timonium, MD: York.

Daniel, S. S., Walsh, A. K., Goldston, D. B., Arnold, E. M., Reboussin, B. A., & Wood, F. B. (2006). Suicidality, school dropout, and reading problems among adolescents. *Journal of Learning Disabilities*, *39*, 507–514.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*, 181–199. doi:10.3102/0013189×08331140.

Durkin, D. (1978–1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, *14*, 481–583.

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C. K., Cable, A., Tackett, K., & Wick, J. (2009). A synthesis of reading interventions and effects on reading outcomes for older struggling readers. *Review of Educational Research*, *79*, 262–300.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*, 906–911.

Gajria, M., Jitendra, A. K., Sood, S., & Sacks, G. (2007). Improving comprehension of expository text in students with LD: A research synthesis. *Journal of Learning Disabilities*, *40*, 210–225.

Gamoran, A. (1989). Measuring curriculum differentiation. *American Journal of Education*, *97*, 129–143.

Gelzheiser, L. M., & Myers, J. (1991). Reading instruction by classroom, remedial and resource room teachers. *The Journal of Special Education*, *24*, 512–526.

Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research*, *71*, 279–320.

Graham, J. W., & Donaldson, S. I. (1993). Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data. *Journal of Applied Psychology*, *78*, 119–128.

Guthrie, J. T., Wigfield, A., & VonSecker, C. (2000). Effects of integrated instruction on motivation and strategy use in reading. *Journal of Educational Psychology*, *92*, 331–341.

Hall, L. A. (2005). Teachers and content area reading: Attitudes, beliefs, and change. *Teaching and Teacher Education*, *21*, 403–414.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, *2*, 88–110. doi:10.1080/19345740802539325

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A practice guide* (NCEE#2008–4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://eric.ed.gov.ezproxy.lib.utexas.edu/PDFS/ED502398.pdf

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Kintsch, W. (2004). The construction–integration model of text comprehension and its implications for instruction. In R. Ruddell & N. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 1270–1380). Newark, DE: International Reading Association.

Klingner, J. K., Urbach, J., Golos, D., Brownell, M., & Menon, S. (2010). Teaching reading in the 21st century: A glimpse at how special education teachers promote reading comprehension. *Learning Disability Quarterly*, *33*, 59–74.

Klingner, J. K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities who use English as a second language. *Elementary School Journal*, *96*, 275–293.

Klingner, J. K., & Vaughn, S. (2000). The helping behaviors of fifth-graders while using collaborative strategic reading (CSR) during ESL content classes. *TESOL Quarterly*, *34*, 69–98.

Klingner, J. K., Vaughn, S., Argüelles, M. E., Hughes, M. T., & Ahwee, S. (2004). Collaborative strategic reading: "Real world" lessons from classroom teachers. *Remedial and Special Education*, *2*, 291–302.

Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. *Elementary School Journal*, *99*, 3–21.

Lapp, M. S., Grigg, W. S., & Tay-Lim, B. S. (2002). The nation's report card: U.S. history 2001. *Education Statistics Quarterly*, *4*, 21–34.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: Wiley & Sons.

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

Mastropieri, M. A., Scruggs, T. E., Bakken, J. P., & Whedon, C. (1996). Reading comprehension: A synthesis of research in learning disabilities. In T. E. Scruggs & M.A. Mastropieri (Eds.),

Advances in learning and behavioral disabilities (Vol. 10, Part B, pp. 277–30). Greenwich, CT: JAI Press.

Mastropieri, M. A., Scruggs, T. E., & Graetz, J. E. (2003). Reading comprehension instruction for secondary students: Challenges for struggling students and teachers. *Learning Disability Quarterly*, *26*, 103–116.

McKeown, M. G., Beck, I. L., &. Blake, R. G. K. (2009). Rethinking comprehension instruction: Comparing strategies and content instructional approaches. *Reading Research Quarterly*, *44*, 218–253.

Miller, R. B., & Hollist, C. S. (2007). Attrition bias. *Encyclopedia of Measurement and Statistics*, *1*, 57–60.

Miller, R. B., & Wright, D. W. (1995). Detecting and correcting attrition bias in longitudinal family research. *Journal of Marriage and Family*, *57*, 921–929.

National Center for Education Statistics, National Assessment of Educational Progress. (2009). Data explorer. Available at http://www.nces.ed.gov/nationsreportcard/nde/

Palincsar, A. S., & Brown, A. L. (1984). The reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*, 117–175.

Paris, A. H., Wasik, B. A., & Turner, J. C. (1991). The development of strategic readers. In R. Barr, M. L. Kamil, B. P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 609–640). Hillsdale, NJ: Erlbaum.

Perez, B. (1998). *Sociocultural contexts of language and literacy*. Mahwah, NJ: Erlbaum.

Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford University Press.

Raphael, T. E. (1982). Question–answering strategies for children. *The Reading Teacher*, *36*, 186–190.

Rosenbaum, J. E. (1980). Social implications of educational grouping. *Review of Research in Education*, *8*, 361–401.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C., & Torgesen, J. (2007). *Reading interventions for adolescent struggling readers: A meata-analysis with implications for practice*. Portsmouth, NH: RMC Research Corporation.

Scanlon, D., Deshler, D. D., & Schumaker, J. B. (1996). Can a strategy be taught and learned in secondary inclusive classrooms? *Learning Disabilities Research and Practice*, *11*, 41–57.

Scanlon, D. M., & Mellard, D. F. (2002). Academic and participation profiles of school-age dropouts with and without disabilities. *Exceptional Children*, *68*, 239–258.

Snow, C. E. (2002). *Reading for understanding: toward a research and development program in reading comprehension*. Santa Monica, CA: RAND Reading Study Group.

Swanson, E. A. (2008). Observing reading instruction for students with LD: A synthesis. *Learning Disability Quarterly*, *31*, 1–19.

Swanson, E. A., & Vaughn, S. (2010). An observation study of reading instruction provided to elementary students with learning disabilities in the resource room. *Psychology in the Schools*, *47*, 481–492. doi:10.1002/pits.20484.

Swanson, E. A., Wanzek, J., Vaughn, S., Haring, C., Ciullo, S., & McCulley, L. (2011). Intervention fidelity in special and general education research journals. *Journal of Special Education*. Advance online publication. doi:10.1177/0022466911419516

Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes*. New York, NY: Guilford.

Sweet, A. P., & Snow, C. E. (Eds.). (2003). *Rethinking reading comprehension*. New York, NY: Guilford.

Talbott, E., Lloyd, J. W., & Tankersley, M. (1994). Effects of reading comprehension interventions for students with learning disabilities. *Learning Disability Quarterly*, *17*, 223–232.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency*. Austin, TX: Pro-Ed.

Valentine, J. C., & McHugh, C. M. (2007). The effects of attrition on baseline comparability in randomized experiments in education: A meta-analysis. *Psychology Methods*, *12*, 268–282.

Vaughn, S., Chard, D. J., Bryant, D. P., Coleman, M., Brenda-Jean, T., Linan-Thompson, S., & Kouzekanani, K. (2000). Fluency and comprehension interventions for third-grade students. *Remedial and Special Education*, *21*, 325–335.

Vaughn, S., Gersten, R., & Chard, D. J. (2000). The underlying message in LD intervention research: Findings from research syntheses. *Exceptional Children*, *67*, 99–114.

Vaughn, S., Klingner, J. K., Swanson, E. A., Boardman, A. G., Roberts, G., Mohammed, S. S., & Stillman-Spisak, S. J. (2011). Efficacy of collaborative strategic reading with middle school students. *American Educational Research Journal.* Advance online publication. doi:10.3102/0002831211410305

Vaughn, S., Martinez, L. R., Linan-Thompson, S., Reutebuch, C. K., Carlson, C. D., & Francis, D. (2009). Enhancing social studies vocabulary and comprehension for seventh-grade English language learners at risk for reading problems. *Elementary School Journal*, *107*, 153–180. doi:10.1086/510653.

Vaughn, S., Moody, S., & Schumm, J. S. (1998). Broken promises: Reading instruction in the resource room. *Exceptional Children*, *64*, 211–226.

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

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2010). *Test of Silent Reading Efficiency and Comprehension (TOSREC) examiner's manual*. Austin, TX: Pro-Ed.