American Educational Research Journal
August 2011, Vol. 48, No. 4, pp. 938–964
DOI: 10.3102/0002831211410305

Efficacy of Collaborative Strategic Reading With Middle School Students

Sharon Vaughn
The University of Texas at Austin
Janette K. Klingner
University of Colorado
Elizabeth A. Swanson
The University of Texas at Austin
Alison G. Boardman
University of Colorado
Greg Roberts
Sarojani S. Mohammed
Stephanie J. Stillman-Spisak
The University of Texas at Austin

The authors conducted an experimental study to examine the effects of collaborative strategic reading and metacognitive strategic learning on the reading comprehension of students in seventh- and eighth-grade English/language arts classes in two sites (Texas, Colorado) and in three school districts. Students were randomly assigned to classes and then classes were randomly assigned to treatment or business-as-usual comparison groups. If a teacher had an uneven number of classes, we assigned extra classes to treatment. The total number of classes randomized was 61, with 34 treatment and 27 comparison. Treatment students received a multicomponent reading comprehension instruction (collaborative strategic reading) from their English/language arts/reading teachers that included teaching students to apply comprehension strategies in collaborative groups for 18 weeks, with approximately two sessions per week. Findings indicated significant differences in favor of the treatment students on the Gates-MacGinitie Reading Comprehension Test but not on reading fluency.

KEYWORDS: learning disabilities, school failure, reading failure, longitudinal/developmental work, children at risk

National and international studies have revealed that significant numbers of adolescents and young adults do not adequately understand complex texts, which impedes their secondary success, access to postsecondary

learning, and opportunities within our increasingly competitive work environment (Biancarosa & Snow, 2004; Kamil et al., 2008). At the national level, recent National Assessment of Educational Progress (NAEP) data showed a minimal gain in fourth- and eighth-grade reading comprehension scores since 1992. Bloom, Hill, Black, and Lipsey (2008) reported that when examining multiple data sources, annual student growth in reading achievement is by far the greatest during the first several grades of elementary school and declines considerably over time with students in high school grades making the least gains. Concerns are also heightened in a report by ACT revealing that only 50% of the ACT tested students are ready to read and understand college-level text (ACT, Inc., 2006).

SHARON VAUGHN is the executive director at The Meadows Center for Preventing Educational Risk, The University of Texas at Austin, 1 University Station, D4900, Austin, TX; e-mail: *srvaughn@austin.utexas.edu*. Dr. Vaughn is currently the Principal Investigator or Co-Principal Investigator on several Institute for Education Sciences, National Institute for Child Health and Human Development, and U.S. Department of Education research grants investigating effective interventions for students with reading difficulties and students who are English language learners.

JANETTE K. KLINGNER is a professor at University of Colorado. Dr. Klingner's research interests include reading comprehension strategy instruction for culturally and linguistically diverse students and professional development that supports the sustained implementation of evidence-based practices in diverse schools and classrooms.

ELIZABETH A. SWANSON is a research assistant professor at The Meadows Center for Preventing Educational Risk, The University of Texas at Austin. Current research interests include response to intervention and instructional interventions designed to remediate reading difficulties.

ALISON G. BOARDMAN is an assistant research professor at University of Colorado. Dr. Boardman's research focuses on effective instructional strategies for struggling readers in general education classrooms and ways to support teachers through high quality professional development and coaching.

GREG ROBERTS is an associate director at The Meadows Center for Preventing Educational Risk, University of Texas at Austin. Dr. Roberts has directed evaluations of several federally funded programs related to literacy, disability, and school change and has extensive experience managing large extant data sets and successfully managed several large data-collection efforts.

SAROJANI S. MOHAMMED is an assistant director at The Meadows Center for Preventing Educational Risk, University of Texas at Austin. Research interests include early and adolescent literacy, and Response to Intervention, with experience in data management and analysis, large scale external program evaluations, technical assistance, and online dissemination.

STEPHANIE J. STILLMAN-SPISAK is a director of assessment at The Meadows Center for Preventing Educational Risk, The University of Texas at Austin. Dr. Stillman-Spisak's work focuses on the development of efficient and effective models of measuring reading comprehension and content knowledge acquisition, particularly among struggling readers.

Unfortunately, there is not much room for optimism as students' reading achievement levels and understanding of complex text are at the lowest that they have been in a decade. The Programme for International Student Assessment (PISA) indicated that among the 30 countries that constitute the Organization for Economic Cooperation and Development (OCED), the United States ranked 21st on literacy-related outcomes. It is likely that a significant challenge for U.S. students is understanding the complex texts utilized in PISA assessments. The goal of comprehending text is not easily attainable for many adolescents, who demonstrate consistent and persistent difficulties with reading for understanding (Edmonds et al., 2009; Williams, 1998, 2000). This may be a result of many contributing factors to comprehension, perhaps including vocabulary knowledge, background knowledge, and the ability to use effective practices (e.g., monitoring understanding) while reading to improve comprehension.

To address the issue of improving reading comprehension for middle school students, we designed an experimental study to examine the effects of a multicomponent reading intervention (collaborative strategic reading; CSR) aimed at improving students' text comprehension. CSR is based on several related and converging research literatures, including a theoretical foundation in cognitive psychology and sociocultural theory, research on good comprehenders, and research regarding effective practices from reading comprehension studies.

Theoretical Foundation for Collaborative Strategic Reading

CSR is theoretically grounded in cognitive psychology (Flavell, 1992; Palincsar & Brown, 1984) as well as sociocultural theory (Perez, 1998; Vygotsky, 1978). Harris and Pressley (1991) emphasized that comprehension strategy instruction can be improved when it utilizes these theoretical perspectives as foundational. Like reciprocal teaching (Palincsar & Brown, 1984), CSR teaches students how to monitor their comprehension and also how to use procedures for clarifying understanding when difficulties arise. Students also learn main idea and questioning practices that assist them in reflecting on text while reading and guiding group responses to text after reading. Cooperative learning practices while implementing comprehension strategies in the context of reading are also a critical component of CSR. Through the collaborative approach emphasized with CSR, student learning is supported by both teachers and peers.

Additionally, CSR includes critical elements identified as enhancing the performance of students with learning difficulties, such as: (a) making instruction visible and explicit, (b) implementing procedural strategies to facilitate learning, (c) using interactive groups and/or partners, and (d) providing opportunities for interactive dialogue among students and between teachers and students (Fuchs, Fuchs, Mathes, & Lipsey, 2000; Gersten,

Fuchs, Williams, & Baker, 2001; Swanson, Hoskyn, & Lee, 1999; Vaughn, Gersten, & Chard, 2000).

Research on Effective Comprehenders

Many of the instructional practices suggested for poor readers were derived from observing both good and poor readers, questioning them about their reading strategies, and asking them to "think aloud" while they read (Dole, Duffy, Roehler, & Pearson, 1991; Heilman, Blair, & Rupley, 1998; Jiménez, Garcia, & Pearson, 1995, 1996). These reports described good readers as coordinating a set of highly complex and well-developed skills and strategies before, during, and after reading so that they could understand and learn from text and also remember what they read (Paris, Wasik, & Tumer, 1991). When compared with good readers, poor readers were considerably less strategic (Paris, Lipson, & Wixson, 1983). Good readers monitor the structure and organization of text, monitor their understanding while reading, make predictions, check them as they read, revise and evaluate them as needed, integrate what they know about the topic with new learning, and summarize and self-check their learning (Jenkins, Heliotis, Stein, & Haynes, 1987; Kamil, 2003; Klingner, Vaughn, & Boardman, 2007; Mastropieri, Scruggs, Bakken, & Whedon, 1996; Pressley & Afflerbach, 1995; Swanson, Hoskyn, & Lee, 1999; Wong & Jones, 1982). Knowledge of what good readers do was integrated into teaching students how to read for understanding with CSR.

Effective Practices From Reading Comprehension Intervention Studies

Several reviews of intervention research have reported positive outcomes for students who are taught to use comprehension strategies (e.g., Edmonds et al., 2009; Gajria, Jitendra, Sood, & Sacks, 2007; Gersten et al., 2001; Mastropieri et al., 1996; Matropieri, Scruggs, & Graetz, 2003; National Institute of Child Health and Human Development, 2000; Scammacca et al., 2007; Swanson et al., 1999; Talbott, Lloyd, and Tankersley, 1994; Vaughn et al., 2000). Other reviews analyze the effectiveness of unique components of comprehension instruction (see Maccini, Gagnon, & Hughes, 2002, for review of technology approaches; Kim, Vaughn, Wanzek, & Wei, 2004, for review of graphic organizers; and for strategy instruction for expository text types, De La Paz & MacArthur, 2003; Gajria et al., 2007). In summary, practices that have been associated with improved outcomes for students include students' capacity to use strategies to figure out the meaning of unknown words, to access prior knowledge in order to inform the text being read, to monitor their understanding during reading, to use or create graphic organizers, to generate questions about what is read, to be familiar with and understand narrative and expository text structures, and to use cooperative learning to increase engagement. In addition, explicit

instruction in reading comprehension strategies is associated with improvements in reading comprehension, particularly for students in secondary grades (Scamacca et al., 2007). A recent study (McKeown, Beck, & Blake, 2009) suggests that extended well-constructed conversations about the content of text may be as or more effective than strategy instruction. We interpret CSR as a multicomponent strategy that addresses both content learning through teacher- and student-led conversations as well as strategic enhancements.

Collaborative Strategic Reading: Review of Previous Studies

Early studies of CSR focused on evaluating effectiveness within science and social studies content area instruction. In one such study (Klingner, Vaughn, & Schumm, 1998), CSR was taught to intact, heterogeneous fourth-grade classes for 45 minutes per day during an 11-day Florida state history unit. The comparison group of intact classes received instruction reflective of the school's typical practice. Students in the CSR group made greater gains in reading comprehension and equal gains in content knowledge. To determine whether these findings would be upheld within science instruction, fifth graders were provided with CSR instruction for 30 to 40 minutes per day, 2 to 3 days per week, over a 4-week period during science (Klingner & Vaughn, 2000). Students frequently engaged in verbal discourse that supported vocabulary and content knowledge development; in addition, students made gains in target vocabulary over time. In a subsequent quasi-experimental study with fourth-grade teachers, teachers participating in the treatment condition were provided CSR training and in-class demonstrations with a comparison group of teachers continuing typical practice instruction. On a norm-referenced measure of reading comprehension, students in the CSR group outperformed students in the typical practice comparison group (Klingner, Vaughn, Argüelles, Hughes, & Ahwee, 2004), and when third-grade teachers received either CSR or partner reading training, students in teachers' classes that participated in CSR and partner reading groups performed equally well on tests of oral reading rate and accuracy as well as reading comprehension (Vaughn, Chard et al., 2000), providing additional evidence for the use of CSR with upper elementary students.

There have also been two studies of CSR at the middle school level. In one study, researchers developed a computer-adapted version of CSR (Kim et al., 2006) used with sixth- through eighth-grade students with learning disabilities who were randomly assigned to either a computer-based CSR intervention or typical school practice comparison group. On a norm-referenced measure of passage comprehension, students in the CSR group outperformed students in the comparison group. CSR was also used as one of several intervention practices designed to enhance overall schoolwide reading

comprehension (Bryant et al., 2000). Students demonstrated gains on word identification but not reading comprehension.

Through a series of studies conducted over a decade, CSR has been developed, implemented, and evaluated using quasi-experimental and descriptive research designs. However, the efficacy of CSR has not previously been determined through a randomized control trial and has been inadequately studied with older students with reading difficulties. The purpose of this study was to determine the efficacy of CSR by implementing a rigorous randomized control trial by working with schools to randomize students to classes and then classes to treatment or "business as usual" conditions during English/language arts/reading classes in seventh and eighth grades. This procedure allowed us to determine the effects of CSR while controlling for the potentially large effect of the teacher. We were also interested in the influence of prior strategy knowledge on treatment effects. We expected that strategy knowledge might moderate comprehension outcomes because prior research suggests that strategy knowledge has an increasingly significant effect on comprehension as students progress through the grades and begin reading more complex texts (Wilson & Rupley, 1997). Wilson and Rupley (1997) used structural equation analytical methods to evaluate within- and cross-grade relations among background knowledge, strategy knowledge, and comprehension of both narrative and expository texts and found that strategy knowledge dominates students' cognitive processes by Grade 6.

Method

Research Design and Questions

We conducted a randomized control trial in two sites (Colorado, Texas) by working with schools to randomize students to classes and then classes to treatment or comparison conditions. We have used this design in previous studies (Vaughn, Wanzek, Woodruff, & Linan-Thompson, in press). We identify several benefits of this randomized block study design, including: (a) Teacher effects are systematically controlled; (b) the sample size of teachers implementing the treatment is smaller since randomization is at the class level, thus allowing for adequate support of teachers to improve quality of implementation; and (c) controlling for contamination that may occur when randomization is at the teacher level and teachers work in teams at the middle schools in the districts in which we work. We addressed the following primary research question: What is the efficacy of CSR with adolescent readers when implemented by well-trained and supported novice CSR implementing teachers (unfamiliar with teaching CSR) compared with a well-documented school-implemented comparison group on reading comprehension? We were secondarily interested in whether prior knowledge of

metacognitive strategies would moderate the effect of CSR on reading comprehension. We also addressed whether CSR was effective for a subgroup of students identified as struggling readers.

Participants and Setting

We implemented this study in three school districts (two near urban and one urban). All districts reflected a diverse student population who were provided daily English/language arts/reading instruction. Teacher and student participant information is provided in the following.

Teachers. The 17 participating teachers (3 males and 14 females) provided seventh- and eighth-grade students in this study with English/language arts/reading instruction on a daily basis. All teachers held a bachelor's degree, with 8 teachers possessing a master's degree. One teacher was pursuing a doctoral degree during the study period. Teaching experience ranged from 1 to 35 years (mean = 9.5; median = 8.5), with 11 teachers holding multiple certifications (e.g., English/language arts, special education, English as a second language).

Students. Students in seventh and eighth grades at six middle schools were randomly assigned to the 61 classes of English/language arts/reading. The 61 classes were randomly assigned within teacher to either treatment or comparison condition. In other words, students were randomly assigned to class and then classes were randomly assigned within teacher. There were 27 comparison and 34 treatment classes (for teachers with an odd number of classes, the additional class was assigned to the treatment condition) and a total of 866 students in the initial analysis sample (student cases with missing cluster or grouping data or with missing data on *all* model-related observed scores were excluded for analysis). Patterns of missing data varied by outcome, so actual sample sizes differed by analysis. For the primary comparison (main effects on the Gates-MacGinitie [MacGinitie, MacGinitie, Maria, & Dreyer, 2004]), the sample size was 782, with 382 students in the comparison condition and 400 in treatment. Of the students, 11% were identified as experiencing reading difficulty, as defined by failing the previous year's state-developed reading assessment *and* scoring at least one standard deviation below the standardized mean on the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999). When passing the relevant state-level high-stakes reading test was used as the standard, 37% were identified as having a reading difficulty. We analyzed data for both groups of struggling students. Students' average age in the CSR group was 13.9 years and in the typical practice (TP) group was 13.7 years. See Table 1 for student demographic information. Group comparability (i.e., the effectiveness of randomization) was confirmed by comparing pretest scores by group on each of the three outcome measures

Table 1
Student Demographics and Pretest Means and Differences for Initial Analysis Sample

	CSR		TP	
	n	%	*n*	%
Gender				
Male	199	43.4	188	46.1
Female	170	37.1	166	40.7
Ethnicity				
Anglo	172	37.6	136	33.3
African American	13	2.8	13	3.2
Hispanic	178	38.9	193	47.3
Asian	2	0.4	9	2.2
Native American	4	0.9	3	0.7
Free or reduced lunch	194	42.4	189	46.3
English language learners	10	2.2	11	2.7
Special education	56	12.2	28	6.9

Note. CSR = collaborative strategic reading; TP = typical practice.

(Gates-MacGinitie, AIMSweb Reading Curriculum Based Measure [AIMSweb], and Test of Silent Reading Efficiency and Comprehension [TOSREC]). For Gates and AIMS, no significant differences were found at pretest (t = –0.562 and 0.997, respectively; Table 4). For TOSREC, the difference between groups at pretest was significant (t = –2.212).

Procedures

Professional development and teacher support. Teachers were trained by the researchers to implement CSR in their treatment classes and were provided procedures for continuing with "business as usual" in their control classes. Teachers were provided an initial training in CSR during a 3-day (6 hours per day) professional development on implementing the treatment practices. This training was provided by the same research team at both sites. The 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 control 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. Each teacher was provided with all necessary materials to implement the treatment, including sample lessons, examples of reading materials, and overheads. We also provided subsequent professional development on three occasions spread throughout the

implementation year (approximately 1.5 hours per meeting) to enhance and refine implementation and use of CSR in their classrooms.

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. Teachers participated in meetings with their research support person during implementation. They were provided feedback about their instruction based on observations and provided opportunities to ask questions, problem-solve, and plan for future lessons. Furthermore, depending on their needs, teachers could request on-site modeling of lesson components, help in finding reading materials, or various other types of additional support throughout the duration of the intervention.

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 class. Teachers were asked to implement the intervention for 50 minutes a day, 2 days a week, for approximately 18 weeks. Teachers reported that the number of sessions they implemented ranged from 24 to 48 sessions. Typical instruction was provided for students who were randomly assigned to comparison classes. Teachers also reported on the percentage of class time they spent doing a variety of language arts activities (e.g., phonics, reading comprehension, reading/fluency, and writing), components of reading comprehension (e.g., brainstorming, main idea, summarization), and grouping configurations (e.g., whole class, small group, partner, one-to-one). One of the critical elements of the design that we believe makes the findings compelling is that students in both the treatment and comparison conditions covered the same curriculum material over the same period of time using the same English/language arts 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 control classes.

Collaborative Strategic Reading. The CSR intervention is comprised of four comprehension strategies that are used before, during, and after reading with the goal of increasing student text engagement and reading comprehension. In this study, we asked teachers to use expository passages for a minimum of 50% of CSR lessons. In addition, we asked teachers to model each strategy to students through extensive use of think-alouds with time for guided practice of each strategy with multiple opportunities for providing feedback to students. If students struggled with mastering the use of

a particular strategy, teachers were instructed to implement short lessons to provide practice and review. Finally, teachers were taught to prepare a text for practicing CSR strategies by dividing the passage into three or four short sections of connected and coherent information. Following is a description of each CSR strategy.

Before reading a given passage, students are asked to engage in the first strategy—previewing. The previewing strategy encompasses four activities to build and activate prior knowledge and to motivate students' interest about the passage topic. First, the teacher introduces the passage topic and pre-teaches any proper nouns or specialized vocabulary that may be difficult for almost all students in the class. Second, students brainstorm what they already know about the topic. Third, students are taught to preview the passage and attend to text features such as headings and graphics to learn as much as possible in a very short period of time. Finally, students predict what they think they will learn from the passage. Students record their brainstormed ideas and predictions on their learning log.

During reading, students are guided to read the first section of the passage. As they read, students engage in the second and third strategies: Click and Clunk and Get the Gist. The Click and Clunk strategy is 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, students are instructed to identify "clunks," or breakdowns in understanding, and record them on their learning logs. After reading the section, students return to the clunks and use the following "fix up" strategies to find the meaning of the word in its context: (1) Reread the sentence without the word—think about what word meaning would make sense. (2) Reread the sentences before and after the clunk, looking for clues to determine the word meaning. (3) Identify key elements in the word (e.g., prefixes, suffixes, a known word part). (4) Identify word parts that may aid in understanding. Also during reading, students are instructed to use a practice called "Get the Gist," which is similar to writing the main idea. Students are taught to restate in their own words the most important point of a section of reading as a way of making sure they understood what they read and remembered what they learned.

After reading, students engage 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 are taught to write three levels of questions. "Right there" questions are those with answers that can be found in one sentence. These questions help students remember facts and focus on the most important information. "Think and search" questions are 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 are taught to use facts from the passage to make inferential conclusions. Students generate and answer each type of question on their learning logs. Finally, students are taught to write in their learning logs a summative statement that includes the most important ideas from the passage. Students are also asked to use the text to justify why these were identified as the most important ideas to remember.

Collaborative grouping. After students develop proficiency using the strategies (4–6 weeks), the teacher assigns them to cooperative learning groups of four to five students. During the cooperative groups, students are actively involved and have multiple opportunities to contribute to the group's understanding of text. Students are assigned roles and serve as critical leaders and collaborators in the effective functioning of the group and the implementation of CSR strategies. For CSR to function smoothly, at least four roles are essential (leader, clunk expert, gist expert, and question expert). Other roles (encourager and timekeeper) are assigned as needed. During a 40- to 50-minute session, students work 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 minutes 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 using the instructional practices. Each member of the group has an assigned role and learns to lead the group through previewing, gist, summary, and so on. Students use their learning logs to record previews, gists, clunks, and summaries. The teacher moves 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.

Observation of treatment and comparison classes. Two researchers conducted observations in treatment and comparison (business as usual) classes to ensure fidelity of treatment and to determine if there was any contamination of instruction in the comparison classes. Procedures used to establish level of intervention fidelity and evidence of contamination of instruction are described next.

Fidelity. Implementation fidelity was monitored by two researchers at each study site during four preannounced observations evenly distributed over the course of intervention delivery. To establish interrater agreement, researchers observed and independently coded in-person observations. Afterwards, researchers resolved disagreements in coding. On subsequent interrater agreement observations, the goal of 90% was achieved. The fidelity measure was developed to correspond with the critical elements of the

treatment and consisted of two sections—procedural fidelity and global ratings of quality. In the procedural fidelity section, teachers were assigned a score for each CSR component ranging from 1 (low alignment with CSR procedures) to 4 (very high alignment with CSR procedures). If an instructional component was not observed and not expected (i.e., the teacher modeled the preview strategy for the first time, so click and clunk would not be expected), a score of "NO" was recorded. If an instructional component was not observed but was expected (i.e., the teacher skipped the brainstorming step), a score of 1 was assigned. CSR instructional components that were assigned a score included brainstorming, predicting, identifying clunks, using fix-up strategies, get the gist, ask and answer questions, and review.

Observers also rated teachers on a set of three global items scored on a scale of 1 (low) to 7 (high) related to (a) quality of teacher instruction, (b) quality of teacher's classroom management, and (c) quality of teacher's CSR implementation. Scores were calculated by taking the average ratings of each teacher across the procedural fidelity and global items of the fidelity measure. In addition to the use of the fidelity measure, researchers took detailed field notes related to student and teacher behaviors considered critical to high-quality implementation of CSR. Teacher behaviors included items such as monitoring student work, providing effective feedback, and lesson pacing. Student behaviors included items such as level of participation in small groups and use of learning logs.

The mean number of sessions teachers' implemented was 34.62 (*SD* = 7.37); range 24 to 48; median number of sessions implemented was 38. The number of sessions implemented along with fidelity scores was included as potential mediating variables in early models. Including these variables diminished model fit significantly, thus, they were not included as mediators in the final (reported) model.

Table 2 presents the mean and standard deviations of scores for the treatment classes and Table 3 for the comparison classes at each of the four fidelity data collection points. Among treatment classes, while procedural fidelity scores were highest at Observation 1, the average procedural fidelity remained relatively stable across the four observations. Global quality of overall instruction and CSR implementation scores peaked at Observation 3, while classroom management scores peaked at Observation 2. Still, the average values for instructional quality were relatively stable across observations. Students worked in small groups during 55 of the 64 total fidelity observations, indicating that more often than not, students were engaged in group work during CSR instruction. Overall, these data suggest that teachers consistently provided moderate quality CSR instruction over time and provide adequate evidence that the procedures were implemented to an extent sufficient to attribute group differences to the implementation of CSR.

Table 2
Mean and Standard Deviations for Fidelity Observations in Collaborative Strategic Reading (CSR) Classes

	Observation 1		Observation 2		Observation 3		Observation 4	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Quality of instruction	4.82	1.19	4.76	1.25	5.18	1.38	4.63	1.67
Quality of classroom management	4.53	1.46	5.12	1.58	4.82	1.33	4.44	1.82
Quality of CSR implementation	4.06	1.56	4.53	1.55	4.65	1.84	4.19	1.64
Procedural fidelity	3.01	0.61	2.85	0.72	2.71	0.67	2.52	0.57

Note. Quality of instruction, quality of classroom management, and quality of CSR implementation measured on a 7-point scale, and procedural fidelity measured on a 4-point scale.

Table 3
Mean and Standard Deviations for Observations in Typical Instruction Classes

	Observation 1		Observation 2		Observation 3		Observation 4	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Quality of instruction	4.81	1.17	4.60	1.40	4.76	1.30	4.58	1.31
Quality of classroom management	4.94	1.18	4.60	1.64	4.76	1.64	5.00	1.48

Note. Quality of instruction and quality of classroom management measured on a 7-point scale.

Evidence of contamination. We were interested in determining whether there was contamination between treatment and comparison classes. To determine this, we relied on information provided by the teachers coupled with information from four observations of the comparison classes over the course of the study. As part of typical practice implemented in schools prior to this study, students received comprehension strategy instruction that was similar to CSR in some ways. For example, some teachers would stop periodically during a reading aloud of text and ask students to write a main idea statement, which has some similarities to writing a gist statement. Also, students were observed previewing text and using context clues and knowledge of morphology to determine meanings of unknown words in text, all of which are strategies that resonate with some components of CSR.

To address possible areas of contamination, teachers met with research support staff early and frequently during the study to help differentiate instruction provided in CSR and comparison classes. In addition, during each booster session, we clarified specific points related to contamination

of the comparison group. In other words, we addressed the question, "What is allowed and what is not allowed in typical practice classes?" Whereas versions of certain reading strategies were preexisting within some typical practice classes, through proactive training by research staff to implement the CSR strategies with high fidelity and to reduce instances of contamination, differential instruction for treatment students and comparison students with respect to the target instructional practices was achieved. Furthermore, since randomization was at the class level, if there were contamination it would reduce the impact of the treatment, thus providing a more rigorous test of the treatment.

Measures

All student measures were administered to students by trained research personnel who were blind to students' condition (treatment or comparison). All student measures were administered prior to treatment and immediately following treatment.

Gates-MacGinitie Reading Test (4th edition; Gates & MacGinitie, 2000). The Gates-MacGinitie Reading Test, Comprehension subtest, is a timed, group-administered assessment of reading comprehension. It consists of expository and narrative passages ranging in length from 3 to 15 sentences. Each passage is read silently. Students 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 test during the 35-minute time limit. Internal consistency reliability ranges from .91 to .93 and alternate form reliability is reported as .80 to .87. This test was administered at the beginning and the end of the year.

AIMSweb Reading Curriculum Based Measure (2007). The AIMSweb maze assessment is a multiple-choice close task that is completed while reading silently. Following the initial intact sentence, every seventh word is replaced with three words in parentheses. Students choose the word that fits in the sentence. Technical adequacy was obtained through rigorous design requirements, including adherence to readability formulas and acceptable Lexile scores within a range acceptable for each grade level (Howe & Shinn, 2002). The AIMSweb maze assessment has an alternate-form reliability ranging from .84 to .90 in Grades 6 through 8. Readability correlations are moderate to high, ranging between .78 and .98, with a median correlation of .90. This test was administered at the beginning and the end of the year.

Metacomprehension Strategy Index (Schmitt, 1988). The Metacomprehension Strategy Index (MSI) was developed to evaluate students' knowledge of reading strategies and has been used in several studies (e.g.,

Lonberger, 1988). The version used in the current study is a group-administered 20-item, four-option, multiple-choice measure that requires students to choose the best strategy for use before, during, and after reading. Strategies include prediction, previewing, purpose setting, self-questioning, using background knowledge, summarizing, and applying fix-up strategies. Internal consistency reliability (Cronbach's alpha) in our sample was 0.69 at pretest and 0.71 at posttest. This is somewhat lower than previously published reliability (e.g., Lonberger, 1988), which is possibly due to the fact that our sample is reflective of local populations rather than national populations (as a norming sample would be).

Test of Word Reading Efficiency (Torgesen et al., 1999). The TOWRE measures students' ability to read words out of context. It includes two separate, individually administered, timed assessments. Sight Word Efficiency measures students' ability to recognize common words quickly, while Phonemic Decoding Efficiency measures students' ability to sound out words quickly and accurately. In each subtest, students are provided a list of words or nonwords that are in order of increasing difficulty. The raw score is determined by the number of words or nonwords read in 45 seconds. Internal consistency exceeds .95 for both subtests. This test was administered only at the beginning of the year. It was used to identify struggling readers.

Test of Silent Reading Efficiency and Comprehension (Wagner, Torgesen, Rashotte, & Pearson, 2010). This group-administered assessment measures a student's contextual reading ability during a timed silent reading task. Students receive a list of sentences that increase in difficulty. Sentence length ranges from 4 to 10 words. Students read each sentence silently and circle yes if the sentence is true and no if the sentence is not true. Students are provided 3 minutes to complete the assessment. The total number of correct items comprises the score. This test was administered at the beginning and the end of the year.

Implementation logs. At the end of each CSR class period, teachers reported the number of minutes spent engaged in CSR strategy instruction and/or practice and wrote notes about that day's instruction. Research support personnel retrieved logs from teachers at the end of every week. These data were used to determine frequency of use of CSR.

Results

Data Analysis Procedures

Quantitative analyses. Multilevel modeling in Mplus 5.1 was used to estimate the effects of treatment and the moderating influence of important

covariates. Multilevel models in Mplus belong to the class of structural equation models (SEM), which offers advantages not available in other analysis strategies (e.g., ANCOVA, hierarchical linear modeling). For example, SEM, and Mplus specifically, handles missing data using a direct full information maximum likelihood (FIML) estimator to compute a likelihood function for each case using all available data, which is more efficient than traditional approaches like list-wise deletion of cases with missing data or imputation of values (Enders & Bandalos, 2001). Also, because multilevel modeling analyzes covariance structures representing different levels of aggregation (e.g., individual and group levels), it is more appropriate than traditional approaches when data are clustered, whether by design (i.e., stratified sampling strategy) or circumstance (e.g., students in schools). Similarly, a multilevel structure in Mplus provides a flexible framework for analyzing the effects of covariates (and the possibility that they differ by group or by level of aggregation) by treating Level-2 dependent variables as latent (i.e., estimated at Level 1). Estimates of model fit are provided as well as a means of evaluating a given model's accuracy in representing patterns among the observed data and as a tool for comparing nested models and evaluating statistical significance (Bovaird, 2007; Mehta & Neale, 2005). Comparison of nested models involves constraining the treatment and comparison groups as equal on parameters of interest and comparing the fit of this constrained model to the fit of a fully specified, unconditional model. If both models fit the data similarly well across groups, they are considered comparable. Conversely, constraints resulting in less adequate fit across groups suggest significant group differences. The difference in χ^2 ($\Delta\chi^2$) between nested models provides the test statistic (Bollen & Curran, 2006), and the difference in degrees of freedom (Δ_{df}) is used to evaluate its significance.

In the present study, *teacher* was treated as a stratum for purposes of assignment, and *classes* (both treatment and comparison) were randomly assigned within teachers. Analytically, this represents a randomized block design with teachers as the blocking variable (Raudenbush, 1997) and students nested in classes. Blocking on teachers reduces the within-teacher heterogeneity and increases overall power, leading to less bias in tests of treatment main effects (Cohen, 1991). A pretest score (cluster-level covariate) was also included in the model as a means of minimizing the conditional group-level variance and further increasing precision and power (Bovaird, 2007). In Mplus, this represents a two-level analysis with complex sampling. We identified teacher as a stratum using the STRATIFICATION command in Mplus. Classes were represented as clusters (CLUSTER command), which define levels in a multilevel model. In the unconditional (i.e., no moderating covariates) student-level model (i.e., %within% in Mplus), posttest scores were regressed on the corresponding grand-mean centered pretest values. Posttest means were modeled as latent factors on the %between% -classes model. Treatment condition was modeled using the multiple groups option

in Mplus, which allowed for formal tests of statistical significance using the nested models comparison described previously. Cross-level interactions between MSI and condition were evaluated using the nested models approach.

Treatment Effects

Main effects were estimated for the Gates-MacGinitie, the AIMSweb maze, and the TOSREC according to the multilevel model described previously. The analyses were conducted with the entire sample and with the subsample of students identified at pretest as struggling readers (in the initial analysis sample, n = 46 in comparison and 49 in treatment) based on failure of the high-stakes state reading assessment and a pretest standard score of less than 85 (i.e., one standard deviation below the mean) on the TOWRE. We also estimated treatment for the group of students failing the state high-stakes test (n = 128 in the comparison and n = 125 in treatment). For the full sample, an unconditional multigroup, multilevel model was fit to estimate posttest class-level means (i.e., Level 2) conditioned on the student-level (Level 1) model and on the earlier described adjustments for clustering and stratification for each of the three outcomes. These models were saturated because there were as many parameters as values to fit (accordingly, they have a χ^2 of 0 and 0 degrees of freedom). The model-estimated (Level-2 latent) standard score average on the Gates-MacGinitie was 95.87 for comparison classes and 97.04 for the treatment conditions. This is equivalent to an (bias corrected Hedges) effect size of g = 0.12. The standard score estimates on AIMSweb were 93.42 and 92.53 for the comparison and treatment conditions, respectively. On TOSREC, raw scores were analyzed. The estimated posttest scores for comparison and treatment were 28.75 and 29.27. Findings are presented in Tables 4 and 6.

To evaluate the statistical significance of these differences, the fit of each of the three unconditional models (i.e., for Gates-MacGinitie, AIMSweb, and TOSREC) was compared to its respective conditional model with posttest estimates constrained as equal across conditions (i.e., the nested models comparison described earlier). The difference in χ^2 ($\Delta\chi^2$) between unconditional and constrained models was used as the test statistic (Bollen & Curran, 2006). For Gates-MacGinitie, this value was 9.91 ($p < .01$), suggesting that participants in CSR outperformed nonparticipants when the effects of clustering and pretreatment differences were explicitly modeled. Group differences on AIMSweb ($\Delta\chi^2$ = 1.13, Δ_{df} = 1) and TOSREC ($\Delta\chi^2$ = .41, Δ_{df} = 1) were not statistically significant (Table 6).

Results for the sample of low-achieving students (with TOWRE as a selection criterion) were similar to those for the total sample (Tables 5, 7). The model-derived posttest score on the Gates-MacGinitie was 87.66 for CSR participants, about 3.14 standard score points greater than initially

Table 4

Descriptive Statistics for All Students in the Initial Analysis Sample

	Pretest						Posttest			
	Comparison		Treatment				Comparison		Treatment	
	M	*SD*	*M*	*SD*	Test of Differences *t*	*p*	*M*	*SD*	*M*	*SD*
Gates-MacGinitie	95.68	13.4	96.35	13.7	−0.562	0.574	95.48	13.4	97.13	13.6
AIMSweb	92.64	12.1	91.91	10.8	0.997	0.321	93.46	11.0	92.92	11.0
Metacomprehension Strategy Index	9.80	3.56	10.15	3.69	N/A	10.01	3.62	10.48	3.91	
Test of Silent Reading Efficiency and Comprehension (TOSREC)	23.49	7.36	24.67	7.17	−2.212	0.027	29.03	7.68	29.22	7.66
Test of Word Reading Efficiency (TOWRE)-Sight Words	90.17	7.40	90.59	8.38	N/A		N/A		N/A	
(TOWRE)-Phonemic Decoding	91.53	11.3	91.20	10.85	N/A		N/A		N/A	

Table 5
Descriptive Statistics for Subsample of Student: Low Word Reading and Low Comprehension

	Pretest *M* (*s*)		Posttest *M* (*s*)	
	Comparison	Treatment	Comparison	Treatment
Gates-MacGinitie	84.37 (10.0)	85.62 (11.2)	84.25 (9.08)	87.74 (9.95)
AIMSweb	85.81 (8.75)	85.13 (8.01)	86.30 (7.97)	86.60 (7.37)
Meta-comprehension Strategy Index	9.30 (3.56)	9.06 (3.58)	9.63 (3.51)	9.73 (4.00)
Test of Silent Reading Efficiency and Comprehension (TOSREC)	18.52 (4.60)	20.30 (5.89)	23.87 (6.34)	24.83 (5.76)
TOWRE – Sight Words	84.48 (5.46)	83.76 (4.62)	N/A	N/A
TOWRE – Phonemic Decoding	81.57 (5.52)	82.55 (5.57)	N/A	N/A

struggling students in the comparison. Though not statistically significant (p = .066), the difference represents an effect size g = 0.36 (about 21% of the 15-point standard deviation used by the Gates-MacGinitie), an effect with considerable practical significance (Rossi, Lipsey, & Freeman, 2004). The results for the state-test-only struggling group (Table 8) were similar; the treatment and comparison groups differed at posttest by about 2.5 standard score points (p = .063; g = .35).

The effect of CSR on reading comprehension (the Gates-MacGinitie) may also depend in part on students' knowledge of strategic reading strategies prior to the program's onset. When scores on the Metacomprehension Strategy Index were modeled as a Level-1 covariate in the full sample, the parameter estimate in the comparison group was .169 (p = .176) compared to .452 ($p < .001$) in the treatment condition, a statistically significant difference ($\Delta\chi^2$ = 5.85, Δ_{df} = 1, p = .016). The interaction of MSI and treatment condition suggests that students with similar levels of metacognitive strategy knowledge at pretest experience better comprehension outcomes when given the opportunity or presented with the challenge of CSR. In other words, when we controlled for student performance on the MSI, students in the CSR group made greater gains than students in the comparison group.

Discussion

We designed this randomized control trial to provide a rigorous experimental study examining the effects of CSR on the reading achievement of middle grade students (seventh and eighth graders). The study was designed so that middle school English language arts teachers participated and their sections of classes were randomized to treatment and control conditions providing a test of CSR controlling for teacher effects. Students in the treatment

Table 6
Level 1 Model-Adjusted Means for the Full Sample

Values Adjusted for Level 1 Model	Comparison	*SE*	Treatment	*SE*	$\Delta\chi^2/\Delta_{df}$	*p*
Gates-MacGinitie	95.87	.534	97.04	.535	9.91/1	.002
AIMSweb	93.42	.447	92.53	.433	1.13/1	.287
Test of Silent Reading Efficiency and Comprehension (TOSREC)	28.75	.61	29.27	.35)	.41/1	.522

Table 7
Level 1 Model-Adjusted Means for Struggling Readers Based on State-Level Test and Test of Word Reading Efficiency (TOWRE)

Values Adjusted for Level 1 Model	Comparison	*SE*	Treatment	*SE*	$\Delta\chi^2/\Delta_{df}$	*p*
Gates-MacGinitie	84.52	1.33	87.66	.961	3.38/1	.066
AIMSweb	86.44	1.42	86.32	.84	.01/1	.920
Test of Silent Reading Efficiency and Comprehension (TOSREC)	24.25	.96	24.63	.68	.15/1	.699

Table 8
Level 1 Model-Adjusted Means for Struggling Readers Based on State-Level Test

Values Adjusted for Level 1 Model	Comparison	*SE*	Treatment	*SE*	$\Delta\chi^2/\Delta_{df}$	*p*
Gates-MacGinitie	90.15	1.06	92.58	1.02	3.46/1	.063
AIMSweb	90.64	1.55	90.76	1.43	.01/1	.920
Test of Silent Reading Efficiency and Comprehension (TOSREC)	25.82	.49	26.83	.71	.25/1	.617

classes outperformed students in the comparison classes on a standardized reading comprehension measure. Considering the rigor of the study and the challenge of positively impacting reading comprehension with older readers, we consider this finding noteworthy. We also examined students' outcomes on a fluency measure with no significant effects for the treatment condition. Regardless of whether students were assigned to the treatment or control condition, they demonstrated similar fluency gains at the end of the year. This finding was not surprising to the research team as the focus of CSR is on improving how students think about and interact with text without

guidance for how to improve students' speed of reading. Previous research with CSR suggests positive effects for low readers and students with learning disabilities in reading (Klingner et al., 2004), thus, we hypothesized that students with low reading skills in the CSR treatment would perform significantly better on reading comprehension than comparison students. Though results were not statistically significant (p = .066), the effect size (g = 0.36) has considerable practical significance and future studies examining the effects on low readers are needed. We also examined the extent to which a measure of metacognitive reading strategies (MSI) would be associated with improved outcomes in reading comprehension. Students who participated in CSR and demonstrated similar scores on metacognitive strategies in reading scored significantly better on reading comprehension than comparison students.

We interpret the findings from this study as suggestive that CSR is a feasible and effective practice that can be readily integrated into reading and language arts instruction with positive impact. We are encouraged about the potential effectiveness of this practice because the positive findings from this efficacy study resulted from treatment implementation conditions that are readily replicable. The treatment was only provided twice a week for approximately 18 weeks, integrating the instructional practices into the curriculum rather than replacing the curriculum, suggesting that in a relatively brief period of time with minimal practice, significant effects were realized. We recognize that schools have limited resources and time and are unlikely to consistently implement instructional practices that require considerable change on the part of teachers and resources on the part of schools (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). In this study, CSR was implemented with 18 hours of professional development—an amount of time that most districts would consider feasible. CSR also requires minimal resources since teachers use the text materials that they typically use and no additional materials for students are needed other than copying learning logs for students to record their responses. Furthermore, CSR is integrated into teachers' existing practices, requiring no curriculum changes. We think that there is considerable evidence that instructional practices need to be feasible and aligned with current curriculum for teachers to use them (Garet, Porter, Desimone, Birman, & Yoon, 2001; Klingner, Arguelles, Hughes, & Vaughn, 2001; Vaughn, Hughes, Schumm, & Klingner, 1998).

Since CSR is a multicomponent intervention, it is interesting to speculate about which of the components might be associated with impact and which components may be less influential. For example, since most English/language arts teachers teach students to get the main idea, which is very similar to how we teach "get the gist," perhaps this element of CSR is no longer necessary and additional time could be spent on teaching students to generate and respond to questions. It is also possible that the collaborative group structure that promotes student engagement and discourse about text is

the essential element and using collaborative groups to enhance text comprehension without teaching the comprehension strategies of CSR would be sufficient. The study does not address which components of CSR are associated with differential impact, but future research could experimentally manipulate and isolate the impact of various components determining their relative effects.

Findings from this study compare favorably with those from a recent study examining the effectiveness of several different comprehension programs with fifth-grade students (James-Burdumy et al., 2009). In this study, schools were randomly assigned to one of four reading comprehension interventions or a control condition. Teachers in the treatment conditions received professional development by the curriculum developers. Findings indicated no significant effects for reading achievement or comprehension for students participating in the treatment. For one curriculum the effects were statistically significant in the negative direction, meaning that the comparison students outperformed those in the treatment. Findings from this study, as well as others with older students (e.g., Vaughn, et al., 2010), revealed that significantly improving reading comprehension is not an easy result to obtain from experimental studies.

Limitations

The randomized design of the study is one of its strengths and the within-teacher design represents an innovative use for controlling teacher and school effects. However, estimates of program effect will be biased (i.e., underestimated) to the extent that undetected or unaddressed treatment contamination was present. While there was movement across treatment conditions during the school year, it was fairly limited, occurred in both directions (i.e., from treatment to comparison and from comparison to treatment), and apparently nonsystematic, therefore, for purposes of analysis, students maintained their initial assignment to condition regardless of subsequent movement.

Future Research

All of the teachers in this study were novice implementers of CSR. While most of the instructional practices are not "novel" and teachers were familiar with the idea of before, during, and after reading strategies—most were unfamiliar with the procedures we taught during professional development for modeling aloud how to think about text and how to consider and compare thinking about main idea and summaries with more expert models. We also realized that few teachers used small groups prior to participating in our study, and when they did, these small groups were not executed with procedures following cooperative grouping, including giving specified roles to students and requiring interactive work around text and responses. For this

reason, we are interested in the extent to which more experienced users of CSR might be associated with student achievement in reading. We have asked all participating teachers in this study to continue for a second year in which they continue to implement CSR with a nonoverlapping sample of students. We also think that studies that investigate the relative impact of components of CSR would be valuable.

While we interpret the effects on a standardized reading comprehension measure as meaningful and the high feasibility of implementation replicable, we are also aware that the effects are small. We think it is worthwhile to ask whether these multicomponent approaches like CSR are the most impactful ways of influencing comprehension or whether interventions that focus more on vocabulary building and/or background knowledge might be even more efficacious.

Note

This research was supported by grant R305A080608 from the Institute of Education Sciences, U.S. Department of Education. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Institute of Education Sciences or the U.S. Department of Education.

References

ACT, Inc. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading*. Iowa City, IA: Author.

AIMSweb Reading-Curriculum Based Measurement. (2007). Retrieved from http://www.aimsweb.com/products/cbm.php.

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

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*, 289–328.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. New York, NY: John Wiley.

Bovaird, J. A. (2007). Multilevel structural equation models for contextual factors. In T. D. Little, J. A., Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 149–182). Mahwah, NJ: Lawrence Erlbaum.

Bryant, D. P., Vaughn, S., Linan-Thompson, S., Ugel, N., Hamff, A., & Hougen, M. (2000). Reading outcomes for students with and without reading disabilities in general education middle-school content area classes. *Learning Disability Quarterly*, *23*, 238–252.

Cohen, P. (1991). A source of bias in longitudinal investigations of change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 18–30). Washington, DC: American Psychological Association.

De La Paz, S., & MacArthur, C. (2003). Knowing the how and why of history: Expectations for secondary students with and without learning disabilities. *Learning Disabilities Quarterly*, *26*, 142–154.

Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research*, *61*, 239–264.

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

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*, 430–457.

Flavell, J. H. (1992). *Cognitive development* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Lipsey, M. W. (2000). Reading differences between low-achieving students with and without learning disabilities: A meta-analysis. In R. Gersten, E. P. Schiller, & S. Vaughn (Eds.), *Contemporary special education research* (pp. 81–104). Mahwah, NJ: Lawrence Erlbaum.

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

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*, 915–945.

Gates, A. I., & MacGinitie, W. H. (2000). *Gates-MacGinitie Reading Tests* (4th ed.). Boston, MA: Riverside.

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.

Harris, K. R., & Pressley, M. (1991). The nature of cognitive strategy instruction: Interactive strategy construction. *Exceptional Children*, *57*, 392–404.

Heilman, A. W., Blair, T. R., & Rupley, W. H. (1998). *Principles and practices of teaching reading* (9th ed.). Columbus, OH: Merrill/Prentice Hall.

Howe, K. B., & Shinn, M. M. (2002). *Standard reading assessment passages (RAPs) for use in general outcome measurement: A manual describing development and technical features*. Eden Prairie, MN: Edformation.

James-Burdumy, S., Mansfield, W., Deke, J., Carey, N., Lugo-Gil, J., Hershey, J.,. . . Faddis, B. (2009). *Effectiveness of selected supplemental reading comprehension interventions: Impacts on a first cohort of fifth-grade students* (NCEE 2009-4032). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Jenkins, J. R., Heliotis, J., Stein, M. L., & Haynes, M. (1987). Improving reading comprehension by using paragraph restatements. *Exceptional Children*, *54*, 54–59.

Jiménez, R. T., Garcia, G. E., & Pearson, P. D. (1995). Three children, two languages, and strategic reading: Case studies in bilingual/monolingual reading. *American Educational Research Journal*, *32*, 67–97.

Jiménez, R. T., García, G. E., & Pearson, P. D. (1996). The reading strategies of bilingual Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly*, *31*, 90–112.

Kamil, M. L. (2003). *Adolescents and literacy: Reading for the 21st century*. Washington, DC: Alliance for Excellent Education.

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 No. 2008-4027). Washington, DC: National Center for

Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practiceguides/adlit_pg_082608.pdf

Kim, A.-H., Vaughn, S., Klingner, J. K., Woodruff, A. L., Klein, C., & Kouzekanani, K. (2006). Improving the reading comprehension of middle school students with disabilities through computer-Assisted collaborative strategic reading (CACSR). *Remedial and Special Education*, *27*, 235–248.

Kim, A.-H., Vaughn, S., Wanzek, J., & Wei, S. (2004). Graphic organizers and their effects on the reading comprehension of students with LD: A synthesis of research. *Journal of Learning Disabilities, 37*, 105–118.

Klingner, J. K., Arguelles, M. E., Hughes, M. T., & Vaughn, S. (2001). Examining the schoolwide "spread" of research-based practices. *Learning Disability Quarterly*, *24*, 221–234.

Klingner, J. K., & 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., Vaughn, S., Argüelles, M. E., Hughes, M. T., & Ahwee, S. (2004). Collaborative strategic reading: "Real world" lessons from classroom teachers. *Remedial and Special Education*, *25*, 291–302.

Klingner, J. K., Vaughn, S., & Boardman, A. (2007). *Teaching reading comprehension to students with learning difficulties*. New York, NY: Guilford.

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.

Lonberger, R. B. (1988). *The effects of training in a self-generated learning strategy on the prose-processing abilities of fourth and sixth graders* (Unpublished doctoral dissertation). State University of New York, Buffalo.

Maccini, P., Gagnon, J. C., & Hughes, C. A. (2002). Technology-based interventions for secondary students with learning disabilities. *Learning Disability Quarterly*, *25*, 247–262.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2002). *Gates-MacGinitie Reading Tests* (4th ed.). Itasca, IL: Riverside Publishing.

Mastropieri, M. A., Scruggs, T. E., Bakken, J. P., & Whedon, C. (1996). Reading comprehension: A synthesis of research in learning disabilities. *Advances in Learning and Behavioral Disabilities*, *10*, 201–227.

Mastropieri, M. A., Scruggs, T. E., & Graetz, J. E. (2003). Reading comprehension instruction for secondary students: Challenges for struggling students and teachers. *Learning Disabilities Quarterly, 26*(2), 103–116.

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

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations models. *Psychological Methods*, *10*, 259–284.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the sub-groups*. Washington, DC: U.S. Department of Health and Human Services, National Institute on Health.

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., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, *8*, 293–316.

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). London: Routledge.

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

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Mahway, NJ: Lawrence Erlbaum.

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

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, *2*, 173–185.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M. S., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). *Interventions for adolescent struggling readers: A meta-analysis with implications for practice*. Portsmouth, NH: RMC Research Corporation, Center on Instruction. Retrieved from http://www.centeroninstruction.org/files/Meta-analysis%20Struggling%20Readers1.pdf

Schmitt, M. C. (1988). The effects of an elaborated directed activity on the metacomprehension skills of third graders. In J. E. Readence and R. S. Baldwin (Eds.), *Dialogues in literacy research*. (pp. 167–189). Chicago, IL: National Reading Conference.

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

Swanson, H. L., Kehler, P., & Jerman, O. (2010). Working memory, strategy knowledge, and strategy instruction in children with reading disabilities. *Journal of Learning Disabilities*, *43*, 24–27.

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). *The Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.

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

Vaughn, S., Cirino, P. T., Wanzek, J., Wexler, J., Fletcher, J. M., Denton, et al. (2010). Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. *School Psychology Review, 39*(1), 3–21.

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., Hughes, M. T., Schumm, J. S., & Klingner, J. K. (1998). A collaborative effort to enhance reading and writing instruction in inclusion classrooms. *Learning Disability Quarterly*, *21*, 57–74.

Vaughn, S., Wanzek, J., Woodruff, T., & Linan-Thompson, S. (in press). Three-tier model for effective interventions. In D. Haager, S. Vaughn, & J. K. Klingner (Eds.), *Response to intervention*. Baltimore, MD: Brookes Publishing.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

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

Williams, J. P. (1998). Improving comprehension of disabled readers. *Annals of Dyslexia*, *48*, 213–238.

Williams, J. P. (2000). *Strategic processing of text: Improving reading comprehension for students with learning disabilities* (Report No. EDO-EC-00-8). Reston, VA: Council for Exceptional Children. (ERIC Document Reproduction Service No. ED 449596)

Wilson, V. L., & Rupley, W. H. (1997). A structural equation model for reading comprehension based on background, phonemic, and strategy knowledge. *Scientific Studies of Reading*, *1*, 45–63.

Wong, B. Y. L., & Jones, W. (1982). Increasing metacomprehension in learning disabled and normally achieving students through self-questing training. *Learning Disability Quarterly*, *5*, 228–240.

Yoon, K. S., Duncan, T., Lee, S. W., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues and Answers REL 33). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Manuscript received May 5, 2010
Final revision received September 7, 2010
Accepted January 28, 2011