Relationship Between Implementation of Collaborative Strategic Reading and Student Outcomes for Adolescents With Disabilities

Alison G. Boardman, PhD[1], Pamela Buckley, PhD[1], Sharon Vaughn, PhD[2], Gregory Roberts, PhD[2], Karla Scornavacco, PhD[1], and Janette K. Klingner, PhD[1]

Abstract

This study examines the interaction between the fidelity of implementation of a set of research-based strategies—Collaborative Strategic Reading (CSR)—and outcomes for students with mild to moderate disabilities using data from two nonoverlapping studies in middle school language arts and reading classrooms (Study 1) and middle school social studies and science classrooms (Study 2). The authors use a definition of fidelity that includes both the amount of CSR instruction delivered by teachers and the quality of implementation. Although there were no main effects for quality or amount of CSR instruction, in both studies there was an interaction effect between quality of implementation and special education status. The study used a within-groups design and multilevel analyses, and the results demonstrate that higher quality CSR instruction was associated with higher reading outcomes for students with disabilities. This finding was consistent across Study 1 and Study 2. Implications for practice and future research are discussed.

Research on instructional strategies is important in education, especially when implementation is subject to the real-world complexities of schools and classrooms. Such intervention studies further our knowledge of what works and advance best practices for helping students achieve (Crawford, Carpenter, Wilson, Schmeister, & McDonald, 2012). Demonstrating program impact, however, is difficult for many reasons, including poor or varying levels of implementation. For students with disabilities, issues of fidelity or the extent to which program participants "receive the full protocol as intended" (Lipsey, 1999, p. 641) may be especially important as many of these students require specialized instruction to make educational gains (Zigmond, Kloo, & Volonino, 2009). Thus, teachers who implement instructional programs with low quality or who pick and choose certain aspects of a model may not be providing instruction that includes those features essential for students with disabilities.

In terms of reading comprehension, several meta-analyses have outlined specific features of instruction that result in positive outcomes for students with learning disabilities. These include such components as explicit instruction in reading strategies; teacher modeling; use of multiple strategies before, during, and after reading; and opportunities to collaborate with peers (e.g., Edmonds et al., 2009; Faggella-Luby, Schumaker, & Deshler, 2006; Gajria, Jitendra, Sood, & Sacks, 2007; Scammacca et al., 2007; Scammacca, Roberts, Vaughn, & Stuebing, 2015; H. L. Swanson, 1999). Yet despite the growing evidence that reading strategies instruction increases reading comprehension outcomes, there is little research on the contribution of the quality of reading strategies instruction on those outcomes, especially for students with disabilities in general education settings. We hypothesize that instruction using a research-based model of reading strategies instruction, Collaborative Strategic Reading (CSR), has the greatest effect on improving reading performance when the model is implemented with fidelity.

The present study examines how fidelity of implementation of CSR is associated with reading outcomes for adolescents, including students with mild to moderate disabilities.

[1]University of Colorado Boulder, USA
[2]University of Texas at Austin, USA

Corresponding Author:
Alison G. Boardman, PhD, University of Colorado Boulder, School of Education, 249 UCB, Boulder, CO 80309, USA.
Email: alison.boardman@colorado.edu

We use data from two large studies of CSR implementation (see Note) to analyze the relation among multiple measures of fidelity and student reading comprehension outcomes.

We asked the following research questions and replicated analysis for two nonoverlapping studies to examine the stability of findings:

1. To what extent was CSR implemented as intended?
2. Is higher CSR instructional quality associated with increased student outcomes for adolescents in CSR classrooms and for a subgroup of students with disabilities?
3. Is the amount of CSR instruction related to student outcomes for adolescents in CSR classrooms and for a subgroup of students with disabilities?

CSR

CSR (Klingner, Vaughn, Boardman, & Swanson, 2012; Klingner, Vaughn, & Schumm, 1998) is a set of research-based strategies designed to improve reading comprehension, enhance students' content area learning, facilitate access to higher level texts, and promote student engagement in reading. While CSR benefits a variety of learners (Klingner et al., 1998; Klingner, Vaughn, Argüelles, Hughes, & Leftwhich, 2004; Vaughn et al., 2011), the model was designed to support struggling readers in mixed-ability classrooms, many of whom are students with disabilities.

CSR is theoretically grounded in cognitive psychology (Flavell, 1992) and combines elements of reciprocal teaching (Palincsar & Brown, 1984) and cooperative learning (Kagan, 1986). CSR includes before, during, and after reading strategies that are first introduced one at a time by the teacher. As students gain proficiency in strategy use, the strategies are combined into a full model in which small heterogeneous groups of students use cooperative learning and CSR structures to guide them through reading (for more information about CSR, see Klingner, Vaughn, et al., 2012). Students learn the text content along with metacognitive awareness and reading strategies. CSR can be integrated into content area instruction using both expository and narrative text, aligning well with new reading initiatives such as the Common Core State Standards (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010) that call for increased access to challenging expository text across the curriculum.

Despite the fit with current instructional expectations and the importance for struggling readers, some teachers find it challenging to implement multicomponent reading strategy models like CSR with high fidelity (e.g., Hilden & Pressley, 2007; Klingner et al., 1998; Klingner et al., 2004). Boardman and Woodruff (2004) found that perceived alignment with the curriculum and the standards measured in high-stakes assessments can influence fidelity. Difficulty finding appropriate texts as well as teachers' instructional decision making related to monitoring student work, providing the appropriate amount of explicit instruction, and scaffolding are additional challenges to fidelity (Hilden & Pressley, 2007). As O'Donnell (2008) pointed out, given that adopters inevitably adapt innovations based on local needs, more studies are needed "to guide future researchers in understanding how fidelity of implementation can be used to adjust or interpret outcome measures" (p. 33).

Previous CSR Research

CSR has been evaluated using experimental and quasi-experimental designs and has yielded positive outcomes for students with learning disabilities, students at risk for reading difficulties, English-language learners (ELL), and average and high-achieving students (Boardman, Klingner, Buckley, Annamma, & Jensen, 2015; Bryant et al., 2000; Klingner et al., 1998; Klingner et al., 2004; Vaughn et al., 2011). Several of these studies described the ways in which components of fidelity might influence implementation of CSR. Klingner et al. (2004) reported variation in both quality and amount of CSR instruction, with some teachers seeming to take up the strategies quickly and easily while other teachers did not implement the model as intended, even with support from researchers. The study involved fourth graders using CSR in social studies classrooms, and results demonstrated that higher student gains were associated with higher quality of CSR implementation for four out of the five teachers in the sample (Klingner et al., 2004). Hitchcock, Dimino, Kurki, Wilkins, and Gersten (2011) found that although teachers reported teaching CSR the recommended two to three times each week, observations of instruction indicated that only 20% of teachers were teaching all the intended CSR strategies and that more than 50% of the teachers were using fewer than half of the intended strategies (Hitchcock et al., 2011). The study yielded no significant treatment effect for CSR instruction in fifth-grade classes, and it is possible that the lack of a treatment effect for CSR may have been due to low fidelity.

More recently, fidelity measures reported in efficacy studies of CSR have used descriptive statistics to describe implementation differences between treatment and control conditions (Boardman et al., 2015; Vaughn et al., 2011). Vaughn and colleagues (2013), however, successfully used fidelity data in an inferential context to quantify the mediating relationship of random assignment and student outcomes. Each teacher taught two or more sections of language arts or reading, and sections were randomly assigned to the treatment or comparison condition (Vaughn et al., 2013). Results supported the authors' expectation that CSR reading strategies would be more prevalent in treatment classes than in comparison classes. They also found

that language arts teachers were less likely than reading teachers to use CSR-like strategies during instruction in their comparison classes, due presumably to reading teachers' more intensive training on effective reading instruction. That is, typical practice for reading teachers was more likely to include CSR-like strategies because they were more extensively trained in the use of effective instructional methods independent of the study in question. CSR was associated with a greater effect on student reading comprehension when implemented in language arts classrooms compared to reading classrooms. The relative absence of CSR-like practices in language arts comparison classrooms may have created a starker contrast in language arts classrooms between treated and untreated settings, leading to more detectable treatment effects. Each of the studies mentioned above (Hitchcock et al., 2011; Klingner et al., 2004; Vaughn et al., 2011; Vaughn et al., 2013) indicated that student outcomes may be related to fidelity of implementation. Additional research on CSR fidelity is needed to explore the relationships across studies and settings.

Defining Fidelity

The validity of intervention study findings is improved when we examine not only whether a program is effective (e.g., if the treatment group makes significantly greater gains than a control group) but also when we have an understanding of the relationship between how a program is implemented and participant outcomes (Crawford et al., 2012). Although there is no single definition, most agree that fidelity encompasses the degree to which a program is implemented as intended (Lipsey, 1999; O'Donnell, 2008). As described by Hulleman and Cordray (2009), "assessments of intervention fidelity involve the specification of a 'gold standard' or basis for comparison—a theory, model, or conception of the educational intervention—to which something is faithful" (p. 90). Thus, when teachers implement a model in the intended way, fidelity is high, and when high fidelity in one group is associated with better outcomes than in groups with lower fidelity, the improved outcomes may relate, at least in part, to the different levels of implementation.

Several criteria have been used to operationalize fidelity to a particular program (Dane & Schneider, 1998; Dusenbury, Brannigan, Falco, & Hansen, 2003; Mowbray, Holter, Teague, & Bybee, 2003; O'Donnell, 2008). These include *dosage* (the extent to which a program was delivered), *adherence* (the degree to which program components are delivered as prescribed), *quality of delivery* (related to teacher instruction), *participant responsiveness* (how well the intervention is received), and *program differentiation.* According to Dusenbury et al. (2003), program differentiation is the "degree to which elements which would distinguish one type of program from another are present or absent" (p. 240), which is how instruction in

a treatment group can be differentiated from instruction in a control group. Mowbray and colleagues (2003) classified these five criteria into two groups that are useful for studying the effect of research-based practices in real-world settings: *fidelity to structure* (e.g., adherence and dosage) and *fidelity to process* (such as the quality of delivery or program differentiation). For the present research, we used this bimodal definition to guide our study of fidelity of CSR and student outcomes.

Fidelity Research in Education

The emphasis on fidelity of implementation is a recent development in education research (O'Donnell, 2008), and demonstrating its relationship to student outcomes will become increasingly relevant to interventionists as they attempt to identify the potent features of a given intervention. For intervention researchers, greater insight on the fidelity/outcome link will be useful in protecting and characterizing the internal validity of efficacy research. For funders of educational research, theoretical and practical work on implementation fidelity will advance the knowledge base and contribute to more reliable findings, better practice, and ultimately better student outcomes.

Method

Research Design

This research reports findings from two CSR studies. The first (Study 1) was part of a multisite randomized control trial consisting of CSR implemented in Grades 7 and 8 language arts and reading classes. The second (Study 2) includes data from a randomized control trial of CSR implemented in Grades 6 through 8 social studies and science classes as part of a district initiative for accelerating reading comprehension districtwide. For the present analysis, we examined outcomes for CSR students in relation to teacher fidelity for Study 1, and then we replicated the analysis with a separate data set with the teachers and students in Study 2. Table 1 presents the similarities and differences between Study 1 and Study 2.

The samples for both studies included the study groups of students who received CSR instruction as part of the larger randomized control trials of CSR. Middle school classrooms, rather than students, were randomly assigned to a condition (i.e., the CSR group or the business-as-usual group) in which half of each teacher's sections were randomly assigned to the CSR condition and the other half were randomly assigned to the business-as-usual condition. For the present research, we were interested in the relationship between CSR implementation (specifically, quality and dosage) and student outcomes. Because students in the business-as-usual condition did not receive any CSR instruction, we modeled effects for CSR study classes only.

Table 1. Description of Study 1 and Study 2.

Study characteristic	Study 1	Study 2
Setting	9 middle schools in 2 urban and 1 suburban school districts in 2 states	3 middle schools in 1 urban school district
Grade levels	7, 8	6, 7, 8
Teachers	20	23
Content areas and number of classes	37 sections of either reading or LA	30 sections of SCI/SS class pairs
Teacher participation	Teachers volunteered to use CSR	Teachers taught CSR as part of a district initiative
Professional development	2 days up front professional development (12 hours) 2–4 after-school booster sessions (6–8 hours) 2–3 individual coaching sessions/month	Same as Study 1
Expected implementation	CSR twice a week in either reading or LA classes	CSR twice per week—once per week in SCI, once per week in SS
Research design: Fidelity study within-subjects	Students in CSR sections only	Same as Study 1
Implementation measure: Dosage	Teachers report number of sessions	Same as Study 1
Implementation measure: Quality	4 classroom observations using IVC	3 classroom observations using IVC

Note. LA = language arts; SCI = science; SS = social studies; CSR = Collaborative Strategic Reading; IVC = *Implementation Validity Checklist* (Vaughn et al., 2011; Vaughn et al., 2013).

This within-groups design has been used in other studies examining the relationship between fidelity of implementation and student outcomes (Crawford et al., 2012; Foster, Oh, Azano, & Callahan, 2012; Hamre et al., 2010; Kopp, Hulleman, Harackiewicz, & Rozek, 2012).

Setting and Participants

Study 1. The first study took place in the metropolitan areas of two states. Teachers and students from nine middle schools located in three school districts volunteered to participate. Three of the middle schools were located in a major suburban school district in one state that serves 11,000 students. The population of this district is 6% White, 82% Hispanic, and 11% African American. Eighty-seven percent of the students receive free or reduced-cost lunch (FRL; a proxy for low-income status), and 32% are ELLs. In the other state, five of the participating middle schools were from a large urban district. The student population is 20% White, 59% Hispanic, and 15% African American. Seventy-two percent of the students receive FRL, and 35% are ELLs. One school in the same state was located in a district that serves 29,000 students. Roughly 28% of this district's students are Hispanic, 66% are White, and 1% are African American. Thirty-two percent of the student population is low-income, and 15% are ELLs.

Study 2. The second study was conducted in three middle schools from the same large urban district that participated in Study 1. Teacher and student participants in this study were part of a district-university collaboration designed to evaluate the effectiveness of CSR. From the beginning, planning and implementation occurred in response to the district's identified needs and took into consideration existing practices and goals for change.

Students. Table 2 presents student demographics for both studies. There were 608 middle school students in the initial sample for Study 1. The majority of students (73%) were Hispanic, and 10% received special education services. For the second study, 553 students were in the initial sample; just more than half (51%) were Hispanic, and 12% received special education services.

Teachers. Study 1 initially included 20 middle school teachers and 37 sections of language arts or reading classes. One teacher in Study 1 with three sections totaling 30 students could not be included in the analysis because insufficient fidelity data were collected on this teacher. Forty percent of the teachers in Study 1 had taught for fewer than 5 years, 25% had taught 5 to 10 years, and 35% had taught for more than 10 years. Teachers in Study 1 volunteered to participate. Study 2 included 23 teachers who formed 12 teacher pairs and 30 sections of social studies/science class pairs. As mentioned previously, these teachers were required to implement CSR as part of a district initiative but chose to participate in the research aspects of the study (e.g., classroom observations). Thirty-six percent of the teachers in Study 2 had taught for fewer than 5 years, whereas 34% had taught 5 to 10 years, and 30% had taught for more than 10 years.

Table 2. Student Characteristics of Initial Sample for Studies 1 and 2.

Characteristic	Study 1[a]		Study 2[b]	
	n	%	n	%
Gender				
Female	265	43.6	275	49.7
Male	332	54.6	278	50.3
Ethnicity				
White	157	25.8	168	30.4
African American	50	8.2	67	12.1
Latino	445	73.2	283	51.2
Pacific Islander	1	0.2	0	0.0
American Indian or Alaska Native	16	2.6	7	1.3
Multiracial but of unknown origin	8	1.3	17	3.1
Asian	17	2.8	11	2.0
English-language learner	271	44.6	80	14.5
Special education	61	10.0	67	12.1
Free or reduced-cost lunch	—	—	364	65.8

Note. The percentages associated with ethnicity for Study 1 may add up to more than 100 because students could be categorized to more than one ethnicity. For Study 1, Gender and Ethnicity, 1.8% of records (*n* = 11) had missing data. The variable English-Language Learners had 4.9% of records (*n* = 30) with missing data, and the Special Education variable had 2.3% (*n* = 14) of records with missing data. Dashes indicate that free or reduced-cost lunch data were collected only in Study 2.
[a]*N* = 608. [b]*N* = 553.

Procedures

Description of treatment intervention. CSR helps students learn specific strategies implemented before, during, and after reading that enhance reading comprehension. These strategies include the following:

1. Before reading a text, teachers present the topic and key vocabulary terms and students activate prior knowledge and predict what they will learn from the reading (*preview*).
2. During reading, students monitor their understanding and take steps to repair comprehension when it breaks down (*click and clunk*). In addition, students find the main idea of short sections of text (*get the gist*).
3. After reading a text, students wrap up the reading by asking and answering each other's questions and reviewing key ideas (*question generation* and *review*).

The teacher introduces the comprehension strategies to the whole class using explicit instruction, modeling, think alouds, and guided practice. Once students know the strategies, they apply them while working in student-led cooperative groups. In addition, students use CSR learning logs to record the reading strategies they are using.

Both studies used the same implementation procedures. Students were to receive the full CSR model (i.e., preview, click and clunk, get the gist, question generation, and review) two times each week throughout the school year.

The studies differed in the content areas in which CSR instruction was delivered. For Study 1, language arts and reading teachers were asked to implement CSR two class periods each week (approximately 50 min each session by the same teacher) throughout the school year. For Study 2, the experimental treatment unit was pairs of social studies and science teachers, and CSR was implemented 1 day a week for about 50 min in social studies and 1 day a week for 50 min in science classes throughout the school year.

Professional development. Teachers across both studies received similar support for their implementation of CSR consisting of 2 days of professional development, two to four after-school booster sessions, and individual coaching (two to three times per month). Professional development included providing support in learning CSR strategies and teaching CSR to students. After-school booster sessions varied in response to teachers' needs and covered topics such as curriculum alignment, fine-tuning instruction for specific strategies, and facilitating cooperative learning in diverse classrooms.

Implementation fidelity. Similar to Mowbray et al. (2013), for the present study, we categorized "dosage" (i.e., the number of CSR sessions) as fidelity to structure. We operationalized fidelity to process using the *Implementation Validity Checklist* (IVC; Vaughn et al., 2011; Vaughn et al., 2013), a rating scale in which trained observers record the presence or absence, as well as the quality, of instruction for each CSR component during multiple classroom observations. Thus, we associated

two fidelity variables with student outcomes, one categorized under the construct of structure (CSR dosage) and the other under process (quality of teacher implementation of CSR instructional strategies).

Fidelity to structure. Teachers across both studies reported the number of CSR teaching sessions to provide a measure of dosage. At the close of each CSR class period, teachers reported the number of minutes spent engaged in CSR instruction. The number of minutes recorded across the school year was summed to produce a measure of "total minutes" of CSR instruction. In addition, we summed the number of times a teacher submitted a log to determine the "total sessions" of CSR. These processes were consistent across both studies. While the teacher logs were combined with classroom observations to estimate the frequency of CSR use, they were limited in that recordings relied on teacher self-reports.

Fidelity to process. Through in-class observations, raters used the IVC to record the extent to which teachers implemented each component of the CSR model and the quality of instruction for each CSR component (see Figure 1). The IVC consists of a three-part observation and is conducted during one session of CSR. The items break down the essential features of CSR in terms of student behaviors, teacher behaviors, and quality of instruction.

The first IVC section focuses on CSR procedural fidelity in which teachers are rated on each component of CSR and scores range from 1 (*inconsistent with the CSR model*) to 4 (*highly aligned*). If the strategy was not observed, the teacher received a rating of zero. In the second section, teacher and student behaviors that are essential to high-quality implementation of CSR receive scores from 1 (*not observed*) to 3 (*observed*).

The third section includes a CSR global quality rating on a scale from 1 (*low*) to 7 (*high*) to measure the overall quality of a teacher's CSR implementation. To determine the CSR global rating, IVC observers were trained to review the fidelity scores, review the scores for the teacher and student behaviors, and combine them into one holistic score encompassing procedural fidelity, quality of instruction, and how well the instruction was received by students. For example, if a rater assigned a teacher mostly 4s on the procedural fidelity scale and 3s on the teacher/student behavior scales, the overall CSR quality rating was expected to be a 6 or 7. A 7 indicates strong CSR implementation with little room for improvement. The following guidelines were used to determine a CSR global rating score:

- *High quality (6–7)*: Most students are actively engaged. The teacher provides feedback appropriate to student needs and helps students gain proficiency at using CSR strategies.

- *Average quality (4–5)*: Students are familiar with CSR but lack a high level of engagement in the activities. The teacher provides inconsistent feedback or inaccurate information about one or more strategies or omits a strategy that should be present.
- *Low quality (1–3)*: Most students are not engaged in or familiar with the CSR strategies. The teacher does not provide the necessary modeling, explanation, or feedback.

For the present research, we used the CSR global quality rating as the "process" measure of fidelity. Other researchers have found the global score to be an adequate measure of teacher quality. For example, Foorman et al. (2006) used a similar global measure of teacher effectiveness and determined the global score to be highly correlated with a checklist of items measuring teacher effectiveness on a separate observation scale.

Four rounds of IVC observations were conducted for Study 1. The first round took place midyear, with subsequent observations occurring approximately once a month until the end of the school year. Study 2 included three rounds of IVC observations for each teacher pair, starting midyear and reoccurring roughly every 6 weeks until near the end of the school year. Thus, for Study 1 we collected four measures of CSR global quality rating scores, and for Study 2 we collected three measures for each teacher, or six measures in total. Two teachers (one in each study), however, had one fewer observation due to scheduling conflicts.

Assessing the reliability of fidelity to process. For both studies, all IVC observers were experts in CSR implementation and participated in trainings to ensure interrater agreement prior to beginning classroom observations. Training on the IVC included using a scoring manual to address implementation fidelity of audio- and video-recorded CSR lessons. For both studies, a lead researcher served as the "gold standard of reliability," a training process that has been used in other studies (e.g., Wanzek et al., 2014, p. 190). Ratings were discussed until consensus with the lead scorer on each IVC item was reached. Researchers then observed CSR classrooms for training purposes and independently coded the in-person observations. Discrepant scores were addressed, and the process was repeated until an interrater agreement of 90% on all IVC items was reached consistently for each of the raters.

For Study 2, we continued trainings while we collected data. We did this by comparing individual team members' scores to one experienced lead observer's score across a percentage of the overall observations. The lead observer simultaneously conducted a classroom observation with an individual team member, and each rater calculated scores separately. The reliability was calculated across 20 observations (or roughly 10% of the total number of observations). Because interrater agreement in the field was 0.80 (compared to 0.90 in the training), field notes

Part I. Procedural Fidelity – Rate the fidelity to the CSR model as well as the quality of instruction for each CSR component.

Very High: 4	Mid High: 3	Mid Low: 2	Low: 1	Not Observed: 0

Before Reading

Preview:

Teacher Preview: Teacher introduces the topic; pre-teaches vocabulary and/or proper nouns; uses English language learner (ELL) strategies during pre-teach; builds background knowledge/connects to students' prior knowledge; sets the purpose for reading.

Student Preview: Students preview the passage; brainstorm what they already know about the topic; predict what they will learn; share with small group or partner.

During Reading

Click & Clunk

ID Clunks: Students identify words or ideas they don't understand; write them in their log.

Use Fix-up Strategies: Students use fix-up strategies to figure out the meaning of the clunk; work together with their group to find meanings of clunks.

Get the Gist

Students identify the most important "who" or "what" in the section and the most important information about the "who" or "what"; write an individual gist; share gists and provide feedback in group.

After Reading

Questions & Answers

Students write leveled questions and answers; ask and answer questions with their group.

Review

Students write 1–2 of the most important ideas from the entire passage; share with their group; teacher leads whole group wrap-up that reviews important content and/or highlights a CSR strategy.

Part II. Teacher and Student Behaviors Related to Reading

Very High: 4	Mid High: 3	Mid Low: 2	Low: 1

Teacher monitors and provides feedback on strategy use.
Teacher monitors student groups.
Teacher reinforces content learning.
Students help each other while working in groups.
All students participate in their small groups.
Evidence that students are proficient at using strategies independently.

Part III. Global Rating

Highest: 7	Mid High: 6	High Avg: 5	Mid Avg: 4	Low Avg: 3	Mid Low: 2	Low: 1

Use ratings from the procedural fidelity scale as well as teacher and student behaviors scale to determine the overall rating of CSR. The scores should be roughly equivalent.

Figure 1. Implementation Validity Checklist (IVC).

were reviewed and compared to ratings by the lead scorer. Any discrepant scores were discussed until agreement was reached, and ratings were then adjusted as needed.

Using data collected in each study, internal consistency for the CSR global rating was .80 for Study 1 and .91 for Study 2.

Measures

Reading outcomes were measured in both studies using the reading comprehension subtest of the fourth edition of the *Gates-MacGinitie Reading Test* (GMRT; MacGinitie, MacGinitie, Maria, & Dryer, 2000). Two parallel forms permit pre- and posttesting. The GMRT is a timed paper-and-pencil, group-administered measure of student achievement in reading. The assessment was administered by trained researchers twice: at the beginning and end of the school year. Internal consistency reliability for the GMRT ranges from .91 to .93, and alternate form reliability is reported as .80 to .87.

Data analysis procedures. The GMRT scores were standardized in both studies to have a mean of 100 and standard

deviation of 15 to ease interpretation. Quality was modeled using the IVC seven-point CSR global rating score averaged across each time point. The total number of minutes that CSR was taught by each teacher (Study 1) or teacher pair (Study 2) was used to model dosage.

Using the Hierarchical Linear Modeling 7.0 software program (Raudenbush, Bryk, & Congdon, 2010), we used multilevel, random-intercept models to estimate the relationship between CSR implementation (dosage and quality) and student outcomes after controlling for student pretest scores and demographics. No random slopes were included, as neither study had hypotheses regarding variance in the effects of Level 1 covariates over teachers or teacher pairs. Student-level covariates (e.g., special education status) were included in both models. Study 2 also included FRL since we had access to these data. For both studies, we ran a three-level model with students nested in classes nested in teachers (Study 1) and students nested in class pairs nested in teacher pairs (Study 2). Gates-MacGinitie pretest score, ELL, and special education status were included at Level 1 (FRL was also included at Level 1 for Study 2). CSR dosage and quality were modeled at Level 2. In addition, we modeled an interaction term between special education status and CSR dosage at Level 2. The Level 2 model also included an interaction term between CSR global quality and special education status. First we analyzed data from Study 1 using a within-groups design. We conducted a separate analysis with data from Study 2, also using a within-groups design. Following the three-level notation described by Raudenbush and Bryk (2002), the model for each study therefore took the following form:

$$\text{Level 1}: \text{GMRT post} = \pi_{0jk}$$
$$+ \pi_{1jk} * (\text{SpEd}_{ijk})$$
$$+ \pi_{2jk} * (\text{FRL}_{ijk})$$
$$+ \pi_{3jk} * (\text{ELL}_{ijk})$$
$$+ \pi_{4jk} * (\text{GMRT pre}_{ijk})$$
$$+ e_{ijk}$$

$$\text{Level 2}: \pi_{0jk} = \beta_{00k} + \beta_{01k} * (\text{Total Minutes }_{jk})$$
$$+ \beta_{02k} * (\text{CSR Global Quality }_{jk})$$
$$+ r_{0jk} \ \pi_{1jk} = \beta_{10k} + \beta_{11k} * (\text{Total Minutes }_{jk})$$
$$+ \beta_{12k} * (\text{ CSR Global Quality }_{jk})$$

$$\text{Level 3}: \beta_{00k} = \gamma_{000} + u_{00k}$$

There were 36 students in Study 1 with missing data (either demographic or GMRT). For Study 2, one entire class (*n* = 30 students) was missing posttest data due to a scheduling error with the testers; an additional 45 students in the initial sample were missing posttest data for random reasons.

Records with missing data were dropped using listwise deletion. We did not evaluate attrition bias because we were analyzing data within treatment groups. Thus, the usual concerns with bias (that randomized groups are no longer comparable due to differential attrition) did not apply.

Results

Research Question 1

To what extent was CSR implemented as intended? Table 3 presents findings from the implementation data collected across both studies. These data include (a) fidelity to structure, operationalized as CSR dosage—a teacher report of the number of times CSR was taught throughout the school year—and (b) fidelity to process, operationalized as CSR quality—the average of observation ratings collected by teachers across each data collection time point.

Dosage varied by study. For Study 1, teachers taught CSR for an average of 17 sessions (*SD* = 9). Teacher pairs in Study 2 taught an average of 42 sessions (*SD* = 5; 21 sessions for social studies teachers and 21 sessions for science teachers). Thus, in Study 2, students received approximately twice as many CSR sessions as they did in Study 1. The CSR components that composed procedural fidelity were, on average, about the same in both studies, though "review" (the final component of CSR) was higher in Study 2 compared to Study 1. In both studies, mean scores for most CSR components (preview, clunks, fix-up strategies, and gist) ranged between ratings of mid-low (i.e., several instructional components are inconsistent and not aligned with CSR) and mid-high (i.e., majority of instructional components are implemented with high fidelity and moderate alignment with CSR). However, mean scores for "ask and answer questions" and "review" ranged from ratings of not observed to low (i.e., most instructional components were not implemented) and mid-low. The teacher and student behaviors, as measured by the IVC, were similar in Studies 1 and 2. Overall, these data indicate that in most cases, the teacher provided adequate instruction in the before and during reading strategies, and students performed those strategies at a proficient level. The overall rating of CSR (i.e., CSR quality) was roughly equivalent across both studies, as measured by the seven-point CSR global rating scale averaged across multiple data collection time points (Study 1: *M* = 4.75, *SD* = 0.84; Study 2: *M* = 5.29, *SD* = 1.08). On average, CSR quality in both studies was considered to be slightly lower than expected, indicating room for improvement in various aspects of CSR.

Completion of the model. Findings from IVC observations indicate that teachers in both studies did not teach all of the CSR components (see Table 4). When averaging the number of components observed across each round of data collection, only 21% of teachers in Study 1 were observed

Table 3. Implementation Data for Studies 1 and 2.

	Study 1		Study 2	
Measure	M	SD	M	SD
Dosage: Total minutes	1,194	442	2,095	361
Dosage: Total sessions	17	9	42	5
IVC Section 1: Procedural fidelity (0–4; 0 = not observed)				
Teacher preview	2.32	0.89	3.10	0.50
Student preview	2.21	0.78	2.91	0.54
Identify clunks	2.58	0.75	3.12	0.60
Use fix-up strategies	2.15	0.65	2.77	0.55
Get the gist	2.20	0.76	2.74	0.58
Ask and answer questions	1.51	0.84	1.77	0.93
Review	0.79	0.84	1.52	0.63
IVC Section 2: Teacher and student behaviors (1–4 scale)				
Teacher monitors and provides feedback	2.91	0.55	3.04	0.63
Teacher reinforces learning	2.34	0.72	3.08	0.43
Teacher monitors student groups	3.07	0.54	3.34	0.51
Students help each other while working in groups	2.74	0.57	2.79	0.76
All students participate in groups	3.23	0.47	3.10	0.66
Proficient at using strategies independently	2.51	0.45	2.80	0.41
IVC Section 3: CSR global quality (1–7 scale)	4.75	0.84	5.29	1.08

Note. IVC = *Implementation Validity Checklist* (Boardman, et al., 2015; Vaughn et al., 2011; Vaughn et al., 2013).

Table 4. Frequency of Collaborative Strategic Reading Components (or Strategies) Observed for Studies 1 and 2.

	Study 1[a]		Study 2[b]	
Strategies observed (number)	n	%	n	%
0–2	0	0.0	0	0.0
3–4	4	21.1	0	0.0
5–6	11	57.9	8	34.8
7 (all core strategies)	4	21.1	15	65.2

Note. The unit of analysis is at the teacher level, and the number of components observed was averaged across each round of data collection. The seven Collaborative Strategic Reading components measured in this study were teacher preview, student preview, identify clunks, use fix-up strategies, get the gist, question generation, and review.
[a] n = 19. [b] n = 23.

teaching all seven of the CSR components (teacher preview, student preview, identify clunks, use fix-up strategies, get the gist, question generation, and review). Just more than half (60%) were observed teaching five or six strategies. For Study 2, about two thirds (65%) of teachers were observed teaching all seven components, and just more than one third (35%) of teachers were observed teaching five or six strategies. The strategy most frequently not observed was review, followed by question generation. This finding was consistent across both studies.

Research Question 2

Is higher CSR instructional quality associated with increased student outcomes? The mean pretest and posttest

reading comprehension scores, as measured by the GMRT, for students with disabilities were 84.8 (*SD* = 11.2) and 85.6 (*SD* = 11.0; Study 1) and 82.65 (*SD* = 11.9) and 83.75 (*SD* = 13.0; Study 2). For students without disabilities, pretest and posttest mean GMRT scores were 97.1 (*SD* = 14.3) and 97.9 (*SD* = 13.7; Study 1) and 93.2 (*SD* = 12.2) and 95.0 (*SD* = 12.3; Study 2). A pre/post growth score of zero, indicating the same score in the fall and spring, represents average growth in reading (MacGinitie et al., 2000). Results show that on average, while students with disabilities received lower mean pretest and posttest scores on the GMRT, all students earned a growth score of more than zero. These findings are presented in Table 5.

Results across both studies showed no statistically significant effect of quality on student outcomes for the overall

Table 5. Descriptive Statistics for Students in Initial Sample, by Special Education Status.

	Study 1		Study 2	
Measure	M	SD	M	SD
Special education				
GMRT pretest	84.82	11.20	82.65	11.94
GMRT posttest	85.62	11.04	83.75	13.01
Non–special education				
GMRT pretest	97.14	14.31	93.17	12.21
GMRT posttest	97.93	13.65	94.97	12.27

Note. GMRT = *Gates-MacGinitie Reading Test* (MacGinitie, MacGinitie, Maria, & Dreyer, 2000).

Table 6. The Effect of CSR Implementation and GMRT Pretest Scores on GMRT Posttest Scores.

	Study 1			Study 2		
Measure	B	SE	p value	B	SE	p value
Intercept, $\beta 0$	93.67	0.73	<.001	97.71	0.63	<.001
Total minutes, $\gamma 010$	0.00	0.00	.876	0.00	0.00	.322
CSR global quality, $\gamma 020$	1.52	0.80	.080	−0.05	0.72	.945
SPED, $\beta 10$	−3.85	1.41	.007	−2.06	1.35	.126
Total minutes, $\gamma 110$	0.00	0.00	.740	0.00	0.00	.489
CSR global quality, $\gamma 120$	3.61	1.59	.024*	2.27	1.14	.046*
FRL, $\beta 20$	—	—	—	0.24	0.98	.808
ELL, $\beta 30$	−0.53	.077	.487	−1.18	1.14	.301
GMRT pretest, $\beta 40$	0.74	0.03	<.001	0.69	0.03	<.001

Note. CSR = Collaborative Strategic Reading; GMRT = *Gates-MacGinitie Reading Test* (MacGinitie, MacGinitie, Maria, & Dreyer, 2000); SPED = special education; FRL = free or reduced-cost lunch; ELL = English-language learner. Dashes indicate that FRL data were not collected for Study 1.
*p < .05.

sample. There was, however, a significant, positive interaction effect between quality of CSR instruction (i.e., fidelity to process) and posttest scores for students with disabilities (see Table 6). This finding indicates that when the quality of CSR instruction was higher, the posttest reading comprehension scores for students with disabilities increased. Specifically, controlling for all covariates, the estimated effect of special education status in Study 1 was −3.85 points ($p < .05$) on the global rating scale. The total estimated effect of a one-point increase in quality of implementation for students with disabilities was 5.12 ($p < .05$); furthermore, this estimate is 3.61 points higher ($p < .05$) than the estimated total effect for students without disabilities. Meanwhile, controlling for all covariates, the estimated effect for students with disabilities in Study 2 was −2.06 points ($p = .126$), while the total estimated effect of a one-point increase in quality of implementation for students with disabilities was 2.22 ($p < .05$), an estimate that is 2.27 points higher ($p < .05$) than the estimated total effect for students without disabilities. Figure 2 illustrates the relationship between the quality of CSR instruction and student outcomes for students with and without disabilities.

Research Question 3

Is the amount of CSR instruction related to student outcomes? Results across both studies showed no statistically significant effect of quantity of CSR implementation (dosage) on student outcomes for the overall sample. Furthermore, there was no significant interaction between dosage and disability status, a finding also consistent across both studies. In other words, there was no relationship either with the whole sample or with the subgroup of students with disabilities between how often a teacher implemented CSR and improvement in student outcomes.

Discussion

We report on two separate studies to address the conditions under which relationships are demonstrated to be replicable in different contexts and with nonoverlapping groups of teachers and students. Replicating findings is important for identifying sound theories of best practice that generalize beyond a specific study. On average, students in the full sample in both studies did not improve as
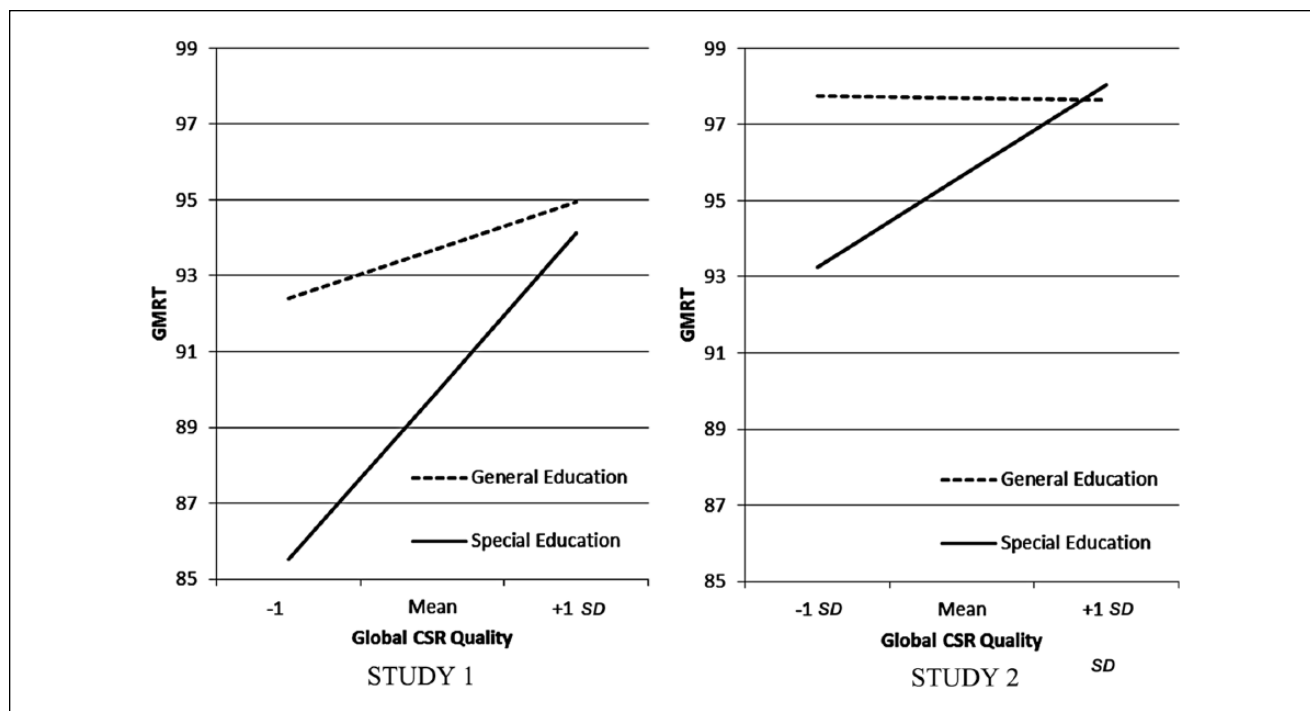
Figure 2. The differential influence of Collaborative Strategic Reading (CSR) global quality (as measured by the CSR global quality rating on the *Implementation Validity Checklist*) on *Gates-MacGinitie Reading Test* (GMRT) scores among students in special education and general education students.
Note. The pre-test GMRT values were mean centered to reflect the population mean of 100. As such the fitted values in this figure provide average values that are higher than the descriptive means presented in Table 5.

a function of either quality or dosage. Results from both studies, however, show a positive relationship that is statistically significant for the quality of CSR instruction and students with mild to moderate disabilities. That is, on average, higher posttest reading comprehension scores for students with disabilities were associated with higher quality of CSR instruction. Thus, the interaction of CSR global quality and disability predicted a statistically significant amount of variance in posttest reading comprehension scores (controlling for pretest scores), a finding replicated across two studies.

While it is intuitive that more instruction in CSR would be associated with greater student reading gains, this did not prove to be the case in this research. Despite wide variation across teachers and nearly twice as much exposure to CSR in Study 2, the amount of CSR instruction (dosage) was not related to changes in student reading outcomes for the full sample or for students with disabilities. This finding is consistent with the findings from previous research. For instance, Hamre and colleagues (2010) found no relationship between dosage and literacy outcomes for the young students in their study. Dane and Schneider (1998) synthesized results from six studies that reported on dosage of implementation for struggling learners (studies were behavioral, academic, or social interventions) and found mixed

results related to the connection between dosage and outcome measures. One explanation could be that a substantial number of sessions must be accumulated before there is an association between dosage and student outcomes. However, in a synthesis that focused only on intensive intervention studies with at least 75 sessions with struggling readers or students with disabilities, no significant differences related to dosage were evident (Wanzek et al., 2013).

Many models, including CSR, provide expectations for the amount of instruction that students should receive. In our studies, students were to receive CSR instruction two times each week, and researchers and coaches worked closely with teachers to integrate CSR regularly into their practice. Still, it should be noted that in both studies, dosage was lower than the expected 50 sessions per student, with great variation across the two studies. In Study 1, students received CSR an average of 17 sessions; in Study 2 students received an average of 42 sessions. Recall that in Study 2, two teachers were teaching CSR, with each teacher delivering approximately 21 sessions. Perhaps the influence of sessions is not cumulative (across teachers) but is more teacher specific. Although the results here suggest that quality may be more important than quantity for students with disabilities, we are as yet unable to provide specific recommendations for best practice related to dosage.

Another way to operationalize fidelity is completion of the full CSR model (i.e., teaching all CSR components, including preview, clunks, fix-up strategies, question generation, and review). Similar to findings reported by Hitchcock and colleagues (2011), IVC observations indicated that teachers were not teaching all the intended CSR reading strategies. In particular, the after-reading strategies, question generation and review were observed in both studies at lower rates than preview, clunks, fix-up strategies, and gist. Our analysis did not examine the relationship between completion of the full CSR model and student outcomes. Additional research is needed to further investigate the influence of completion of the model.

Our findings highlight the importance of quality of CSR instruction for students with disabilities and perhaps for other struggling readers. For this research, we used similar models to examine the relationship between fidelity and student outcome data from two nonoverlapping CSR studies. Interaction effects between quality of implementation and student reading comprehension outcomes for students with disabilities were significant and positive in both studies, while there was no significant association between fidelity and student outcomes for the full sample of students in either study. Fundamental principles of CSR are that teachers offer explicit instruction in each of the reading strategies and that they support the development of student-managed groups. These supports can be particularly beneficial for students with learning disabilities (Edmonds et al., 2009; Faggella-Luby et al., 2006; Gajria et al., 2007; Scammacca et al., 2007). Although we were unable to determine the contribution of individual aspects of the CSR model, our results suggest that quality of CSR instruction is important for students with disabilities in general education settings.

For several studies, researchers have reported that teacher quality is more strongly related to student outcomes than other measures of fidelity (Dane & Schneider, 1998) and that these associations may be the most evident when there is the greatest variation in fidelity (O'Donnell, 2008). Still, there are other studies that have shown that fidelity to process is not associated with student outcomes (e.g., Crawford et al., 2012). These researchers note the difficulty in measuring fidelity to process. It may be that quality is important but that it was not measured accurately. Issues of fidelity are multifaceted and involve the teachers, their students, and the ability of observers to objectively measure the quality of instruction. Still, we are encouraged by the association between CSR quality and outcomes for students with disabilities. CSR is intended to support those students who are in need of the instructional features that compose the model (i.e., explicit instruction in reading strategies; opportunities to use reading strategies before, during, and after reading; peer discussion; resources to scaffold learning throughout the reading; appropriate teacher feedback).

Limitations

The research design used in this study was not experimental as there was no random assignment of students or teachers to treatment or control conditions. As a result, findings cannot justify causal inferences about the effect of CSR quality on improved student reading achievement. Rather, our research suggests that there is a positive relationship between CSR quality and reading comprehension outcomes for students with disabilities.

Another limitation is related to the challenge of identifying reliable and valid measures of teacher quality. Subjectivity is inherent in the IVC observation tool. It is also possible that the IVC misses critical aspects of effective instruction that are not picked up as part of the CSR model. Studies concerning fidelity to CSR should therefore examine the internal structure of the IVC items empirically and in relation to expected student and teacher outcomes, thereby focusing on the validity of the instruments. Still, as Mowbray and colleagues (2003) explained, compared to structural components of an intervention, process criteria require more subjective judgments because they are often based on aspects such as observations and interviews. Thus, process measures such as the IVC observation tool used in this study are inherently more difficult to measure reliably. In addition, they are costly and time intensive. While further attention to the collection of fidelity measures is important to evaluation studies, the limitations outlined in this section highlight the inherent challenges of measuring fidelity to a treatment or intervention.

Implications for Practice and for Future Research

General education teachers in secondary settings are being asked to include more challenging expository text reading in their content classrooms. In addition, they are being asked to provide instruction that increases both content learning and reading skills and that is differentiated to meet the needs of learners who represent a range of reading abilities and content knowledge. These instructional goals are weighty and leave many teachers feeling underprepared for the challenges they are faced with in schools and classrooms with limited resources and support. Despite research that supports the use of instructional models such as CSR that embed reading supports and features of effective instruction within content learning, students with disabilities in general educational classrooms most frequently receive instruction that is whole group and is not differentiated to meet their instructional needs (E. A. Swanson, 2008). To bridge the research-to-practice gap, providers of professional development must understand the features of implementation associated with improving student outcomes.

Despite relatively weak findings for the full sample in Study 1 and a nonsignificant effect of teacher quality on

full sample student outcomes in Study 2, it is notable that students with disabilities appeared to be more sensitive to the quality of instruction than students without disabilities. Future research is needed to determine why there was a stronger association between teacher quality and reading outcomes for these students. Another finding from this study that warrants additional research is that despite lower dosage and slightly lower quality of instruction, the standardized coefficients in Study 1 were nearly double the standardized coefficients in Study 2. In other words, the trajectory from pretest to posttest (or academic improvement) of CSR students with disabilities in Study 1 was nearly twice as steep as the trajectory of CSR students with disabilities in Study 2 (see Figure 2). We need more information to understand the difference in magnitude in Study 1 (implemented with reading and language arts teachers) compared to Study 2 (implemented with social studies and science teachers). In addition, more research is needed that expands how we study fidelity and to design studies that test the results presented here with other interventions and in different contexts. Echoing the results reported in O'Donnell (2008), our findings demonstrate the need for education researchers to more readily measure and report on how fidelity relates to outcomes before drawing conclusions about best practices for helping students learn.

Authors' Note

Declaration of Conflicting Interests

Funding

Note

Both studies used a randomized control trial design to measure whether exposure to Collaborative Strategic Reading (CSR) resulted in improved student reading outcomes when compared to students in a business-as-usual condition that did not receive CSR instruction. Results for Study 1 are reported in Reutebuch, Stillman-Spisak, Solis, Boardman, and Klingner (2012), and results for Study 2 are reported in Boardman et al. (2015).

References

Boardman, A. G., & Woodruff, A. L. (2004). Teacher change and "high-stakes" assessment: What happens to professional development? *Teaching and Teacher Education*, *20*(6), 545–557.

Boardman, A., Klingner, J., Buckley, P., Annamma, S., & Jensen, C. (2015). Collaborative Strategic Reading in content classes: Results from year 1 of a randomized control trial. *Reading and Writing: An Interdisciplinary Journal, 28*(9), 1257–1283.

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.

Crawford, L., Carpenter, D., Wilson, M., Schmeister, M., & McDonald, M. (2012). Testing the relation between fidelity of implementation and student outcomes in math. *Assessment for Effective Intervention*, *37*, 224–235.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*, 23–45.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research Theory and Practice*, *18*, 237–256.

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

Faggella-Luby, M., Schumaker, J. S., & Deshler, D. D. (2006). Embedded learning strategy instruction: Story-structure pedagogy in heterogenous secondary literature classes. *Learning Disability Quarterly*, *30*, 131–147.

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

Foorman, B. R., Schatschneider, C., Eakin, M. N., Fletcher, J. M., Moats, L. C., & Francis, D. J. (2006). The impact of instructional practices in Grades 1 and 2 on reading and spelling achievement in high poverty schools. *Contemporary Educational Psychology*, *31*, 1–29.

Foster, L. H., Oh, S., Azano, A., & Callahan, C. (2012, March). *Adherence: A more nuanced usage of fidelity of implementation's core component to capture variation in treatment effects.* Paper presented at the spring conference of the Society for Research on Educational Effectiveness, Washington, DC.

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

Hamre, B. K., Justice, L. M., Pianta, R. C., Kilday, C., Sweeney, B., Downer, J. T., & Leach, A. (2010). Implementation fidelity of MyTeacherPartner literacy and language activities: Association with preschoolers' language and literacy growth. *Early Childhood Research Quarterly*, *25*, 329–347.

Hilden, K. R., & Pressley, M. (2007). Self-regulation through transactional strategies instruction. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *23*, 51–75.

Hitchcock, J., Dimino, J., Kurki, A., Wilkins, C., & Gersten, R. (2011). *The impact of collaborative strategic reading on the reading comprehension of Grade 5 students in linguistically diverse schools* (NCEE 2011-4001). Washington, DC: U.S. Department of Education, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences.

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.

Kagan, S. (1986). Cooperative learning and sociocultural factors in schooling. In Bilingual Education Office, California State Department of Education (Ed.), *Beyond language: Social and cultural factors in schooling language minority students* (pp. 231–298). Los Angeles: California State University, Evaluation, Dissemination and Assessment Center.

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

Klingner, J. K., Vaughn, S., Boardman, A., & Swanson, E. (2012). *Now we get it! Boosting comprehension with Collaborative Strategic Reading*. San Francisco, CA: Jossey Bass.

Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. *Elementary School Journal*, *99*, 3–22.

Kopp, J. P., Hulleman, C. S., Harackiewicz, J. M., & Rozek, C. (2012, March). *Applying the five-step model of fidelity assessment to a randomized experiment of high school STEM intervention*. Paper presented at the spring conference of the Society for Research on Educational Effectiveness, Washington, DC.

Lipsey, M. W. (1999). Can rehabilitative programs reduce the recidivism of juvenile offenders? An inquiry into the effectiveness of practical programs. *Virginia Journal of Social Policy & the Law*, *6*, 611–641.

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

Mowbray, C., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*, 315–340.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core state standards*. Washington, DC: Author.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relation to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, *78*, 33–84.

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

Raudenbush, S. W., & Bryk, T. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, T., & Congdon, R. (2010). *HLM 7 hierarchical linear and nonlinear modeling* [software]. Skokie, IL: Scientific Software International.

Reutebuch, C., Stillman-Spisak, S., Solis, M., Boardman, A., & Klingner, J. (2012, April). *Effects of Collaborative Strategic Reading on struggling readers, including English learners: Year 3*. Paper presented at the Council for Exceptional Children annual conference, Denver, CO.

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., 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.

Scammacca, N., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in Grades 4–12: 1980–2011. *Journal of Learning Disabilities, 48,* 369–390. doi:10.1177/0022219413504995

Swanson, E. A. (2008). Observing reading instruction for students with learning disabilities: A synthesis. *Learning Disability Quarterly*, *31*, 115–133.

Swanson, H. L. (1999). Reading research for students with LD: A meta-analysis of intervention outcomes. *Journal of Learning Disabilities*, *32*, 504–532.

Vaughn, S., Klingner, J., Swanson, E. A., Boardman, A. G., Roberts, G., Mohammed, S., & Still-Spisak, S. J. (2011). Efficacy of collaborative strategic reading with middle school students. *American Educational Research Journal*, *48*, 938–964.

Vaughn, S., Roberts, G., Klingner, J., Swanson, E., Boardman, A., Stillman-Spisak, S. J., & Leroux, A. (2013). Collaborative strategic reading: Findings from experienced implementers. *Journal of Research on Educational Effectiveness*, *6*, 137–163.

Wanzek, J., Vaughn, S., Kent, S., Swanson, E. A., Roberts, G., Haynes, M., & Solis, M. (2014). The effects of team-based learning on social studies knowledge acquisition in high school. *Journal of Research on Educational Effectiveness*, *7*, 183–204.

Wanzek, J., Vaughn, S., Scammacca, N. K., Metz, K., Murray, C. S., Roberts, G., & Danielson, L. (2013). Extensive reading interventions for students with reading difficulties after Grade 3. *Review of Educational Research*, *83*(2), 163–195.

Zigmond, N., Kloo, A., & Volonino, V. (2009). What, where, and how? Special education in the climate of full inclusion. *Exceptionality: A Special Education Journal*, *17*, 189–204.