Abstract Title Page
Not included in page count.

Title:

The relationship between implementation of Collaborative Strategic Reading and student outcomes for adolescents with disabilities.

Authors and Affiliations:

Alison Boardman, University of Colorado at Boulder

Pamela Buckley, University of Colorado at Boulder

Andrew Maul, University of Colorado at Boulder

Sharon Vaughn, University of Texas at Austin

Abstract Body

Limit 4 pages single-spaced.

Background / Context: *Description of prior research and its intellectual context.*

Collaborative Strategic Reading (CSR) is a set of research-based strategies designed to improve reading comprehension, enhance students' content area learning, facilitate access to higher-level texts, and to promote student engagement. CSR is theoretically grounded in cognitive psychology (Flavell, 1992) and combines elements of reciprocal reading (Palincsar & Brown, 1984) and cooperative learning (Kagan, 1986). CSR helps students develop metacognitive awareness and learn specific strategies associated with enhanced reading comprehension. It can be integrated across content areas since the strategies complement existing curriculum and align well with new textbooks. CSR has been implemented in diverse upper elementary, middle school, and high school general education classrooms, as well as in special education and reading intervention classes.

CSR has been evaluated over the last 15 years and has yielded positive outcomes for students with learning disabilities, students at-risk for reading difficulties, English language learners, as well as average and high achieving students (Bryant et al., 2000; Klingner et al., under review; Klingner et al., 2004; Klingner, Vaughn, & Schumm, 1998; Vaughn, et al., 2011). Fidelity – or the degree to which a program is implemented as intended (Lipsey, 1999; O'Donnell, 2008) – has been reported in CSR studies to describe implementation differences between treatment and control conditions (Klingner et al., under review; Vaughn et al., 2011). Vaughn and colleagues (2013) examined the effects of CSR implementation on student outcomes by empirically linking procedural fidelity to student outcomes. Each teacher taught two or more sections of language arts or reading and sections were randomly assigned to treatment or comparison condition (Vaughn, et al., 2013). Among the study's results were that CSR was associated with a greater effect on student reading comprehension when implemented in language arts classrooms compared to reading classrooms. Each of these studies indicates that student outcomes may be related to fidelity of implementation. Additional research on CSR fidelity is needed to explore these relationships across studies and settings.

Purpose / Objective / Research Question / Focus of Study: *Description of the focus of the research.*

The present study examines how fidelity of implementation of CSR is associated with reading outcomes for students with mild to moderate disabilities using data from two nonoverlapping studies conducted in middle school language arts and reading classrooms (study 1) and social studies and science classrooms (study 2). We use a definition of fidelity that includes both the amount of CSR instruction delivered by teachers and the quality of implementation. Our primary research questions include: 1) Is higher CSR instructional quality associated with increased student outcomes for adolescents in treatment classrooms, and for a subgroup of students with disabilities?, and 2) Is the amount of CSR instruction related to student outcomes for adolescents in treatment classrooms, and for a subgroup of students with disabilities? We replicated the analysis with two different samples (studies 1 and 2). We compared results to examine the stability of estimated relationships and to answer our secondary research question: How do the characteristics of study participants and program contexts influence the replication of findings?

Setting: *Description of the research location.*

Study 1: The first study took place in nine low income middle schools in the metropolitan areas of Denver, Colorado and Austin, Texas. Three of the middle schools were located in one district in Texas. In Colorado, five of the participating schools were from the same large, urban school district and one school was located in a smaller urban district.

Study 2: The second study was conducted in three middle schools located in the same large, urban school district in Colorado. Teacher and student participants were part of a school-district-and-university collaboration designed to evaluate the effectiveness of CSR within the context of a district initiative and scale up study. 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.

Population / Participants / Subjects: *Description of the participants in the study: who, how many, key features, or characteristics.*

Teachers: Study 1 initially included 20 middle school teachers and 41 sections of language arts or reading classes. The participants' average years of teaching experience was 10.12 (SD = 8.25). For study 2, each social studies class roster was paired with a science class roster within each school to create one social studies/science teacher pair. Thus, the study included 23 teachers that formed 12 teacher pairs and 30 sections of social studies/science class pairs. Thirty-six percent of the teachers in study 2 had taught for less than five years, whereas 34% had taught five to 10 years and 30% had taught for more than 10 years.

Students: Table 1 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 Latino and 10% received special education services. For the second study, 533 students were in the initial sample. Just over half were Latino (51%) and 12% received special education services.

Intervention / Program / Practice: *Description of the intervention, program, or practice, including details of administration and duration.*

Both studies used the same process where teachers first introduced each CSR strategy to the whole class using explicit instruction, modeling, think alouds and guided practice. Once students had learned the strategies, they applied them while working in cooperative groups and reading content area text with their peers. The studies differed in their administration and frequency. For study 1, language arts or reading teachers taught CSR once a week for approximately 50 minutes each week throughout the school year. The experimental treatment unit for study 2 was pairs of social studies and science teachers, and CSR was implemented one day a week for about 50 minutes in social studies and one day a week for 50 minutes in science classes throughout the school year. Teachers in each study received similar support for their implementation of CSR consisting of two days of professional development, two after school booster sessions and individual coaching (2-3x/month). Professional development consisted of 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, support in text selection, and managing cooperative learning in diverse classrooms.

Research Design: *Description of the research design.*

The samples for both studies included the treatment group from larger randomized control trials of CSR. In each of the larger studies, we conducted a cluster-randomized controlled trial with middle school teachers participating and serving as their own comparisons. Classrooms rather than students were randomly assigned to condition (i.e., the CSR group or the business-as-

usual group) in which half of each teacher's sections were randomly assigned to CSR and the other half were randomly assigned to business-as-usual. For the present research, we are interested in the relationship between CSR implementation (specifically quality and dosage) and student outcomes. Thus, we modeled effects for CSR treatment classes only. 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; Hambre et al., 2010; Kopp, Hulleman, Harackiewicz, & Rozek, 2012).

Data Collection and Analysis: *Description of the methods for collecting and analyzing data.*

Quality Implementation Measure: Teacher quality for both studies was assessed using the CSR Global Quality Rating Score measured through the *Implementation Validity Checklist* (IVC; Vaughn et al., 2011). For study 1, four IVC observations were conducted; meanwhile, study 2 included three IVC observations for each teacher or six observations in total. The IVC consists of items that break down the essential features of CSR in terms of student behaviors, teacher behaviors, and quality of instruction.

The first IVC section focuses on procedural fidelity in which teachers are rated on each component of CSR and scores range from one (inconsistent with CSR procedures) to four (highly aligned with CSR procedures). 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 zero (not observed) to three (observed). The third section included a global quality rating on a scale from one (low) to seven (high) to measure the overall quality of teacher's CSR implementation. To determine the CSR Global Rating, IVC observers were trained to review the fidelity scores and scores for the teacher and student behaviors and to combine them into one holistic score encompassing both procedural fidelity, quality of instruction, and how well the instruction was received by students.
For both studies, all IVC observers were experts in CSR implementation and participated in trainings to ensure interrater agreement of 90% or higher on the IVC items prior to beginning classroom observations. In addition, 10% of observations throughout each study were completed by two researchers to confirm interrater agreement in the field. Also, internal consistency for the CSR Global Rating on the IVC was .80 for study 1 and .91 for study 2.

Dosage Implementation Measure. Teachers in both studies completed a log to monitor the dosage or amount of CSR intervention children received. In this log, teachers reported the days they used CSR and the number of minutes they taught CSR each day.

Gates-MacGinitie Reading Test (GMRT). The student outcome for both studies was reading comprehension ability, as measured by the reading comprehension subtest of the Gates-MacGinitie Reading Test, Fourth Edition (MacGinitie, MacGinitie, Maria, & Dryer, 2004). Two parallel forms permit pre- and post-testing. The GMRT is a timed paper and pencil, group-administered survey that assesses student achievement in reading. It was administered by trained researchers 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. Multilevel (random-intercept) models, as implemented in the software program Hierarchical Linear Modeling 7.0 (HLM; Raudenbush, Bryk & Congdon, 2010) were used to estimate the effects of CSR on outcomes controlling for pretest scores and student demographics. For study 1, students were regarded as nested in language arts or reading classes nested in teachers. For study 2, students were nested in social studies and science section pairs nested in teacher pairs. Since we were interested in certain subgroups, student-level

covariates (special education, SpEd; and English language learner, ELL) were included in both models. Study 2 also modeled free or reduced lunch (FRL) status (a proxy for low-income) since we had access to these data. Using Raudenbush and Bryk's (e.g., 2002) three-level notation, both models took the following form:

Level 1 $\text{GMRT post} = \pi 0jk + \pi 1jk*(\text{SpEdijk}) + \pi 2jk*(\text{FRLijk}) + \pi 3jk*(\text{ELLijk}) + \pi 4jk*(\text{GMRT preijk}) + \text{eijk}$

Level 2 $\pi 0jk = \beta 00k + \beta 01k*(\text{Total Minutes jk}) + \beta 02k*(\text{CSR Global Quality jk}) + \text{r0jk}$
$\pi 1jk = \beta 10k + \beta 11k*(\text{ Total Minutes jk}) + \beta 12k*(\text{ CSR Global Quality jk})$

Level 3 $\beta 00k = \gamma 000 + \text{u00k}$

The GMRT scores were standardized in both studies to have a mean of 100 and standard deviation of 15. Quality was modeled using the 7-point CSR Global Rating scale score from the IVC averaged across each time point. The total number of minutes CSR was taught by each teacher (study 1) or teacher pair (study 2) was used to model dosage.

Findings / Results: *Description of the main findings with specific details.*

Dosage varied by study (Table 2). 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). Quality, however, was roughly equivalent (study 1: M=4.75, SD=0.84; study 2: M=5.29, SD=1.08).

Results across both studies show no statistically significant main effect of quality or dosage and no interaction effect between dosage and special education status. There was, however, a significant, positive interaction effect between quality and posttest scores for students with disabilities (Table 3). Controlling for all covariates, the estimated effect of special education status in study 1 is -3.85 points ($p < .05$). The total estimated effect of a one-point increase in quality of implementation for students with disabilities was 5.12; 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 of disability in study 2 is -2.06 points ($p = .126$), while the total estimated effect of a one-point increase in quality of implementation for special education students was 2.22, an estimate that is 2.27 points higher ($p <. 05$) than the estimated total effect for students without disabilities. Note 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.

Conclusions: *Description of conclusions, recommendations, and limitations based on findings.*

This study addresses the conditions under which causal relations are demonstrated to be replicable in different contexts (i.e., teachers volunteered to teach CSR in study 1 whereas teachers were required to teach CSR in study 2) and among diverse groups of teachers (study 1 included language arts and reading teachers whereas study 2 involved social studies and science teachers). Both studies found that higher quality CSR instruction is associated with higher student outcomes for students in special education. The standardized coefficients in study 1, however, were nearly double those in study 2. Further research is needed to determine possible explanations for these findings.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

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.

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

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 presentation at the Spring Conference of the Society for Research on Educational Effectiveness, Washington, DC.

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.

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.* Los Angeles: Evaluation, Dissemination and Assessment Center, California State University.

Klingner, J., Boardman, A., Buckley, P., Annamma, S., Jensen, C., & Maul, A (under review). Collaborative Strategic Reading in Content Classes: Results from Year 1 of a Randomized Control Trial. Manuscript submitted to the *American Educational Research Association Journal.*

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.

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 presentation 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.

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. & Congdon, R. (2010). *HLM 7 hierarchical linear and nonlinear modeling.* Scientific Software International.

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

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., 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., Mohammed, S. & Leroux, A. (2013): Collaborative Strategic Reading: Findings From Experienced Implementers, *Journal of Research on Educational Effectiveness, 6:2,* 137-163.

Appendix B. Tables and Figures
Not included in page count.

Table 1

Student Characteristics of Initial Sample for Study 1 (N = *608*) *and Study 2* (N = *533*)

	Study 1		Study 2	
	n	%	n	%
Gender				
Female	265	43.6	275	49.7
Male	332	54.6	278	50.3
Ethnicity				
Anglo	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 Learners	271	44.6	80	14.5
Special Education	61	10.0	67	12.1
Free or Reduced Lunch	--	--	364	65.8

Note. The percentages associated with ethnicity for study 1 may add up to more than 100 because students were allowed to be categorized to more than one ethnicity.

Table 2
Implementation Data for Study 1 and Study 2

Measure	Study 1		Study 2	
	M	*SD*	*M*	*SD*
Dosage – Total minutes	1194	442	2095	361
Dosage – Total minutes	17	9	42	5
Quality – Scale of 1 to 7 (Average across rounds)	4.75	.84	4.65	.99

Note. Study 1 collected 4 rounds of quality ratings on each teacher, and study 2 collected 3 rounds of quality ratings on each teacher pair for a total of six observations.

Table 3

The Effect of Dosage, Quality, Student Demographics and Gates Pretest Scores on Gates Posttest Scores for Study 1 and Study 2

	Study 1			Study 2		
	B	*SE*	*p value*	*B*	*SE*	*p value*
INTERCPT2, β0	93.67	0.73	<0.001	97.71	0.63	<0.001
Total Minutes, γ010	0.00	0.00	0.876	-0.00	0.00	0.322
CSR Global Quality, γ020	1.52	0.80	0.080	-0.05	0.72	0.945
SPED, β10	-3.85	1.41	0.007	-2.06	1.35	0.126
Total Minutes, γ110	-0.00	0.00	0.740	0.00	0.00	0.489
CSR Global Quality, γ120	3.61	1.59	0.024	2.27	1.14	0.046
FRL, β20	--	--	--	0.24	0.98	0.808
ELL, β30	-0.53	0.77	0.487	-1.18	1.14	0.301
GMRT pretest, β40	0.74	0.03	<0.001	0.69	0.03	<0.001

Note. SpEd – Special Education; FRL – Free and Reduced Lunch; ELL – English Language Learner; GMRT – Gates-MacGinitie Reading Test. FRL data was not collected for Study 1.