Abstract Title Page

Title: Collaborative Strategic Reading: Replications with Consideration of the Role of Fidelity.

Authors and Affiliations:

Sharon Vaughn
Greg Roberts
Colleen Reutebuch
The Meadows Center for Preventing Educational Risk, The University of Texas at Austin

Abstract Body

Background /Context: While there have been numerous calls over the years for an increased focus on reading comprehension (e.g., Snow, 2002; Biancarosa & Snow, 2006), it appears that adolescents' reading comprehension is not improving (National Center for Educational Statistics, 2009). Collaborative Strategic Reading (CSR; Klingner, Vaughn, Dimino, Schumm & Bryant, 2001) is a multicomponent reading intervention aimed at improving students' text comprehension, with evidence of its efficacy established through quasi-experimental research studies conducted over the last 15 years. Recently, two 1-year randomized controlled trials were conducted to determine the efficacy of CSR with seventh and eighth grade students. The Year 2 replication study was identical to the original Year 1 study except that the student sample was different and the teachers were more experienced in the Year 2 study.

Purpose / Objective / Research Question / Focus of the Replication Study: In year 1, we conducted a study to examine the effects of CSR on the reading comprehension for adolescent readers. During year 2, we replicated the year 1 study with a new cohort of students *taught by the same teachers* (therefore, teachers who were more experienced with the intervention) to determine the relative effects of CSR with relatively experienced teachers on the reading comprehension outcomes of students. Our primary research question for the replication study was to determine the efficacy of CSR on middle school students' reading comprehension when taught by relatively experienced teachers (year 2) and to contrast the impacts to those obtained when similar middle school students were taught by inexperienced CSR teachers (year 1).

Secondarily, we were interested in exploring the role of implementation fidelity on students' outcomes and comparing fidelity across the two years. The importance of implementation fidelity is widely acknowledged (e.g., Hulleman & Cordray, 2009; Swanson et al., in review). When applied to both treatment and comparison conditions, fidelity measures can be used to investigate not only the extent to which implementation of the treatment has occurred, but also the extent to which this treatment is different from the condition to which it is being compared, sometimes called "achieved relative intervention strength" (Hulleman & Cordray, 2009: 88). Building on this growing body of research, this study incorporated fidelity into multilevel analyses in order to more fully explore and document the relationship between implementation of CSR strategies by experienced teachers and middle school students' reading comprehension.

Setting: This study was conducted in 6 middle schools in 3 school districts (two near urban and one urban) in Texas and Colorado that reflected a diverse student population who were provided daily English/language arts/reading instruction.

Population / Participants / Subjects: During year 1, teachers provided seventh and eighth grade students in this study with English/language arts/reading instruction on a daily basis. Seventeen teachers and 61 classes (treatment = 34, control = 27) participated in the study during year 1. Due to changes in teaching assignments, five teachers did not continue after Year 1. Therefore, twelve teachers and 48 classes (treatment = 26, control = 22) participated during Year 2. Students in this study were two separate cohorts of seventh and eighth graders enrolled in English/language arts/reading classes. Student participants totaled 782 (treatment = 400, control = 382) students in Year 1 and 528 (treatment = 264, control = 264) in Year 2.

Intervention / Program / Practice: <u>Description of Intervention</u>. 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. The *Preview* activities aim to build and activate prior knowledge and to motivate students' interest about the passage topic. 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 (e.g., "Read the sentence before and after the clunk. Look for cues.") Students also *get the gist,* which is similar to 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 understand what they read and remember what they learned. *Wrap up* takes place after reading, to help students generate and answer questions about what they have read, and summarize key ideas presented in the text. Initially, the teacher presents the strategies to the whole class using explicit instruction, modeling, and teacher think-alouds. After students develop proficiency using the strategies (4–6 weeks), the teacher assigns them to cooperative learning groups of four to five students.

 <u>Administration and Duration</u>. Teachers reported that the number of sessions they implemented ranged from 24 to 48 sessions for Year 1 and between 18 and 61 sessions for Year 2. During both years focused professional development was offered 3 days before school began, followed by three 90-minute booster sessions distributed throughout implementation. Thus the key difference between the implementation in the initial study and the Year 2 Replication study was level of teacher experience with the intervention. Other factors remained constant.

Research Design: We conducted a randomized control trial at two sites (Colorado and Texas) for both studies. Students in seventh and eighth grade English and reading classrooms (61 classes during Year 1 and 48 classes during Year 2) were randomly assigned to a class and then classes were randomly assigned a teacher. For teachers with an odd number of classes, the additional class was assigned to the treatment condition. Typical instruction (business-as-usual-instruction) was provided for students who were randomly assigned to comparison classes.

 Because the same teacher provided instruction for both the treatment and comparison conditions with the students randomly assigned to condition, we controlled the effect of an individual teacher accounting for a significant amount of variance. To guard against treatment contamination into comparison classrooms, 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.

Data Collection and Analysis: The same set of pretest and posttest measures were administered prior to treatment and immediately following treatment for both. The reading achievement battery included the TOWRE, the Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner, Torgesen, Rashotte, & Pearson, 2010), Test of Sentence Reading Efficiency (TOSRE; Wagner, Torgesen, Rashotte, in press), AIMSweb Maze passages for 7th and 8th Grades (AIMSweb Maze-CBM, 2009), and the Gates-MacGinitie Reading Test (Gates & MacGinitie, 2000). We also collected data on student characteristics (e.g., language and special education status, age, gender, ethnicity, reading proficiency) to examine comparability of groups. Students were considered "struggling" 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 Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999). All student

measures were administered by trained research personnel who were blind to students' condition (treatment or comparison). Fidelity measures included the Internal Validity Checklist (IVC; Vaughn et al., 2011), evidence of strategy use in typical comparison classrooms, and implementation logs.

     Data Analysis. For 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 offered the advantage of a direct full information maximum likelihood (FIML) estimator of missing data, more appropriate modeling of the covariance structures of clustered data, a flexible framework for analyzing the effects of covariates, and estimates of model fit. The teacher was treated as a stratum for the purposes of assignment, and classes (both treatment and comparison) were randomly assigned among teachers. Analytically, this represents a randomized block design with teachers as the blocking variable (Raudenbush, 1997) and students nested in classes. A pretest score was 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. Classes were represented as clusters, which define levels in a multilevel model. We modeled posttest means as latent factors on the between-classes model and treatment condition using the multiple groups option in Mplus, which allowed for formal tests of statistical significance using a nested models comparison.

     For fidelity analysis in Year 2, confirmatory factor analysis with categorical indicators was used to estimate factors related to fidelity and spillover and the mediating effect of implementation was evaluated. Six latent variables were specified, each corresponding to a key CSR-aligned teacher practice: *Brainstorming*, *Preview*, *Click and clunk*, *Fix-up* Strategies, *Question generation*, and *Summarizing*. A higher-order factor representing overall alignment of intended and enacted models was also specified. Observers scored fidelity on a 4-point scale during each of the eight observations, with a score of 1 representing the absence of a given element and a score of 4 representing its full implementation according to the intended model.

Findings / Results: Initial Study (Year 1). Main effects were estimated for the Gates-MacGinitie, the AIMSweb maze, and the TOSREC according to the multilevel model. The analyses were conducted with the entire sample and with the sample of students identified at pretest as struggling readers. Group differences on AIMSweb ($\Delta\chi^2 = 1.13$, $\Delta_{df} = 1$) and TOSREC ($\Delta\chi^2 = .41$, $\Delta_{df} = 1$) were not statistically significant. The model-estimated (Level-2 latent) posttest average standard score on the Gates-MacGinitie was 95.87 for comparison classes and 97.04 for the treatment conditions ($\Delta\chi^2 = 9.91$, p < .01). This is equivalent to an (bias corrected Hedges) effect size of g = 0.12. Results for the sample of low-achieving students (with TOWRE as a selection criterion) were similar to those for the total sample. The model-derived posttest score on the Gates-MacGinitie was 87.66 for CSR participants, about 3.14 standard score points greater than initially 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).

     Replication (Year 2). The findings of the replication study revealed no statistically significant impact for students in CSR classes over students in their typical classes on reading comprehension. The difference in levels of fidelity across the two groups (i.e., CSR and comparison) was statistically significant ($\Delta\chi^2 = 34.97$, $\Delta_{df} = 1$, p < .001), suggesting that CSR was more prevalent, on average, in the treatment classes. There were no statistically significant

group differences on the AimsWeb standard score (b = -.51, p = .72), or the TOSRE standard score (b =.84, p =.44). The unstandardized Gates coefficient was .50 (p = .836), and the standardized value, β, was .10, about 10% of a standard deviation. While not significant, the effect size in year 2 approximates the effect in year 1. There were no statistically significant group differences for low-performing readers (51 students across 27 classes within 11 teachers) on the Gates McGinnitie (b =.071), AimsWeb (b =-2.91, p =.117), or TOSRE (b =1.58, p =.536). Further, as anticipated, the mediated model results related to posttest means do not differ substantively from those of the intent-to-treat analyses. There were no statistically significant differences in estimated Gates-McGinnitie posttest scores ($\Delta\chi^2$ = .259, Δdf = 1, p=.61), on the AEMSWeb ($\Delta\chi^2$ = 1.54, Δdf = 1, p=.28), or the TOSRE ($\Delta\chi^2$ = Δdf = 1, p=). Similarly, when the regression of student outcomes on fidelity was constrained as equal across treatment groups (a constraint that *hypothetically* eliminates any difference due to assignment), the expected pattern emerged. Differences on the Gates McGinnitie ($\Delta\chi^2$ = 4.18, Δdf = 1, p=.04) and TOSRE were statistically significant, suggesting that fidelity of implementation mediated the effect of assignment on outcomes and that the fidelity was reasonably well established and that spillover was relatively minimized. The difference for AEMSWeb was not statistically significant ($\Delta\chi^2$ = 3.08, Δdf = 1, p=.08).

Conclusions: We replicated the initial Collaborative Strategic Reading study in Year 1 with a 1-year randomized trial in Year 2. Similar effect sizes were evident across the two trials. In the second experiment, we focused on documenting implementation and demonstrating its prediction to student outcomes. We collected fidelity data on features of the normative program model that are essential, hypothetically, to its effect, as a means of identifying critical program components. Evaluating the relative effects of a program's components is useful, generally, but may be particularly meaningful in the context of a replication (Walker, 1971), where an estimate of the treatment's effect (from the Year 1 trial) can be considered according to the relative contributions of the model's key components (i.e., Year 2 trial). In this case, the effect size was replicated (largely) across two years. The Year 2 trial suggested that implementation was high in the treatment conditions and relatively low in the comparisons. Our findings also indicate that implementation was consistent across program elements; treatment conditions had high implementation of all program elements, on average. An experimental manipulation of program elements would be necessary to address this issue fully. Nonetheless, our findings represent a model for replicating effective intervention in the context of evaluating the relative contribution of its respective components.

Appendices

Appendix A. References

AIMSweb Maze-Curriculum Based Measurement (n.d.). Retrieved from http://www.aimsweb.com/measures-2/maze-cbm/.

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.

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

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: Erlbaum.

Durkin, D. (1978-1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly, 14*(4), 481-583.

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

Gelzheiser, L. M., & Myers, J. (1991). Reading instruction by classroom, remedial and resource room teachers. *The Journal of Special Education, 24*(4), 512-526.

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*(1), 88-110. doi:10.1080/19345740802539325

Johnson, D. W., & Johnson, R. T. (1989) Cooperative learning: What special educators need to know. *The Pointer*, 33, 5-10.

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

Klingner, J. K., Vaughn, S., Dimino, J., Schumm, J. S., & Bryant, D. P. (2001). *From clunk to click: Collaborative Strategic Reading.* Longmont, CO: Sopris West.

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

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

Scanlon, D., Deshler, D. D., & Schumaker, J. B. (1996). Can a strategy be taught and learned in secondary inclusive classrooms? *Learning Disabilities Research and Practice, 11*(1), 41-57.

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

Swanson, E. A., Wanzek, J., Vaughn, S., Haring, C., Ciullo, S.,& McCulley, L. (in review). Intervention fidelity in special and general education research journals. *Journal of Special Education.*

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *The Test of Word Reading Efficiency (TOWRE).* Austin, TX: Pro-Ed.

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

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

Appendix B. Tables and Figures

None.