Abstract Title Page

Title: Impact of the Teacher Study Group Professional Development Program on Student Vocabulary and Observed Teaching Practice: A Replication in First Grade Classrooms.

Authors and Affiliations:

Russell Gersten
Joseph Dimino
Madhavi Jayanthi
Rebecca Newman-Gonchar
Mary Jo Taylor
Instructional Research Group

Abstract Body

Limit 4 pages single-spaced.

Background / Context: The ESEA Blueprint for Reform (U.S. Department of Education, 2010) states that teachers need "effective, ongoing, job-embedded, professional development that is targeted to student and school needs… [and] aligned with evidence of improvements in student learning." Unfortunately, the professional development approaches advocated, though sensible and compelling in theory, have rarely been widely field-tested and evaluated using rigorous research techniques. The situation is slowly changing (e.g., Garet et al., 2008; Powell, Diamond, Burchinal, & Koehler, 2010), but often the results have been mixed.

One promising professional development effort in recent years has been the Teacher Study Group (TSG) (Gersten, Dimino, Jayanthi, Kim, & Santoro, 2010). This approach to professional development is an attempt to orchestrate several of the "best practices" in professional development–linkage to core curriculum, concreteness, establishment of collegial networks, and ongoing related activities–into a feasible model for use in elementary schools. The goal of the TSG professional development program is to enhance instruction by helping teachers integrate research-based instructional strategies into their existing curriculum (Gersten, Woodward, & Morvant, 1992).

The effectiveness of the TSG professional development program on teacher and student outcomes in the areas of comprehension and vocabulary was initially examined via a small-scale randomized controlled trial (Gersten et al., 2010). Despite the weak statistical power due to small sample size, the findings demonstrated significant positive impacts on observed teaching practice[2] (.58, $p < .01$) and teacher knowledge[3] (.73, $p < .05$). The study also yielded potentially positive impacts on student vocabulary[4] (.44, $p < .10$).

In the earlier small-scale randomized controlled trial (Gersten et al., 2010), post-doctoral-level staff with strong backgrounds in reading research conducted the TSG sessions. The next logical step in the evolution of this line of research is to conduct a replication study in which TSG sessions are facilitated by literacy coaches or mentor teachers from the schools, rather than by research staff. This replication study would help determine the feasibility of literacy personnel as facilitators of the TSG sessions and also help examine the impact of the TSG sessions led by literacy personnel on teacher and student outcomes.

To this end, a replication study (i.e., a large-scale randomized controlled trial) was conducted with first grade teachers to examine the effectiveness of the TSG program led by literacy personnel on observed teaching practice, teacher knowledge, and student vocabulary achievement.

Purpose / Objective / Research Question / Focus of Study: The purpose of the study is to examine the impact of the Teacher Study Group, focused on effective vocabulary instruction, on teacher knowledge, observed teaching practice, and student vocabulary achievement when implemented with first grade teachers in Title I schools. Our major research questions were:

[2] Measured using the *Observation Measure for Vocabulary Instruction* (Gersten et al., 2010).
[3] Measured using the *Content Knowledge for Teaching Reading* (Phelps & Schilling, 2004).
[4] Measured using the *Oral Vocabulary* subtest of the *Woodcock Diagnostic Reading Battery*.

- *Question 1*: What is the impact of the TSG on teacher knowledge and teaching practice when compared to the professional development efforts being provided by the states and districts?
- *Question 2*: What is the impact of the TSG on students' vocabulary knowledge when compared with students in classes receiving existing professional development efforts?
- *Question 3*: What role does fidelity of implementation of the TSG protocols by the facilitators play in moderating the impacts of TSG on teacher and student outcomes?
- *Question 4*: What contextual factors facilitate or hinder effective implementation of the TSG?

Setting: The randomized controlled study was conducted in 61 Title 1 schools (31 treatment and 30 control) from 16 school districts in four states: California, Ohio, Illinois, and Texas. The mean percentage of students eligible for free and/or reduced lunches was 77% in both treatment and control schools and the mean percentage of third grade students at or above proficient on the state reading test was 64% in treatment schools and 57% in control schools.

Population / Participants / Subjects: The sample consisted of 182 first grade teachers (94 treatment and 88 control) and a randomly selected sample of 1811 students (940 in treatment and 871 in control). Table 1 and 2 provide detailed information on the student and teacher demographics. (please insert Tables 1 and 2 here)

Intervention / Program / Practice: The TSG intervention is a concentrated professional development effort designed to improve first grade teachers' teaching practice and increase student vocabulary outcomes. The TSG intervention consists of 11 interactive sessions held at the school site twice a month, starting in October. Sessions were conducted before or after school for approximately 75 minutes, to maximize instructional time during the school day. The set of topics covered in the TSG sessions include: Words in Context; Selecting Words to Teach; Student Friendly Definitions; Examples, Contrasting Examples, and Concrete Representations; Activities to Promote Word Learning; Using Context to Determine Word Meaning; and Reviewing and Extending Word Learning.

During each session, a five-phase recursive process was used to explore a research-based vocabulary concept and integrate it into the teachers' lesson planning: (1) *Debrief:* Participants describe the collaboratively-planned lesson they taught, report on any changes they made while teaching the lesson, and discuss how students responded. (2) *Discuss the Focus Research Concept*: Participants review, reflect and discuss the new research concept. (3) *Compare the Focus Research Concept with Practice*: Participants compare how the research aligns with the instructional recommendations for teaching content vocabulary in their curriculum. (4) *Plan Collaboratively:* Participants collaboratively plan a lesson by incorporating the focus research concept into the lesson. (5) *Assignment*. Participants are asked to implement the lesson they planned collaboratively before the next TSG session.

TSG sessions were facilitated by literacy personnel from each school. The TSG facilitator's guide (Dimino & Taylor, 2009) was used to provide the facilitator with a specific "game plan" for leading participants through the five-phase recursive TSG process. Facilitators attended a 2-day training and were provided with ongoing support as needed.

Research Design: A multi-site cluster randomized trial design was used, in which schools were randomly assigned within participating school districts (Donner & Klar, 2000; Shadish, Cook, & Campbell, 2002) to treatment and control condition. First grade teachers in treatment schools participated in the TSG professional development. The control condition (business-as-usual) constituted school- or district-instituted professional development. Teachers in the control condition were not engaged in the TSG or did not have access to the materials made available to teachers in the TSG condition during the course of the study. Data on the professional development activities (hours, type, and content) in reading, particularly in vocabulary, in both the TSG and control conditions was collected to compare the nature of professional development activities available in both conditions.

Baseline Equivalence**.** Random assignment of schools yielded treatment and control groups that were similar at baseline on all demographics and pretest measures except gender of the student sample. There was a higher percentage of females in the control condition (χ^2 =5.67, p=.015). See Tables 1 and 2 above for more details.

Data Collection and Analysis: In each district, data were collected from both the TSG and control schools at the same time, by observers and testers blind to condition, to guard against bias entering the data collection process. Prior to the start of the intervention in the fall, data were collected on teacher demographics (e.g., teacher experience and education). At the end of the study in the spring, a cadre of trained observers and data collectors collected data on teaching practice using the *Observation Measure of Vocabulary Instruction* (*OMVI*). All teachers (from both TSG and control conditions) were observed once during the entire language/arts block; 50% of the teachers were observed twice. Inter-observer reliability data was collected on 20% of all observations. Immediately after the end of the TSG intervention, teachers completed the vocabulary knowledge measure (*Content Knowledge for Teaching Reading*).

Student measures (*Woodcock Johnson* and *Group Reading Assessment and Diagnostic Evaluation*) were administered by trained evaluation staff to a randomly selected sample of students. Student pre-test data were collected 4 weeks after the start of the school. Post-test data were collected 4-6 weeks before the end of the school year. Student and school demographic data (EL status, free and reduced lunch, AYP status, ethnicity) was gathered from the school databases.

Data Analysis. Hierarchical linear modeling (HLM) was used to perform the main impact analyses as the data used are of a nested nature, that is, students and teachers nested within schools (Raudenbush & Bryk, 2002). For both confirmatory and exploratory analyses of the TSG's impact on teacher outcomes, a two-level HLM model was used, with teachers at level 1 and schools at level 2. The impact analyses for student outcomes were based on a three-level HLM model with students at level 1, teachers at level 2, and schools at level 3. Exploratory analyses include an examination of relevant mediating factors and potential moderator variables.

In addition to the statistical significance of the intervention's effects, the magnitude of the effects was gauged. Specifically, the effect size was computed as a standardized mean difference (Hedges' *g*) by dividing the adjusted group mean difference by the unadjusted pooled within-group standard deviation of the outcome measure.

Findings / Results: Preliminary analyses indicate positive significant impacts at the teacher level, with a *g* of 0.43 (p<.001) for teaching practice and a *g* of 0.31 (p<.001) for teacher

knowledge. These are slightly smaller than the impacts when expert facilitators conducted the TSGs in the small-scale randomized controlled trial, but still appreciable and significant[5].

Preliminary estimates for student outcomes show non-significant impacts. While the small-scale randomized controlled trial showed promise for positive effects on student vocabulary growth[6], the finding, unlike the finding on teaching practice, was not replicated. The research team is surprised by the lack of significant findings. Reasons for the lack of effects could be that the standardized measures used are not sensitive enough to detect changes or that the vocabulary instruction did not transfer beyond the target words that were taught (Elleman, Lindo, Morphy, & Compton, 2009; Pearson, Hiebert, & Kamil, 2007). Another reason could be that the majority of schools in the first study were inner city schools with extremely high levels of poverty and low academic achievement, while the larger replication study included a much broader range of schools—rural, suburban, and schools with varying levels of poverty and academic achievement. The study team is in the process of exploring site variability in outcomes and potential relationships between procedural fidelity of the TSGs and outcomes. The research team is also exploring if the impacts are higher for students or classes entering with the most severe entry-level skills.

Fidelity of Implementation. Mean fidelity of treatment, calculated on 20% of the sessions, was 88% (median = 90%; range = 54% to 100%). The procedural fidelity data indicate that school-based facilitators varied in their implementation fidelity more than the expert facilitators used in the previous small-scale study[7], as one might expect. In fact, there were some cases of very weak procedural fidelity. Nonetheless, the fidelity data suggest that trained school level personnel can, in most cases, facilitate the TSGs.

Conclusions: The purpose of this replication study was to examine the impact of the Teacher Study Group, focused on effective vocabulary instruction, on teacher knowledge, observed teaching practice, and student vocabulary achievement when implemented with first grade teachers in Title I schools by literacy personnel. Preliminary findings demonstrate the potential of the TSG intervention to bring about significant positive impacts at the teacher level (a proximal outcome for a professional development intervention). In this session, the presenters will (a) share the findings and discuss lessons learned from this replication, and (b) discuss possible reasons for differences in impacts from this replication study and the previous small-scale study. However, unlike the original study, this study did not detect any significant improvement in student vocabulary knowledge or related aspects of reading. We hypothesize that this may be due to the less knowledgeable and skilled facilitators or the group, but remain puzzled by the results. Any results of ongoing secondary analyses will be shared with the audience, if exploratory analyses reveal any relevant information that might help us better understand reasons for this phenomenon.

---

[5] Impacts in the small-scale RCT were .58 ($p < .01$) for observed teaching practice and .73 ($p < .05$) for teacher knowledge (Gersten et al., 2010).

[6] In the small-scale study, an effect of .44, ($p < .10$) was found on the *Oral Vocabulary* subtest of the *Woodcock Diagnostic Reading Battery.*

[7] In the small-scale study, fidelity means for each TSG session ranged from 83.3% to 93.8%, with a mean of 86.5% (Gersten et al., 2010).

Appendices

Appendix A. References.

Dimino, J., & Taylor, M. J. (2009). *Learning how to improve vocabulary instruction through teacher study groups*. Baltimore, MD: Paul H. Brooks Publishing Co.

Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomized trials in health research*. London, England: Arnold.

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). Instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2,* 1-44. doi: 10.1080/19345740802539200

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., . . . Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement* (NCEE 2008-4030). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gersten, R., Dimino, J., Jayanthi, M., Kim, J., & Santoro L. (2010). Teacher study group: Impact of the professional development model on reading instruction and student outcomes in first grade classrooms. *American Educational Research Journal, 47*, 694-739.

Gersten, R., Woodward, J., & Morvant, M. (1992). Refining the working knowledge of experienced teachers. *Educational Leadership*, 49, 34-39.

Pearson, P., Hiebert, E., & Kamil, M. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, *42*(2), 282-296.

Phelps, G., & Schilling, S. (2004). Developing measures of content knowledge for teaching reading. *The Elementary School Journal, 105*, 31-48.

Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on head start teachers and children. *Journal of Educational Psychology, 102*, 299-312.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2010). *ESEA blueprint for reform.* Washington, DC: Author. Retrieved from http://www2.ed.gov/policy/elsec/leg/blueprint

Appendix B. Tables and Figures

Table 1. Baseline equivalence of teacher demographics

	Intervention (n = *94)*	*Control* (n = *88)*	χ^2	t	p[a]
Years Teaching–Total[a] (SD)	10.1 (7.26)	9.9 (7.73)		0.25	.805
Years Teaching–Grade 1[a] (SD)	74.5 (15.63)	73.6 (16.00)		0.38	.875
Years Teaching–Current School[a] (SD)	10.1 (7.26)	9.9 (7.33)		0.25	.805
Gender – Female	94.7	94.3	0.0115		.915
Percentage race/ethnicity[b]			0.7304		.948
American Indian/Asian/Multiracial/Other	5.3	5.7			
Black	5.3	4.5			
Hispanic	25.5	25.0			
White	60.6	59.1			
Not Reported	3.2	5.7			
Highest Education Level Attained			0.6501		.722
BA	37.2	40.9			
MA	26.6	21.6			
>MA	36.2	37.5			

* Statistical significance set at $p < 0.05$, however no comparisons reached this level.

Note: Percentages may not sum to 100 because of rounding. *SD* is standard deviation; *t* is the *t*-statistic resulting from a two-sample t-test; *p* is the probability level associated with the level of the *t*-statistic or χ^2.

a. 1 treatment teacher did not report.

b. Districts reported race/ethnicity in seven categories: American Indian, Asian, Black, Hispanic, Other, White and Multiracial. Due to small sample sizes, the American Indian, Asian, Multiracial and Other categories have been collapsed in this table. Unless otherwise noted, Black includes African American, Hispanic includes Latino, Asian includes Native Hawaiian or Other Pacific Islander, and American Indian includes Alaska Native.

Table 2. Baseline equivalence of student and school level variables

	Intervention (n = *940)*	*Control* (n = *871)*	χ^2	t	p[a]
	Baseline Assessments				
WIF_Score (SD)	10.5 (12.56)	9.6 (11.5)		1.63	.103
LNF_Score (SD)	49.0 (15.70)	73.6 (16.17)		0.64	.520
WJ_Read_Pre (SD)	451.3 (13.60)	450.7 (13.11)		0.84	.401
WJ_Oral_Pre (SD)	457.07 (14.14)	457.06 (14.14)		0.03	.976
GRADE_List_Pre (SD)	13.46 (2.98)	13.46 (3.01)		0.03	.980
GRADE_Word_Pre (SD)	17.13 (6.35)	16.84 (6.37)		0.97	.335
	Demographics				
Gender – Female[a]	47.9	53.5	5.8672		.015*
	School-Level Demographics[b]				
	Intervention Schools (n = *31)*	*Control Schools* (n = *30)*		t	p[a]
Percentage race/ethnicity American Indian/Asian/Multiracial/Other	8.9 (8.07)	9.0 (10.42)		-0.03	.976
Black	14.6 (19.00)	12.9 (15.09)		0.41	.681
Hispanic	33.5 (39.98)	40.6 (41.92)		-0.68	.498
White	42.9 (35.93)	37.6 (35.80)		0.58	.559
Proportion of students proficient on 3rd grade Reading	63.6 (25.3)	57.3 (25.06)		0.98	.330
Proportion of LEP[c]	38.2 (29.30)	40.4 (6.32)		-.0.24	.813
Proportion eligible for Free-Reduced School Lunch	76.8 (15.74)	76.5 (15.50)		0.07	.944

* Statistical significance set at $p < 0.05$.

Note: Percentages may not sum to 100 because of rounding. *SD* is standard deviation; *t* is the *t*-statistic resulting from a two-sample t-test; *p* is the probability level associated with the level of the *t*-statistic or χ^2.

a. Gender was not reported for 2 treatment and 3 control students.

b. These represent average percentages reported at the school level. Each percentage is an average of the school-level averages within that condition.

c. LEP findings are reported for 18 treatment and 20 control schools that reported valid data.

Districts reported race/ethnicity in seven categories: American Indian, Asian, Black, Hispanic, Other, White and Multiracial. Due to small sample sizes, the American Indian, Asian, Multiracial and Other categories have been collapsed in this table. Unless otherwise noted, Black includes African American, Hispanic includes Latino, Asian includes Native Hawaiian or Other Pacific Islander, and American Indian includes Alaska Native.