Looking More Deeply: Fidelity of Implementation as a Critical Component in Evaluating Intervention Impacts

Margaret J. Hornbacher, Beverly J. Dretzke, Kristin A. Peterson, and Meghan C. Hickey

Center for Applied Research and Educational Improvement

University of Minnesota

Paper Presented at the Annual Meeting of the American Educational Research Association,

New York, March 28, 2008

Correspondence to:
Judy Hornbacher, Ph.D.
University of Minnesota
Center for Applied Research and Educational Improvement (CAREI)
College of Education and Human Development
360J Education Sciences Building
56 East River Road
Minneapolis, MN 55455-0364
E-mail: horn0017@umn.edu
Phone: 612-626-7261; Fax: 612-625-3086

Abstract

This study investigated the use of fidelity of implementation measures in concert with standardized tests in a matched-pairs, quasi-experimental design for evaluation of a pilot intervention program that was designed to increase second and third graders' mathematics and reading achievement. Although students in pilot classrooms characterized by high fidelity generally performed at the same level as control students , students in low fidelity pilot classrooms performed at a significantly lower level compared to both control students and students in high fidelity pilot classrooms. We found that the fidelity measures allowed more in-depth analysis of the intervention's component parts and increased the confidence with which the project's major questions could be addressed, thereby providing more useful information to school district personnel.

Looking More Deeply: Fidelity of Implementation as a Critical Component in Evaluating Intervention Impacts

In any school district, there can be a great deal of variation in performance from one student to the next and from one school to another. In a world of high stakes accountability, teachers, administrators, and policy makers have begun to examine factors that contribute to achievement outcomes and to implement a wide variety of programs and interventions designed to address the needs of learners and the demands of ever increasing accountability standards.

Prior to the 1970's, program and intervention adoption was seen as an "empty vessel" endeavor in that it was assumed that professionals, after completion of training, would launch a new reform with absolute fidelity and would continue implementing with unwavering exactitude (e.g., Blackmore, 1998; O'Donnell, Lynch, Watson, & Rethinam, 2007; Rogers, 2003). Adopters were "considered to be rather passive acceptors of an innovation, rather than active modifiers of a new idea" (Rogers, 2003, p. 180). However, hallmark studies of educational programs, such as the Rand report on the implementation of educational innovation, found that teacher-proof programs did not exist in practice and that there was a consistent lack of implementation integrity in school programs (Berman & McLaughlin, 1976). Furthermore, recent research has shown that teachers frequently make modifications, even when the program is highly specified and program developers have demanded strict compliance (e.g., Datnow & Castellano, 2000).

The important concept of fidelity of implementation first appeared in the literature mainly as a tool to account for negative or ambiguous findings (e.g., Hohmann & Shear, 2002; Sechrest, West, Phillips, Redner, & Yeaton, 1979). Additionally, researchers have emphasized fidelity of implementation as a means to "adopter-proof" programs and describe how well aspects of a

program or intervention were being put into practice as well as document deviations from or differences within a particular model (Bond, Becker, & Drake, 1997). This has allowed researchers to determine if unsuccessful outcomes were due to a failure of the model or a failure to implement the model as intended (e.g., Chen, 1990; Sánchez, Steckler, Nitirat, Hallfors, Cho, & Brodish, 2007).

Fidelity of implementation or program integrity measures have been primarily used in the mental health and prevention fields (e.g., Bond, Evans, Salyers, Williams & Kim, 2000) with some effort made to codify measures and criteria. Dane and Schneider (1998), in their meta-analysis of the use of program integrity measures in primary and early secondary prevention studies, concluded that there are five primary aspects of program integrity: 1) *Adherence* or the extent to which specified program components are delivered as prescribed in program manuals; 2) *Exposure* or an index that may include any of the following: a) the number of sessions implemented; b) the length of each session; or (c) the frequency with which the program techniques were implemented; 3) *Quality of delivery* or a measure of qualitative aspects of program that are not directly related to the implementation of prescribed content; 4) *Participant responsiveness* or a measure of participant reaction to program sessions; and 5) *Program differentiation* or a manipulation check that is performed to safeguard against the diffusion of treatments to ensure that the subjects in each experimental condition received only planned interventions.

An increasing emphasis on fidelity of implementation in education studies has been prompted in part by requirements of the U.S. Department of Education. In the 2007 Request for Applications, the Institute of Education Sciences (IES) states, “Investigators should propose research designs that permit the identification and assessment of factors impacting the fidelity of

implementation" (p. 60). Dusenbury, Brannigan, Falco, and Hansen (2003) also stressed the importance of including fidelity measures in evaluation work, especially of that conducted in schools. They concluded that it is important to include a variety of measures of fidelity (e.g., adherence, dose, quality of program delivery) and to identify the critical elements of effective programs.

Current Study

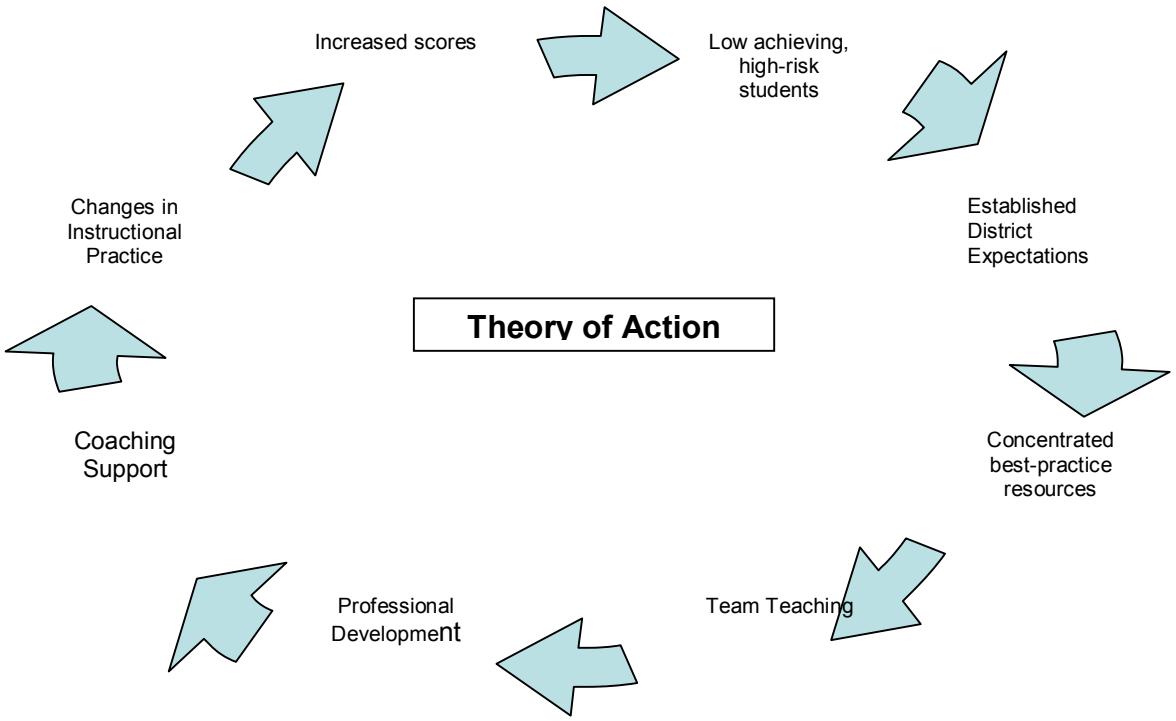
In 2005, a large Minnesota school district contracted with the Center for Applied Research and Educational Improvement (CAREI) at the University of Minnesota to evaluate the efficacy of a pilot program for which the district had received state funding. The district, like so many others, was seeking ways to improve student achievement in reading and mathematics and to narrow the achievement gap between at-risk students and those with fewer risk factors.

Our task was to examine the many components of the district intervention plan and the relationship, if any, of those various components to student achievement in the second- and third-grade classrooms in three pilot sites. The goal of the evaluation was to be able to provide the client district with strong data that could inform its decisions about revising and/or scaling up the model.

The district had adopted the *Balanced Literacy* (Reading: Planning Guide for Balanced Literacy, 2006) curriculum for reading and *Investigations Math* (Investigations in Number, Data, and Space, 2007) for the mathematics curriculum. There was a strong assumption on the part of the district that these curricula were being faithfully implemented in the classrooms. CAREI's evaluation addressed adherence to the components of the curricula as well as the impact of implementing strategies within those curricula.

The components of the pilot intervention included a) teaming, b) a prescribed teaching time schedule, c) classroom coaching, and d) intense, frequent professional development focused on math and literacy interventions within existing curricula. The hypothesis was that, using the district mathematics and reading curricula and being exposed to teaming, coaching, and professional development, teachers would change their practices resulting in increased student achievement in mathematics and literacy (See Figure 1).

Figure 1. Theory of Action



The initial evaluation design included descriptive classroom observations but data analysis focused on standardized achievement test outcomes. By the end of the first year, we became aware that, while comparisons between test scores of pilot program students and matched controls yielded interesting results in some subgroups, they did not provide information on the components of this complex intervention that might actually be influencing student outcomes. Therefore, in the second year, we added fidelity of implementation and fidelity of

process measures to the observation protocol. Fidelity of implementation in the present study was akin to Dane and Schneider's (1998) concepts of adherence and exposure, and fidelity of process was akin to their concept of quality of delivery. We expected that fidelity of process measures would show a stronger relationship with student achievement outcomes than fidelity of implementation measures.

Four major questions were addressed in this research study. First, how faithfully would the second- and third-grade teachers adhere to the structural components of the prescribed intervention program (i.e., fidelity of implementation)? Second, to what extent would the teachers employ the strategies that were included in their professional development sessions (i.e., fidelity of process)? Third, would the teams function as truly collaborative teacher teams rather than as independent teachers in the same classroom? Fourth, what components of the intervention program would be significantly related to student achievement outcomes, and would process measures prove to be more important than implementation measures?

Methods

Participants

School district characteristics. The school district investigated in this study serves approximately 41,000 students who reside in 13 suburban communities located in a Midwestern state. The district has two kindergarten centers (early childhood), 28 elementary schools (grades K-5), seven middle schools (grades 6-8), and five high schools (grades 9-12). The overall student population is 82% White, 8% Black, 6% Asian, 3% Hispanic, and 1% Native American. Twenty-three percent of students receive free or reduced-priced meals, 7% are described as English Language Learners (ELL), and 12% receive special education services.

Pilot site characteristics. The three pilot program sites were selected by the district because they enrolled a higher percentage of students living in poverty than the district as a whole, were failing to make adequate yearly academic progress, and did not have an intervention program already in place. The pilot program was implemented in the second- and third-grade classrooms at these schools.

Pilot site 1 served 507 students in grades K-5 and had four second-grade and three third-grade classrooms. The student population was 37% White, 37% Black, 18% Asian, 7% Hispanic, and 2% Native American. Over half the students (54%) were eligible for free and reduced-priced meals. Twenty-four percent were ELL and 13% received special education.

Pilot site 2 served 702 students in grades K-5 and had five second-grade and five third-grade classrooms. The student population was 81% White, 8% Asian, 7% Black, 3% Hispanic, and 1% Native American. Thirty percent of students were eligible for free or reduced-price meals, 12% were described as ELL, and 11% received special education services.

Pilot site 3 served 397 students in grades K-5 and had three second-grade and three third-grade classrooms. The student population was 73% White, 11% Black, 8% Asian, 7% Hispanic, and 1% Native American. Forty-three percent of students were eligible for free and reduced-priced meals, 18% were described as ELL, and 12% received special education services.

Teachers. Twenty-three classroom teachers served as participants in this study. Twelve were second-grade teachers and 11 were third-grade teachers. An additional 18 team teachers also participated. All 18 were certified teachers. Eleven of the team teachers were assigned to either a second-grade or a third-grade class, and seven had assignments at both grade levels. The teachers' years of experience ranged from 1 to 37 years (M = 10.86 years) with teachers teaching at their current school from between 1 and 30 years (M = 6.15 years). On average, classroom

teachers had more years of teaching experience (*M* = 13.5 years, range 2 - 30 years) than team teachers (*M* = 7.92 years, range 1 - 37 years).

Observation Schedule

The data used to measure fidelity were gathered through classroom observations in the 12 second-grade and 11 third-grade pilot classrooms. The purpose of the observations was to determine the presence or absence of the required pilot program components and the degree to which processes and strategies taught to teachers during professional development were being used in the classroom. In addition, the observations documented the extent to which teacher team members worked together in a collaborative manner.

All pilot classrooms were observed two times during the school year, in October 2006 and May 2007, by one of three trained observers. The teachers were told that they would be observed during the semester, but they were not given the specific date. Each observation spanned the entire 2½-hour block that included reading instruction, mathematics instruction, and intervention time. An additional observation was made in four pilot classrooms in January 2007 for the purpose of obtaining data to compute inter-rater agreement (median *alpha* = .85, range .73 – 1.0). Observer retraining sessions were completed after each round of observations in October 2006, January 2007, and May 2007.

Observation Instruments

The observation instrument (See Appendix A) was developed by the researchers in collaboration with district administrative and curriculum staff. It included eight components of the pilot program: classroom procedures and management, teaming, support staff, guided reading/center time block, guided reading strategies and skills, center work/independent time, *Investigations Math*, and intervention time. Observers also noted adherence to the time allocation

requirements of the program. For example, the instructional block was to include 1 hour of *Balanced Literacy*, 1 hour of *Investigations Math*, and 30 minutes of intervention time. Observers also noted teacher behavior (e.g., *Students are given clear objectives for their learning*), and student behavior (e.g., *Every student is actively engaged in appropriate center work the entire time they are not in guided reading*).

Teacher Surveys

Paper-pencil surveys were administered in December 2006 and May 2007 to classroom and team teachers. On the December survey, the teachers were asked to provide descriptive information about their teaching assignment and years of experience. In addition, the December 2006 survey presented 39 questions regarding attitudes about the pilot program, professional learning communities, centers and interventions, training and professional development, instructional coaches, and academic intervention specialists. Responses were made using a 6-point Likert-type scale ranging from *strongly disagree* (1) to *strongly agree* (6).

The May 2007 teacher survey consisted of 83 items that measured attitudes on the same topics as the December 2006 survey but with the addition of items regarding students' independent work and the teachers' perceived importance of each of the components of the pilot program. Importance was rated using a 4-point scale ranging from *not important at all* (1) to *very important* (4). The survey response rate for both December 2006 and May 2007 was 95% with 39 of 41 teachers responding to both surveys.

Student Achievement Measures

Student achievement was measured using the *Measures of Academic Progress* (MAP) reading and mathematics tests for second- and third-grade students. The MAP test is a computerized adaptive test that was administered to students in September 2006 and in April and

May 2007. A gain score for both reading and mathematics was computed for each student by subtracting the fall MAP score from the spring MAP score.

Third graders' achievement was also measured with the *Minnesota Comprehensive Assessment-Series II* (MCA-II) reading and mathematics tests administered in May 2007. The MCA-II are paper-pencil tests developed by the state to measure students' progress towards academic standards. These tests are also used to determine federal adequate yearly progress (AYP) status. MCA-II results are reported in terms of four proficiency levels: (1) Does not meet proficiency, (2) Partially meets proficiency, (3) Meets proficiency, and (4) Exceeds proficiency. For analysis purposes, percent proficient included students who scored at or above proficiency (level 3 or level 4).

Results

Student pairing procedure. Students at pilot and comparison schools were eligible for inclusion in the analysis if they met the following criteria: (a) attended a minimum of 155 out of the possible 172 school days, and, (b) had fall (pretest) scores on the *Measures of Academic Progress (MAP*) reading and mathematics tests. For the analysis, each second- and third-grade student enrolled in a pilot classroom was matched with another student enrolled in the same grade but at another elementary school in the district. Matching variables were ethnicity (American Indian, Asian, Hispanic, Black, or White), eligibility for free or reduced price lunches (eligible or non-eligible), ELL status (ELL or non-ELL), and disability status (disabled or not disabled). Students in the subgroups defined by these four variables were ordered according to the sum of their fall MAP reading and mathematics scores.

The control pair mate selected for each pilot student was a non-pilot student who matched on all four demographic variables and had an identical MAP reading and mathematics sum. For

cases where the sum was not identical, the pairing was made with a non-pilot student who had the same MAP reading score. When exact matches still could not be made, the pairings were made with adjacent non-pilot students in an alternating above and below order. In the case of ties, a pair mate was selected randomly. For some combinations of demographic variables, the subgroups were extremely small, containing only one or two pilot students and no potential control students (e.g., Hispanic, free/reduced lunch eligible, non-ELL, special education student). In these situations, pairings were made with students in different ethnicity and/or lunch eligibility categories. This happened in only eight cases, affecting the pairing of three second-grade and five third-grade pilot students.

The second-grade sample included 484 students (mean age 7 years 6 months, range 6 years 10 months to 8 years 6 months). The second-grade student sample was 68% White, 16% Black, 12% Asian, 4% Hispanic, and less than 1% Native American. Two hundred fifty-eight second graders were boys and 226 were girls. Thirty-nine percent received free or reduced-priced meals, 18% were described as English Language Learners (ELL), and 11% received special education services.

The third-grade sample included 502 students (mean age 8 years 6 months, range 7 years 7 months to 9 years 10 months). The third-grade student sample was 66% White, 20% Black, 8% Asian, 5% Hispanic, and 1% Native American. Two hundred fifty-seven third graders were boys and 245 were girls. Thirty-nine percent received free or reduced-priced meals, 14% were described as ELL, and 14% received special education services.

Preliminary analyses of mathematics and reading test scores. Hierarchical linear modeling (HLM) analyses were carried out on the MAP mathematics and reading gain scores of second and third graders with free/reduced lunch eligibility, ELL status, disability status,

ethnicity, and school-level free/reduced lunch eligibility as covariates. The instructional program (pilot versus control) did not have a statistically significant effect in either second-grade or third-grade classrooms. HLM analyses were also carried out on the third graders' MCA-II mathematics and reading scores with fall MAP, free/reduced lunch eligibility, ELL status, disability status, ethnicity, and school-level free/reduced lunch eligibility as covariates. Again, the instructional program did not have a statistically significant effect.

Fidelity scoring. (See Appendix B for scoring rubrics.) Observation data were used to compute three fidelity of implementation (FOI) and four fidelity of process (FOP) scores for each classroom. The three FOI measures reflected the structural requirements of the pilot program apart from any quality assessment. The four FOP measures reflected the quality of the classroom instruction provided by both the classroom teacher and the team teacher. The classroom and team teachers were rated separately on FOP measures, and the average was used in the analysis.

FOI measures. FOI-Reading included the time allocated for reading instruction and literacy-focused center activities (0 to 3 points), presence of required reading groups (0 to 3 points), and proportion of students engaged in center time activities (0 to 3 points). *FOI-Math* included the time allocated for *Investigations Math* (0 to 3 points) and the presence of the three required components (introduction, exploration activities, and summary) (0 to 3 points). *FOI-Intervention* included the time allocated for intervention as well as the presence of both a large and a small group (0 to 3 points).

FOP measures. *FOP-Reading* assessed features of reading instruction and center time activities. For reading instruction, points were awarded for clarity of expectations, use of comprehension strategies, teacher-student interaction, and engagement level of students (0 to 4

points). For center time activities, points were awarded for clarity of instructions, smoothness of transitions, organization of materials, and adequate variety of literacy-based learning materials (0 to 4 points). *FOP-Math* included assessment of clarity of instructions, organization of materials, use of exploration strategies, and use of strategies to give students opportunities to share reflections and summaries (0 to 4 points). *FOP-Intervention* included assessment of clarity of instructions, use of large and small groups, teacher monitoring, and engagement level of students (0 to 4 points). The score for *FOP-Team* reflected how well the teachers were collaborating rather than working independently with respect to evidence of mutual planning, use of the same curriculum, goal-based interactions, and mutual monitoring and adjustment of intervention activities (0 to 4 points).

Bivariate correlations among the fidelity measures. The correlations between fidelity measures for second- and third-grade classrooms are displayed in Table 1. The relationships between FOI and FOP measures for mathematics, reading, and intervention were fairly strong in both second- and third-grade classrooms. However, only the mathematics fidelity measures (i.e., FOI-Math and FOP-Math) were significantly correlated at both grade levels. In addition, reading fidelity measures were significantly correlated with intervention and teaming at both grade levels. More specifically, in second-grade classrooms, FOP-Reading was significantly correlated with FOI-Intervention, FOP-Intervention, and FOP-Team, while FOI-Reading was significantly correlated with FOP-Intervention. In third-grade classrooms, FOI-Reading was significantly correlated with FOI-Intervention, FOP-Intervention, and FOP-Team, whereas FOP-Reading was significantly correlated with FOP-Intervention.

Table 1
Correlations Between Fidelity Measures in Pilot Classrooms

	FOP Math	FOI Reading	FOP Reading	FOI Interv.	FOP Interv.	FOP Team
Second-Grade Classrooms (n = 12)						
FOI-Math	.906**	.017	-.039	.447	.389	-.050
FOP-Math	—	-.086	.081	.369	.376	.040
FOI-Reading		—	.678*	.833**	.736**	.186
FOP-Reading			—	.601*	.701*	.728**
FOI-Intervention				—	.944**	.157
FOP-Intervention					—	.318
FOP-Team						—
Third-Grade Classrooms (n = 11)						
FOI-Math	.882**	.097	.398	-.009	.434	.529
FOP-Math	—	.273	.497	.122	.551	.599
FOI-Reading		—	.481	.645*	.678*	.651*
FOP-Reading			—	.533	.780**	.379
FOI-Intervention				—	.610*	.211
FOP-Intervention					—	.490
FOP-Team						—

* Correlation is significant at the .05 level (two-tailed).
** Correlation is significant at the .01 level (two-tailed).

Descriptive statistics for fidelity measures. In addition to the separate implementation and process measures, we summed the implementation and process measures for mathematics, reading, and intervention to provide a Total Math score, a Total Reading score, and a Total Intervention score. Total Math, Total Reading, and Total Intervention were then added together to provide a Composite score.

Descriptive information about the fidelity scores for second-grade and third-grade classrooms is displayed in Table 2. Inspection of these summary measures indicated that there was quite a bit of variability in both second- and third-grade pilot classrooms with respect to adherence to the pilot program requirements. This variation was present in both implementation and process measures. In addition, although we did not test variances for statistically significant differences, the standard deviations for second-grade classrooms were larger than those for third-grade classrooms across all fidelity measures with the exception of FOP-Team. Therefore, both

second- and third-grade teachers varied with respect to adherence to program requirements, and third-grade teachers appeared to be more similar to each other than second-grade teachers.

Selection of high and low fidelity classrooms. High and low fidelity pilot classrooms were identified for each of the eleven fidelity measures shown in Table 2. Within each grade level, the fidelity scores were rank ordered and, whenever possible, classrooms in the highest one-third of the distribution were labeled high fidelity, and classrooms in the lowest one-third were labeled low fidelity. Due to numerous tie scores in the distributions for some of the fidelity measures, however, the number in the high and low fidelity groups was sometimes somewhat less than or greater than one-third.

Analysis. Mean differences between pilot and control classrooms on the MAP reading and mathematics assessments were tested for statistical significance via a Bonferroni planned comparison procedure with the type I error probability set at .05 for each comparison. The same analysis was carried out on the third graders' MCA-II proficiency attainment.

Second-grade MAP mathematics. The MAP mathematics mean gain scores for second-grade classrooms are displayed in Table 3. Fidelity groupings based on implementation resulted in two statistically significant comparisons: Pilot students in high FOI-Reading classrooms outperformed pilot students in low FOI-Reading classrooms, and control students outperformed pilot students in low FOI-Intervention classrooms. Fidelity groupings based on process also resulted in two statistically significant comparisons: Control students scored higher than pilot students in low FOP-Intervention classrooms, and pilot students in the high FOP-Team classrooms scored higher than pilot students in low FOP-Team classrooms. Finally, comparisons based on fidelity sums resulted in three statistically significant differences. When fidelity groupings were based on Reading Total, pilot students in the high fidelity classrooms

significantly outperformed pilot students in the low fidelity classrooms. When fidelity groupings were based on Intervention Total, control students significantly outperformed students in the low fidelity pilot classroom. And when groupings were based on the Composite, pilot students in the high fidelity classrooms significantly outperformed pilot students in the low fidelity classrooms.

Second-grade MAP reading. The MAP reading means for second-grade classrooms are displayed in Table 4. Three comparisons carried out on MAP-reading scores were statistically significant. For two of these, when the fidelity classification was based on FOI-Reading or on the Composite, high fidelity pilot students significantly outperformed controls. For the third, when the fidelity classification was based on FOP-Math, low-fidelity pilot students scored significantly higher than controls.

Table 2.
Descriptive Statistics for Fidelity Measures in Pilot Classrooms

	Mean	SD	Median	Minimum	Maximum
Second-Grade Classrooms (n = 12)					
FOI-Math	10.08	2.15	10.5	6	12
FOP-Math	5.48	2.01	5.5	2	8
Math Total	15.56	4.06	16	8	20
FOI-Reading	16.17	2.08	17	12	18
FOP-Reading	11.92	3.20	11.5	6	16
Reading Total	28.08	4.86	28.75	18	34
FOI-Intervention	4.67	1.92	5.5	0	6
FOP-Intervention	5.33	2.57	5.5	0	8
Intervention Total	10.00	4.43	11	0	14
Composite	53.65	10.38	54.5	32.5	66
FOP-Team	5.50	2.11	6	2	8
Third-Grade Classrooms (n = 11)					
FOI-Math	10.32	1.23	10	9	12
FOP-Math	5.82	1.90	6.5	3	8
Math Total	16.14	3.04	16.5	12	19.5
FOI-Reading	16.55	1.76	18	14	18
FOP-Reading	12.18	2.42	11.5	8	16
Reading Total	28.73	3.61	28.5	22	34
FOI-Intervention	5.45	0.82	6	4	6
FOP-Intervention	5.64	1.36	5	4	8
Intervention Total	11.09	1.97	11	8	14
Composite	55.95	7.24	55	46	67.5
FOP-Team	5.18	2.36	6	0	8

Table 3

Comparisons of Second-Grade MAP Mathematics Gain Scores among Low Fidelity Pilot, High Fidelity Pilot, and Control Classrooms

	Low Fidelity Pilot			High Fidelity Pilot			Control		
	N[a]	Mean	SD	N[a]	Mean	SD	N	Mean	SD
FOI-Math	76 (4)	15.07	7.82	106 (5)	14.42	7.83	242	15.15	7.50
FOP-Math	56 (3)	15.64	8.15	86 (4)	14.45	7.90	242	15.15	7.50
Math Total	76 (4)	15.06	7.82	106 (5)	14.41	7.83	242	15.15	7.50
FOI-Reading	59 (3)	12.81	7.33	85 (4)	16.88+	8.29	242	15.15	7.50
FOP-Reading	56 (3)	15.64	8.15	86 (4)	14.45	7.94	242	15.15	7.50
Reading Total	59 (3)	12.81	7.33	84 (4)	15.39+	8.33	242	15.15	7.50
FOI-Intervention	77 (4)	13.00	7.34	124 (6)	14.61	7.88	242	15.15#	7.50
FOP-Intervention	77 (4)	13.00	7.34	80 (4)	13.70	7.85	242	15.15#	7.50
Intervention Total	77 (4)	13.00	7.34	80 (4)	13.70	7.85	242	15.15#	7.50
Composite	56 (3)	12.59	7.84	62 (3)	15.69+	8.68	242	15.15	7.50
FOP-Team	82 (4)	12.80	6.89	102 (5)	15.13+	8.31	242	15.15	7.50

[a]The number in parentheses is the number of classrooms.
+Statistically significant difference between high pilot and low pilot
Statistically significant difference between low pilot and control

Table 4

Comparisons of Second-Grade MAP Reading Gain Scores among Low Fidelity Pilot, High Fidelity Pilot, and Control Classrooms

	Low Fidelity Pilot			High Fidelity Pilot			Control		
	N[a]	Mean	SD	N[a]	Mean	SD	N	Mean	SD
FOI-Math	76 (4)	17.53	9.89	106 (5)	17.00	9.56	242	15.31	9.60
FOP-Math	56 (3)	19.84#	9.86	86 (4)	17.02	0.30	242	15.31	9.60
Math Total	76 (4)	17.53	9.89	106 (5)	17.00	9.56	242	15.31	9.60
FOI-Reading	59 (3)	15.24	10.76	85 (4)	18.55*	10.52	242	15.31	9.60
FOP-Reading	56 (3)	15.31	991	86 (4)	16.62	9.34	242	15.31	9.60
Reading Total	59 (3)	15.24	10.76	84 (4)	16.62	9.34	242	15.31	9.60
FOI-Intervention	77 (4)	15.81	10.17	124 (6)	16.06	9.35	242	15.31	9.60
FOP-Intervention	77 (4)	15.81	10.17	80 (4)	15.34	9.15	242	15.31	9.60
Intervention Total	77 (4)	15.81	10.17	80 (4)	15.34	9.15	242	15.31	9.60
Composite	56 (3)	15.55	10.54	62 (3)	18.42*	9.40	242	15.31	9.60
FOP-Team	82 (4)	14.12	9.81	102 (5)	16.77	9.63	242	15.31	9.60

[a]The number in parentheses is the number of classrooms.
*Statistically significant difference between high pilot and control
Statistically significant difference between low pilot and control

Third grade MAP mathematics and reading. The MAP mathematics means for third-grade classrooms are displayed in Table 5, and the MAP reading means are displayed in Table 6.

None of the pairwise comparisons for either MAP mathematics or reading test performance of third graders were found to be statistically significant.

Table 5
Comparisons of Third-Grade MAP Mathematics Gain Scores among Low Fidelity Pilot, High Fidelity Pilot, and Control Classrooms

	Low Fidelity Pilot			High Fidelity Pilot			Control		
	N[a]	Mean	SD	N[a]	Mean	SD	N	Mean	SD
FOI-Math	96 (4)	13.04	6.43	60 (3)	12.53	6.28	251	13.17	7.73
FOP-Math	81 (3)	13.59	6.73	63 (3)	13.34	7.54	251	13.17	7.73
Math Total	96 (4)	13.04	6.43	84 (4)	12.63	7.23	251	13.17	7.73
FOI-Reading	47 (2)	10.79	6.85	137 (5)	12.78	6.84	251	13.17	7.73
FOP-Reading	76 (3)	12.92	7.37	84 (4)	12.83	7.43	251	13.17	7.73
Reading Total	90 (4)	14.01	6.66	60 (4)	12.82	6.60	251	13.17	7.73
FOI-Intervention	47 (2)	13.15	6.50	155 (6)	13.54	6.87	251	13.17	7.73
FOP-Intervention	47 (2)	13.15	6.50	34 (2)	13.44	6.07	251	13.17	7.73
Intervention Total	72 (3)	12.69	6.61	86 (4)	13.29	6.74	251	13.17	7.73
Grand Total	95 (4)	13.18	6.69	84 (4)	12.83	7.43	251	13.17	7.73
FOP-Team	64 (3)	15.09	6.48	60 (3)	12.82	6.60	251	13.17	7.73

[a]The number in parentheses is the number of classrooms.

Table 6
Comparisons of Third-Grade MAP Reading Gain Scores among Low Fidelity Pilot, High Fidelity Pilot, and Control Classrooms

	Low Fidelity Pilot			High Fidelity Pilot			Control		
	N[a]	Mean	SD	N[a]	Mean	SD	N	Mean	SD
FOI-Math	96 (4)	11.02	8.13	60 (3)	9.90	7.77	251	10.36	8.35
FOP-Math	81 (3)	11.60	7.74	63 (3)	9.63	7.93	251	10.36	8.35
Math Total	96 (4)	11.02	8.13	84 (4)	9.90	7.79	251	10.36	8.35
FOI-Reading	47 (2)	10.79	6.85	137 (5)	10.85	8.57	251	10.36	8.35
FOP-Reading	76 (3)	11.32	8.17	84 (4)	10.58	8.38	251	10.36	8.35
Reading Total	90 (4)	11.31	7.57	60 (4)	10.85	8.59	251	10.36	8.35
FOI-Intervention	47 (2)	10.79	6.85	155 (6)	11.08	8.53	251	10.36	8.35
FOP-Intervention	47 (2)	10.79	6.85	34 (2)	9.62	8.04	251	10.36	8.35
Intervention Total	72 (3)	10.94	7.41	86 (4)	10.26	8.34	251	10.36	8.35
Composite	95 (4)	11.39	7.75	84 (4)	10.58	8.38	251	10.36	8.35
FOP-Team	64 (3)	11.73	7.59	60 (3)	10.85	8.59	251	10.36	8.35

[a]The number in parentheses is the number of classrooms.

Third-grade MCA-II mathematics proficiency. The MCA-II mathematics proficiency means for third-grade classrooms are displayed in Table 7. Comparisons based on FOI groupings resulted in one statistically significant comparison in which controls outperformed pilot students

in low FOI-Math classrooms. Comparisons based on FOP groupings resulted in three statistically significant comparisons: Pilots in low FOP-Math classrooms performed less well than controls, and pilots in low FOP-Team classrooms performed less well than either controls or pilots in high FOP-Team classrooms. Comparisons based on the summed measures produced three statistically significant results: Controls outperformed pilot students in classrooms categorized as low fidelity with respect to the Math Total, Reading Total, and Composite.

Table 7
Comparisons of Third-Grade MCA-II Mathematics Percent Proficient among Low Fidelity Pilot, High Fidelity Pilot, and Control Classrooms

	Low Fidelity Pilot		High Fidelity Pilot		Control	
	N[a]	%	N[a]	%	N	%
FOI-Math	96 (4)	66.7	60 (3)	75.0	251	78.7#
FOP-Math	81 (3)	60.5	63 (3)	64.1	251	78.7#
Math Total	96 (4)	66.7	84 (4)	70.2	251	78.7#
FOI-Reading	47 (2)	68.1	137 (5)	77.4	251	78.7
FOP-Reading	76 (3)	73.7	84 (4)	72.6	251	78.7
Reading Total	90 (4)	65.6	60 (4)	78.3	251	78.7#
FOI-Intervention	47 (2)	68.1	155 (6)	74.8	251	78.7
FOP-Intervention	47 (2)	68.1	34 (2)	76.5	251	78.7
Intervention Total	72 (3)	68.1	86 (4)	79.1	251	78.7
Composite	95 (4)	64.2	84 (4)	74.6	251	78.7#
FOP-Team	64 (3)	62.5	60 (3)	78.3+	251	78.7#

[a]The number in parentheses is the number of classrooms.
+Statistically significant difference between high pilot and low pilot
Statistically significant difference between low pilot and control

Third-grade MCA-II reading proficiency. The MCA-II reading proficiency means for third-grade classrooms are displayed in Table 8. Two comparisons based on FOP groupings were significant: Pilot students in low FOP-Reading classrooms outperformed pilot students in high FOP-Reading classrooms, and control students outperformed pilot students in low FOP-Team classrooms. None of the comparisons based on FOI groupings or fidelity sum scores were found to be statistically significantly.

Table 8
Comparisons of Third-Grade MCA-II Reading Percent Proficient among Low Fidelity Pilot, High Fidelity Pilot, and Control Classrooms

	Low Fidelity Pilot		High Fidelity Pilot		Control	
	N[a]	%	N[a]	%	N	%
FOI-Math	96 (4)	72.9	60 (3)	75.0	251	75.6
FOP-Math	69 (3)	65.4	57 (3)	63.5	251	75.6
Math Total	96 (4)	72.9	84 (4)	71.1	251	75.6
FOI-Reading	47 (2)	70.2	137 (5)	78.8	251	75.6
FOP-Reading	76 (3)	80.3[+]	84 (4)	66.7	251	75.6
Reading Total	90 (4)	66.7	60 (4)	70.0	251	75.6
FOI-Intervention	47 (2)	70.2	155 (6)	75.5	251	75.6
FOP-Intervention	47 (2)	70.2	34 (2)	70.6	251	75.6
Intervention Total	72 (3)	70.8	86 (4)	83.7	251	75.6
Composite	95 (4)	68.4	84 (4)	67.5	251	75.6
FOP-Team	64 (3)	60.9	60 (3)	70.0	251	75.6[#]

[a]The number in parentheses is the number of classrooms.
+Statistically significant difference between high pilot and low pilot
Statistically significant difference between low pilot and control

Summary of fidelity group comparisons. In all seven statistically significant comparisons of MAP mathematics achievement means for second-grade fidelity groups, the students in low fidelity pilot classrooms were outperformed by either controls or high fidelity pilots. For two of the statistically significant comparisons of second-grade MAP reading, high fidelity pilots had a performance advantage over controls, and for one statistically significant comparison, low fidelity pilots had a performance advantage over controls. None of the comparisons conducted on MAP mathematics or MAP reading scores of third graders were statistically significant. A total of seven comparisons on the third-grade MCA-II mathematics proficiency means were found to be statistically significant, and in all seven comparisons, pilot students in low fidelity classrooms were outperformed by either controls or by high fidelity pilots. For one statistically significant comparison of third graders' MCA-II reading proficiency means, low fidelity pilots had a performance advantage over controls, and for the other, controls had a performance advantage over high fidelity pilots.

Teacher survey. Teachers were asked to complete paper-pencil survey questionnaires in the fall and in the spring. The only items that are presented here appeared in the spring survey. For these items, the teachers were asked to rate the importance of nine components of the compensatory pilot program (see Table 9). The teachers' responses indicated that they believed that the most important component of the pilot program was team teaching, followed by training in guided reading, intervention, and Advantage Math. They felt that the least important program component was the prescribed instructional time schedule.

Table 9
Teacher Survey Responses to "How important are the following components of the Compensatory Pilot Program to its success?"

	Not At All Important	Slightly Important	Important	Very Important
Team teaching	0% (n = 0)	5.3% (n = 12)	7.9% (n = 3)	86.8% (n = 33)
Guided reading training	2.8% (n = 1)	11.1% (n = 4)	27.8% (n = 10)	58.3% (n = 21)
Intervention training	8.1% (n = 3)	10.8% (n = 4)	29.7% (n = 11)	51.4% (n = 19)
Advantage Math training	5.6% (n = 2)	13.9% (n = 5)	30.6% (n = 11)	50.0% (n = 18)
CGI training	2.9% (n = 1)	29.4% (n = 10)	32.4% (n = 11)	35.3% (n = 12)
Center time training	11.4% (n = 4)	28.6% (n = 10)	28.6% (n = 10)	31.4% (n = 11)
Instructional coaches	16.2% (n = 6)	32.4% (n = 12)	40.5% (n = 15)	10.8% (n = 4)
Academic specialists	18.4% (n = 7)	36.8% (n =14)	36.8% (n = 14)	7.9% (n = 3)
Prescribed instructional schedule	27.8% (n = 10)	41.7% (n =15)	25% (n = 9)	5.6% (n = 2)

Discussion

Adherence to the Structural Components

The FOI measures that we included in the evaluation enabled us to assess the teachers' adherence to the structural components of the compensatory program. We discovered, not

unexpectedly, that implementation fidelity varied in both reading and mathematics and in second-grade as well as third-grade classrooms. However, it appeared that third-grade teachers were less variable and exhibited a higher degree of adherence than second-grade teachers. An important implication of findings regarding implementation fidelity is that complete fidelity cannot be assumed and some resistance on the part of the teachers should be expected (e.g., Datnow & Castellano, 2000). Consequently, results of fidelity investigations can be used not only to document degree of fidelity, but also to identify areas where implementation is somewhat problematic and where additional training and support may be needed (e.g., Noell, Witt, LaFleur, Mortenson, Ranier, & LeVelle, 2000). In the project described in the present study, for example, information obtained from observations of mathematics instruction revealed that teachers frequently omitted the summarizing component of *Investigations Math* where students were supposed to share their reasoning and problem-solving strategies with each other. Therefore, the district offered additional professional development specifically related to the implementation and importance of student summarizing and reflection.

Fidelity of Process

Reading and mathematics instruction. Similar to results regarding FOI measures, teachers varied with respect to process fidelity in both reading and mathematics instruction and in both second- and third-grade classrooms. The process measures for reading and mathematics primarily reflected clarity, organization, and appropriate use of strategies, and variation on these dimensions from one classroom to the next would be expected. When a number of observations taken at different times in the same classroom all result in very low process ratings, however, the data signal the need for consultation on the part of an academic coach.

The district described in the present study did employ classroom coaches who were responsible for observing classroom instruction in the pilot schools and providing feedback to the teachers. The pilot teachers, however, perceived coaching to be only moderately important for the success of the program. Russo (2004) has suggested that classroom coaches can be very effective because, unlike conventional conferences and professional development days, the exchange is one-on-one and is related to a specific teacher's classroom work. On the other hand, Neufeld and Roper (2003) suggest that it may take a long time for teachers to change their practices, and coaches may observe no evidence that teachers have adopted their suggestions. Perhaps the impact of coaching could be enhanced by ensuring that both coaches and teachers understand how coaching fits into the "big picture" of the instructional program, and by establishing benchmarks that will encourage collegial discussion about the relationship between teaching practice and achievement outcomes (Neufeld & Roper, 2003). It might be valuable for school administrators and coaches to become familiar with the observation protocols utilized by the program evaluators. As an aside, principals of the pilot schools were given training in use of a modified version of our observation protocols and rubrics. The principals who have started using these instruments claim that they are very valuable and have markedly changed how they observe classroom instruction.

Teaming. Teaming was the biggest change implemented in the pilot program and the most costly. The district hoped that teaming would ultimately provide the biggest return in terms of student achievement. Our analysis showed that the full range of ratings was attained on the fidelity of collaborative teaming measure. While the top rating indicated that the teachers were clearly working as a collaborative team, ratings in the middle of the scale indicated that different variations of teaming were being utilized. For example, in one variation, both teachers provided

instruction using the same plan and the same materials but without any interaction indicating mutual adjustment based on students' performance. Or in another variation with somewhat less collaboration, the teachers provided instruction on the same topic but followed different plans and used different materials.

Cook and Friend (1995) state that the degree of collaboration in teacher teams is often based on trust and comfort with one another, and the lowest rating in our scale, which was given only a couple of times, indicated that the teamers exhibited lack of trust and antagonism. Prior to the implementation of teams, however, the pilot teachers received no training or orientation specifically focused on team teaching, and Cook and Friend have indicated that successful teaming requires communication skill development, knowledge about collaborative planning, as well as assessment of one's readiness for collaboration. Although, based on our analysis, we would have recommended that the district provide professional development on these topics, district administrators have already decided to discontinue teaming because it was very costly to place two licensed teachers in a classroom and because the pilot students who experienced a high degree of collaborative teaming achieved no better than students in control classrooms where no teaming occurred. Interestingly, results of the teacher survey indicated that the vast majority of the pilot teachers thought that teams were very important for the success of the compensatory program.

Relationship between fidelity and student achievement

If the compensatory education program were especially effective, we might expect to see overwhelming evidence that students in the high fidelity pilot classrooms were achieving at a significantly higher level than students in control classrooms. This result, however, was only obtained for two statistical comparisons on second-grade reading achievement. In contrast to

expectations, the overall pattern in the results indicated that pilot students in low fidelity classrooms were especially at a disadvantage with respect to mathematics achievement, both compared to high fidelity pilots and to control students. Through our conversations with district personnel, we learned that, with the exception of the prescribed time structure and team teaching, many control schools were implementing instructional programs that had components very similar to the components of the pilot schools' compensatory program. Observations that we recently carried out in a sample of the control classrooms confirmed that this was indeed the case. Therefore, we should probably not be disappointed that students in high fidelity pilot classrooms and control students performed at the same level, especially because all three pilot schools made AYP in the second year of the project whereas this was not the case in year one. Nevertheless, we should be concerned about instruction in low fidelity pilot classrooms and should attempt to identify the specific reasons for low fidelity and then prescribe appropriate responses in the form of coaching, professional development, or program modifications.

We note that performance differences related to fidelity were observed for third graders on the MCA-II tests but not the MAP tests. We believe this was likely due to the differences in test content in that the MCA-II test has a fairly strong emphasis on assessment at the application, analysis, synthesis, and evaluation levels of Bloom's Taxonomy (Bloom & Krathwohl, 1956; Minnesota Department of Education, 2006). Frequently, teachers of low math fidelity classrooms omitted the summary portion of *Investigations Math* that requires students to use higher order thinking skills and articulate mathematical reasoning, which may have translated to poorer performance on the MCA-II mathematics test.

An interesting finding in second-grade classrooms was that fidelity of implementing the reading program components was significantly related to mathematics as well as reading

achievement. We conjecture that, because the *Investigations Math* curriculum placed a heavy emphasis on story problems for application of problem solving and reasoning skills, appropriate exposure to the reading components gave the second-grade pilot students a performance advantage in mathematics. This, of course, is a conjecture that would need further investigation.

Recommendations for Future Fidelity Observations

We had expected that measures of process fidelity (e.g., Did the teachers rigorously use the strategies that were taught to them?) would be more strongly related to achievement outcomes than measures of implementation fidelity (Did the teachers follow the time schedule?) However, the data did not reveal much variation between our implementation and process measures. The most likely explanation is that the measures, in essence, measured the same thing. In other words, if teachers implemented the structure of the pilot program, they were probably implementing the processes within the structure.

Conclusion

There were limitations to the study. All components of the project were chosen in advance by the client district. The school sites were not randomly selected. They were chosen by the district in advance of the evaluation based on state and district assessments and the absence of other intervention programs. Additionally, no single intervention strategy except the prescribed time structure and team teaching was unique to the pilot schools. Nevertheless, we found that using fidelity measures in concert with learning outcome measures in a matched-pairs, quasi-experimental design for evaluation of a school district's intervention allowed more in-depth analysis of the intervention's component parts and increased the confidence with which the project's major objectives could be addressed, thereby providing more useful information to school district personnel.

While there is no single best framework for assessing fidelity of implementation, and while fidelity studies are time and labor intensive, researchers should, whenever possible, include fidelity measures in evaluation studies in order to give clients the most thorough information possible so that they can draw sound conclusions regarding the effectiveness of programs for improving educational outcomes.

References

Reading: Planning guide for balanced literacy (Teacher's ed.). (2006). Boston, MA: Houghton Mifflin

Berman, P. & McLaughlin, M. W. (1976). Implementation of educational innovation. *The Educational Forum, 40,* 345–370.

Blackmore, J. (1998). The politics of gender and educational change: Managing gender or changing gender relations? In A. Hargreaves, et al. (Eds.), *International Handbook of Educational Change* (pp. 460-481). London: Kluwer Academic Publishers.

Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Handbook I, The cognitive domain.* New York: David McKay.

Bond, G.R., Becker, D.R., & Drake, R.E. (1997). A fidelity scale for the individual placement and support model of supported employment. *Rehabilitation Counseling Bulletin, 40,* 265-284.

Bond, G.R., Evans, L., Salyers, M.P., Williams, J., & Kim, H.W. (2000). Measurement of fidelity in psychiatric rehabilitation*, Mental Health Services Research, 2(2),* 75-87.

Chen, H. (1990). *Theory-driven evaluations.* Thousand Oaks, CA: Sage.

Cook, L., & Friend, M. (1995). Co-teaching: Guidelines for creating effective practices. *Focus on Exceptional Children, 28(3)*, 1-16.

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.

Datnow, A., & Castellano, M. (2000). Teachers' responses to Success for All: How beliefs, experiences, and adaptations shape implementation. *American Educational Research Journal, 37(3),* 775-799.

Dusenbury, L., Brannigan, R., Falco, M., and Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings, *Health Education Research*, *18(2),* 237-256.

Hohmann, A. A., & Shear, M. K. (2002). Community-based intervention research: Coping with the "noise" of real life in study design. *American Journal of Psychiatry, 159(2),* 201-207.

Institute of Education Sciences. (2007). *Request for Applications* (No. IES-NCER-2008-01). Retrieved online from http://ies.ed.gov.

Investigations in Number, Data, and Space. (2007). Cambridge, MA: Pearson Scott Foresman TERC.

Minnesota Department of Education (2006, August). Minnesota Comprehensive Assessments Series II (MCA-II): Test specifications for reading & mathematics. Retrieved February 28, 2008, from *http://education.state.mn.us/mdeprod/groups/Assessment/documents/Report/006367.pdf*

Noell, G. H., Witt, J. C., LaFleur, L. H., Mortenson, B. P., Ranier, D. D., & LeVelle, J. (2000). Increasing intervention implementation in general education following consultation: A comparison of two follow-up strategies. *Journal of Applied Behavior Analysis, 33(3),* 271-284.

Neufeld, B., & Roper, D. (2003). *Coaching: A strategy for developing instructional capacity: Promises and practicalities*. Washington, D.C.: Aspen Institute Program on Education and Annenberg Institute for School Reform.

O'Donnell, C. L., Lynch, S. J., Watson, W. A., & Rethinam, V. (2007, April). FOI as quality of delivery and student responsiveness and its relationship to classroom achievement. In C. L. O'Donnell (Chair), *Analyzing the relationship between fidelity of implementation (FOI) and student outcomes in a quasi-experiment.* Symposium conducted at the annual meeting of the American Educational Research Association, Chicago, IL.

Rogers, E. (2003). *Diffusion of innovations.* New York: Free Press.

Russo, A. (2004, July/August). School-based coaching: A revolution in professional development or just the latest fad? *Harvard Education Letter*. Retrieved February 27, 2008, from http://www.edletter.org/past/issues/2004-ja/coaching.shtml

Sánchez, V., Steckler, A., Nitirat, P., Hallfors, D., Cho, H., & Brodish, P. (2007). Fidelity of implementation in a treatment effectiveness trial of Reconnecting Youth. *Health Education Research, 22(1),* 95-107.

Sechrest, L., West, S.G., Phillips, M.A., Redner, R., & Yeaton, W. (1979). Some neglected problems in evaluation research; Strength and integrity of treatments. In L. Sechrest, S.G., West, M.A. Phillips, R. Redner, & W. Yeaton (Eds.), *Evaluation studies review annual* (Vol.4, pp. 15-35). Thousand Oaks, CA: Sage.

Appendix A. Observation Instrument

Pilot Program Study Observation Record – Spring 2007

Date_____ **School**_____

Classroom _____ **New Team** _____**yes** _____**no**

Classroom Procedures and Management	Notes	
	Lead teacher	Team teacher
• The classroom is set up to match the required time schedule of the Pilot i.e., one hour in Reading of which 20 minutes is spent with each student in Guided Reading and the balance of the time is spent in Independent work or with ELL or Title teachers; one hour Investigations Math; 30 minutes in Intervention time		
• There are clear classroom routines for transitioning from one activity to the next. Transitions are brief, orderly and intentional		
• Students are given clear objectives for their learning. Teachers convey the message that the work is important, sensible and doable. Teachers clearly convey to students the purpose and procedures of each lesson or block of time.		
• Students are given examples, demonstrations and opportunities to interact with material		
• Students are given clear instructions on what to do in each of the three pilot components. Students have the materials they need and little or no time is wasted in choosing materials		
• Both teachers monitor student activities and behavior		
• Desks are arranged in clustered groups to encourage and allow collaboration		
• Students listen to the teacher and classmates during large group, small group and transition times. There is clear communication between the teachers and students at all times		

Teaming	Notes	
	Lead teacher	Team teacher
• There is obvious collaboration between the two teachers		
• Both teachers conduct Guided Reading groups, 20 minutes in length		
• Both teachers are either working with small groups or conducting a larger group during Intervention Time, and Exploration Math		
Support Staff		
• Are there support staff in the room?		
• How many other people?		
• What do they do?		
• How do they work with students?		
Guided Reading/ Center Time Block	**Notes**	
	Lead teacher	**Team teacher**
• **Frequency/Duration** Every child meets with a teacher for GR for 20 minutes. If a student leaves during GR time, question when the child gets their GR time. If there is less time or an alternative to this structure, ask why		
• **ELL and Title I** Students who are in ELL or Title I reading groups also participate in classroom teacher led GR groups Is ELL push in or pull out? Title I push in or pull out? Special Education?		
• **Purpose/Focus** for each GR group is clear. Teachers reference previous learning, and actively reference specific learning strategies during the lesson. If GR lesson is prior to students' Center Time, teachers give clear instructions as to what the group is supposed to work on when they return to their desks. If GR group is after Center Time, there is some reference made to the work they had been doing in Center time. Teachers may even meet briefly with a GR group before Center Time to make sure students know what to do. Both teachers are seated with a view of the room and are responsive to what is happening in the room		

Guided Reading Strategies and Skills	Notes	
	Lead teacher	Team teacher
• One student reading aloud to teacher while others whisper read to themselves • RRR not used • Frame words with fingers • Find verbs • Re-reading • Predicting • Mapping • Retelling • Using the easel or small white boards • Identifying challenges within the text • Vocabulary • Readers Theatre • Phonics • Text features • Decoding • Fluency		
Comprehension Strategies to be used • Activating Background Knowledge • Visualizing • Determining Importance • Questioning • Inferring • Summarizing/Synthesizing • Monitoring comprehension		
Are the Comprehension Strategies posters at or near the GR tables? Do teachers use the vocabulary and language of the Comprehension Strategies? Do they refer to GR posters as they discuss the strategies?		
On-going Assessment • Running Record • Other		
Center Work/Independent Time	**Notes**	
	Lead teacher	**Team teacher**
• Every student is actively engaged in appropriate center work the entire time they are not in GR.		
• When students are not with a teacher for GR, they are engaged in one of the following ways: 1. Independent Reading of leveled materials 2. Writing related to Reading 3. Listening/Computer 4. Word Work		

Examples of possible work: • Continuous Independent Reading of leveled books, either fiction or non-fiction • Rereading material from GR group • Poetry reading • Keeping records of their reading • Partner reading • Readers Theatre practice • Reading Response journals: reflecting on their own reading • Research: fact finding related to non-fiction reading and response • Literacy listening books on computer • Focus on Vocabulary (Word Work) • Making works (Word Work)		
The following activities are not acceptable during Center/Independent time: Math, Handwriting practice, Computer games or non-literacy sites, keyboarding)		
• What instructions do teachers give students at the end of their GR groups to direct their independent time? Do they give instructions at the beginning to the GR session?		
• How are students choosing their reading materials? Do they go to big bins in groups? How are the books at the bins categorized so students know which levels to use? Does each student have a solid file folder with books already in it? Etc.		

Investigations Math	Notes	
	Lead teacher	Team teacher
1. Introduction: Begins with 10-15 minute introduction which should include the following: • Written (and stated) statement of lesson objectives • Connections to previous learning • Introduction of Exploration • Clarification of expectations		
2. Exploration: Class split in half or in small groups • Teachers introduce the lesson and activities clearly and in an organized fashion • Students actively engaged • Students participate in "math talk" with classmates and teacher • Both teachers move among groups and check for understanding, record observations that will be helpful for next lesson • Activities are modified when needed • Teachers/students model with manipulatives/visual aides • Students organized in variety of flexible groups • Teachers use open-ended questions in components of the lesson; ask students rather than tell. Teacher should have 1-2 questions that every child can answer by the end of Exploration • Students are comfortable sharing work and questions • Evidence that teachers have a grasp of the mathematics of the lesson		
3. Summary/Sharing: may be whole or split group. • Sharing encouraged by use of open ended questions, clarifying questions and paraphrasing of student thinking • Students pre-selected to share their work • Students use academic vocabulary to share ideas. They can summarize lesson objectives and new understandings • There is encouragement to accept multiple approaches to a problem		
Question: How often does S/S occur?		

Things that should **not** be seen during Math Exploration: • Worksheets of isolated fact and computation (does not include Investigations Student Activity Books or Practice Counts) • High proportion of single answer questions • Teachers doing the math • 1 or 2 students in a group doing the math on behalf of rest of group • Desks in rows facing front of room • Teachers imposing their own math concepts on students or explaining how the problem was solved based on how they would have done it rather than how the student chose to do it • "Anything goes" sharing of ideas or strategies • Teachers walking students through the tasks step by step		
Intervention Time	**Notes**	
	Lead teacher	**Team teacher**
• **30 minutes spent on intervention work** • Can be a continuation of Center time activities. Should not be a continuation of Math- that lesson should come to a close with all students • No Whole Group instruction • One teacher meets with individual or small group while second teacher meets with remaining students in the room (2nd Grade = lead teacher does Literacy, team teacher does Math) as a group or individually		
Possible Activities for large group • Literature Circles • Content area modeled or shared reading • Partner Reading • Readers Theatre • Independent Research Activities (very clearly defined) • Differentiated Math Games • Practice Counts • Big Books • Word Work • Handwriting (not every day) • Spelling (not every day) • Expansion on other subject areas, but not the daily lesson in that area • Continuation of Center activities		

Small Group Instructional Strategies • Add+Vantage Math • No re-teaching of lesson or Practice Counts • Lesson has a clear focus and simply stated instructional objective • Lesson has a displayed group poster outlining the objective; this is added to each day by group • Students encouraged to monitor their own progress • Teacher embeds formative assessment into each lesson and keeps records of student progress		
What does each teacher focus on and how did you figure it out?		

Additional Notes/Comments:

Appendix B. Scoring Rubric

Pilot Study **Fidelity of Implementation Rubric**		
Fidelity of Implementation Components	**FOI scale**	**Rubric**
1 hour of Investigations Math	0-3	3=Doing a full hour of Investigations Math 2= Less than 1 hour of Investigations Math 1= Doing math for an hour but not Investigations 0= Not doing math
Investigations *Math*: presence of the three components: Introduction, Exploration, Summary/Sharing	0-3	3=all three present 2=two present 1= one 0= none
1 hour of Guided Reading and literacy focused Center Time	0-3	3=Doing a full hour of GR and Center Time 2=Less than 1 hour of Reading Block 1=Doing reading for an hour but not Guided Reading/Center Time 0=Not doing GR/ Center Time
GR: Every child has 20 minutes in GR	0-3	3=Yes and students who have a group with ELL or Title or Sp Ed (either pull in or push out) also have a GR group with classroom teacher 2= Some students pulled out and do not have a 20 minute GR group with a classroom teacher when they return 1= Some groups are 20 minutes, others are not; 0= Students not engaged in GR groups
GR: Center Time: students are engaged in activities when not in GR	0-3	3=Nearly all students meaningfully engaged most of CT 2=Many students are engaged but a significant number are wandering, talking, off task during CT 1=Very few students engaged appropriately for CT 0= Students not engaged during CT
30 Minutes Scheduled Intervention time		3=Full 30 minutes, both large and small group 2=Less than 30 minutes but exists 1=Only 1 group 0=No discernable Intervention Time

Pilot Study Fidelity of Process Rubric		
Fidelity of Process Components	**FOP scale**	**Rubric**
Teaming: evidence of collaboration	0-4	4=Team clearly has planned together and prepared similar materials; monitor and adjust with each other during the lessons 3=Have planned but do not interact during lessons 2=Teachers co-existing in the room; same curriculum but teaching with different materials 1=Team planning not evident except for the general structure of the day; no interaction; different materials, different approaches 0=Same as 1 but with obvious antagonism and/or lack of trust
GR Groups: quality of groups, use of strategies, skills; clarity of instructions and purpose	0-4	4=Purpose of group is very clear; use of comprehension strategies is apparent in vocabulary and conversation with students;, instructions clear, good interaction between teacher and students; students are pushed to answer/do themselves and go beyond the obvious; skills are part of the group; this applies whether the group is working on Readers Theatre, preparation for presentations, new materials, etc. 3=The same as 4 but missing one component 2=Groups take place; there is little reference to the comprehension strategies; work on the reading material perfunctory or superficial with little reference to prior knowledge, skill building, etc 1=Group purpose not immediately clear;; no reference to Comp Strategies and not much, if any skill work; students not asked to come up with their own answers, lots of teacher talk, little student response. 0=Goals not at all clear; if goal is stated, not accomplished during the group; disruptions from students being taken out of the lesson/classroom; lack of focus/coherence

GR Center Time: clarity of transitions and instructions; organization of materials, engagement of students; types and variety of work	0-4	4=Very clear instructions as to what groups are to do in CT; students get materials quickly, know what they are to do and do it with very little prompting during CT ; all choices are strong, literacy based activities; all choices are explained clearly so that students can access materials and know what to do with little wasted time. 3=The same as 4 but with one component missing 2=CT takes place. Choices are spoken rather than written down; free reading materials are not necessarily leveled, and there may be considerable wandering about taking one book after another; there may be games instead of reading/writing related activities; students are not on task the whole time 1=Like 2 but even less structured, more inappropriate choices, little monitoring of what students are doing; little concentration on part of students. 0=CT is basically a free time for students; no accountability for the time; activities not literacy based; no instruction; no monitoring of students
Investigations Math: division of class in exploration; clarity of instructions in each component and good transitions; materials ready to go; use of manipulatives; modifications as needed; open ended questioning	0-4	4=Instruction is carefully thought out with materials ready to go; good introduction as to the purpose of the lesson, reference back to previous work; very clear explanation as to what is going to take place during Exploration; transition to Exploration made easily and class is divided into two groups or teams for Exploration; work in Exp goes forward as explained with teachers monitoring, adjusting, asking open ended questions; Summary and Sharing takes place sometime during the lesson. 3=The same as 4 but with one component missing 2= Math lesson takes place although it may not be in the three components of Investigations. It may be totally full class with no dividing the class into groups; Students participate but the purpose may not be very clear, nor the instructions or the transitions 1=It is Math; unclear instruction; class does not work in teams or in two groups; mostly paper and pencil, single answer work sheets with no Investigations, no inquiry based questioning, no opportunity for students to reflect on learning 0=No attempt to use Investigations curriculum and expectations; students do very little or no math during the hour.

Intervention Time: Large Group & small group; clear expectations and instruction; clear purpose; students engaged; goals clear; goals written	0-4	4=The small group (may be one individual) is skill based, using manipulative and other strategies in Math or focused skills in reading (like fluency or decoding); students are engaged, can tell they know where they are in their skills, they can talk about what they are doing; teacher modifies activities if not working. The Large Group is group working with a teacher on some skill, or on shared reading, finding meaning or doing individual/partners work that is a continuation of Center Time. Teacher interacts with students and monitors work of large group. It is not a new lesson. 3= Same as 4 but with one component missing 2= The small group takes place but the Large group is more diffuse, less differentiated. Probably no large group work. There is a minimum of contact with or monitoring by teacher 1=Interventions are not readily discernible in the classroom. Students may be pulled out during this time, there may be one small group but it may not last the whole time and its purpose is not clear 0= No effort to have an organized Intervention Time