

Abstract Title Page

Title: Implementation Work at Scale: An Examination of the Fidelity of Implementation Study of the Scale-Up Effectiveness Trial of Open Court Reading

Authors and Affiliations: Kate Sullivan (AIR), Nance Bell (AIR), Debra Hughes Jones (Consultant), Sarah Caverly (AIR), and Michael Vaden-Kiernan (AIR)

Abstract Body

Background / Context:

The development of approaches for measuring fidelity of implementation (FOI) in studies of educational interventions has progressed in recent years as the focus of educational research has shifted toward rigorous experimental trials. In order to understand the "black box" in traditional experimental designs, assessment of the degree to which an intervention is actually implemented (and how well) and the degree to which implementation in treatment and control groups differs contributes to understanding the overall impact of a specific program. While there is no single, universal approach used in FOI studies, there is a growing consensus about the necessity of addressing a series of specific issues in this area. In general, these issues include: the a priori specification of program models and/or theories of change (Nelson, et al., 2012; Cordray and Pion, 2006), the identification of "critical components" in specific interventions that are assumed to be associated with desired change (O'Donnell, 2008; Century, et al., 2010), the development of precise and reliable measures of these critical components (O'Donnell, 2008; Munter, et al., 2014), and, ultimately, the process of linking the overall assessment of fidelity to program impacts (Nelson, et al, 2012; Hulleman, C. S., & Cordray, D. S., 2009). These priorities were defined in the broader arena of FOI work intended to accompany randomized control trial research and have become the commonly-accepted organizing principles for designing and carrying out implementation research, especially for efficacy trials. However, there are fewer published reports documenting program implementation within the context of effectiveness trials at scale (Pas and Bradshaw, 2012). This study addresses this gap by applying the organizing principles of FOI to a national effectiveness trial of the Open Court Reading (OCR) program and reports on the findings from these efforts.

Purpose / Objective / Research Question / Focus of Study:

The FOI study that is the focus of this report was conducted as a component of a scale-up effectiveness trial of the SRA/McGraw-Hill Open Court Reading program (described further below). The overall purpose of the FOI study was to support and provide context for findings from the larger experimental impact study of OCR. To accomplish this task, the FOI study addressed the following research questions: 1) In what ways were OCR students' experiences similar or different to those of students receiving other reading curricula?, and 2) Was there significant variation in implementation fidelity of OCR among classrooms/teachers, schools, and districts?

The first question addresses the issue of general fidelity (i.e., comparing treatment and control group implementation), and the second examines specific fidelity (i.e., variation of implementation within the treatment condition). General and specific fidelity are both critical areas of investigation to include in FOI study designs intended to inform scale-up effectiveness trials.

Setting:

The FOI study was conducted during two years of OCR implementation in the 49 elementary impact study schools (kindergarten through 5th grade) in seven districts across the country.

Population / Participants / Subjects:

The primary sample for the FOI study included teachers in both treatment and control classrooms whose students were selected for assessment in the impact study (approximately 150 teachers in each year of the study). Student achievement data utilized in the FOI study came from assessment data on approximately 4,500 students each year.

Intervention / Program / Practice:

This FOI study was a component of the cluster randomized controlled effectiveness trial of the OCR program. The OCR impact study included 49 elementary schools across seven districts located in the midwest, south, and western regions of the U.S. Participating schools were randomized into treatment (N=25) and control (N=24) conditions, with blocking at the district level. Random assignment determined whether K-5 students in participating schools were to receive the OCR curriculum (treatment condition) or the standard reading curriculum otherwise being used in the school or district (control condition). The study followed two grade cohorts across two years, assessing a total of approximately 4,500 students from the designated grade levels in the study schools in the fall and spring (students in grades K and 3 were assessed in Year 1 and grades 1 and 4 students in Year 2). Results from the Group Reading Assessment and Diagnostic Evaluation (GRADE) were used as the indicator of student achievement outcomes. While some subgroup differences in reading achievement were revealed, findings from the main impact study revealed no significant overall effects in either the first or second year of the evaluation.

Research Design:

The FOI study was designed to address questions of both general fidelity (what was implemented in treatment vs. control classrooms) and specific fidelity (why, how, and under what conditions OCR instruction was delivered more or less effectively in treatment classrooms). The focus of the general fidelity analyses was on between-group (i.e., treatment and control) differences, while specific fidelity analyses were concerned with within-group (i.e., treatment classrooms only) differences. The conceptual models we used to inform the FOI study design incorporated four of the five dimensions of program integrity identified by Dane and Schneider (1998): adherence, exposure, quality of delivery, and participant responsiveness. In order to further specify the variables of interest, we grouped them into the superordinate categories of structure and process (O'Donnell, 2008; Mowbray, et al., 2003). The specific indicators used to measure structural components included: program materials and program strands observed, minutes of instruction, and teacher stability. For process components we used summative fidelity ratings from classroom observation data and general teacher effectiveness and student responsiveness scores. Contextual factors outside of the classroom that might influence FOI were also taken into account (e.g., school, teacher, and student demographics, administrative support for implementation, and other factors at the state and district levels). In addition to measuring fidelity components, the study design included fully-specified, multilevel models to examine potential linkages between levels of fidelity and student achievement.

Data Collection and Analysis:

Extensive data for the structural and process components of general and specific fidelity were collected from multiple sources, including: archival data, teacher surveys, classroom observations, interviews, and research team member field notes. Archival, survey, and observation data were collected from teachers in both treatment and control groups, and interviews were conducted at the end of each year with treatment teachers only.

The analytical approach for the general fidelity portion of the study conceptualized general fidelity as a single latent variable underlying structural and process fidelity. Data were analyzed using Latent Class Analysis (LCA) (Hagenaars & McCutcheon, 2002; Lanza, Flaherty, & Collins, 2003), which empirically identifies subgroups of teachers based on commonalities or patterns in teachers' structural and process fidelity components. Using the seven indicators of structural and process fidelity as manifest variables for the underlying latent construct of general fidelity, teachers with similar instructional practices were ultimately grouped into categories based on their level of general fidelity in delivering the OCR intervention.

To answer FOI study questions about specific fidelity, variation in fidelity was explored for the group of teachers classified as OCR implementers by LCA models. We conducted separate analyses for the indicators of structural fidelity, which are multi-method and multi-form. The indicators of process fidelity were amenable to analysis as a single latent variable (Abry et al., 2015). Using confirmatory factor analysis, a single latent variable was defined from the measures of teacher effectiveness, student engagement, and quality of delivery. We reviewed univariate analyses for structural indicators and the process latent variable and also explored potential links between structural and process fidelity and student achievement.

Findings / Results:

Results from general fidelity analyses showed, first, that the demographic characteristics of teachers in the year 1 and year 2 samples were very similar. Table 1 displays the descriptive statistics for overall structural and process fidelity indicators by latent class for each evaluation year. (insert Table 1 here) Most teachers remained at their schools for the entire year, resulting in high levels of teacher stability. Teachers also reported high levels of reading instruction. Generally, all teachers were rated as very strong on measures of teacher effectiveness and student engagement. There were no differences for these indicators by grade or treatment condition. Levels of OCR materials, program coverage, and quality of program delivery, however, differed sharply by treatment condition.

Latent class analyses revealed that a two-class model best fit the data. The largest class (54% of teachers in both years) consisted of teachers who showed no evidence of implementing OCR (the non-implementing or NIC class). The second class (46%) consisted of teachers showing strong evidence of program implementation (the implementing or OCR class). Teacher demographics were equivalent across the NIC and OCR classes, and there were no differences by grade taught. As expected, there was significant overlap between treatment condition and latent class membership. All teachers in control schools were grouped into the NIC class, showing no

evidence of program contamination in control schools. Among teachers in treatment schools, more than nine out of ten teachers were members of the OCR class.

Overall, findings from specific fidelity analyses revealed that implementation of the OCR program among treatment teachers was relatively high. Indicators of dosage showed that the foundation for implementation was sufficient (95% of teacher reported 60+ minutes of daily reading instruction) and teacher stability was high (all but three teachers remained at their school for the entire year). Indicators for adherence did point to some deviation from program recommendations – on average teachers covered two of the three OCR instructional strands. The most commonly skipped strand was Language Arts. Similarly, 75% of key program materials were observed classrooms. The latent indicator of process fidelity was based on standardized scores from quality of program delivery, teacher effectiveness, and student engagement. Process fidelity did not vary by grade and showed an even distribution across schools and districts.

Analyses of potential linkages between structural and process fidelity and student achievement revealed that none of the indicators for structural fidelity (dosage and adherence) were predictive of student reading outcomes. However, there was a relationship between teachers' process fidelity and students' achievement. (insert Table 2 here) As shown in Table 2, in both years, students with teachers that scored higher on process fidelity had significantly higher reading outcomes compared to students whose teachers had lower process fidelity. On average, a one standard deviation shift in process fidelity yielded a 1.5 point gain in reading outcomes (Cohort 1 effect size = 0.11; Cohort 2 effect size = 0.09). These findings confirm O'Donnell (2008) and Mowbray, et al. (2003), who emphasize that, while often more difficult to measure, process fidelity components are more likely to be related to student outcomes.

Conclusions:

The general fidelity analyses provided clear evidence of teacher adherence to assigned condition. There was no indication of contamination in the control group and, on the whole, treatment teachers implemented OCR with relatively high fidelity. These findings validate the integrity of findings in the broader impact study. Results from specific fidelity analyses highlight the importance of taking the next step from general fidelity (necessary, and of primary interest in, effectiveness trial studies) to the more focused analysis of variation among implementing teachers. By breaking down the indicators of specific fidelity according to structure and process fidelity, we found that variation in process implementation had important implications for student achievement, but that structural fidelity was not significantly related to either process fidelity or to student achievement outcomes. The study of FOI within in the context of effectiveness trials is both conceptually and empirically complex and, in many ways, unique. We will further discuss specific aspects related to conceptual issues, research questions, and methodological challenges faced and how we have begun to address these.

Appendices

Appendix A. References

- Abry, T., Hulleman, C. S., & Rimm-Kaufman, S. E. (2015). Using indices of fidelity to intervention core components to identify program active ingredients. *American Journal of Evaluation, 36*(3): 320-38.
- Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation, 31*, 199-218.
- Cordray D. S., & Pion G. M. (2006). Treatment strength and integrity: Models and methods. In Bootzin RR, McKnight PE (eds). *Strengthening Research Methodology: Psychological Measurement and Evaluation*. Washington, DC: American Psychological Association: 103–124.
- 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.
- Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied Latent Class Analysis*. Cambridge University Press.
- 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 Intervention Effectiveness, 2*, 88-110.
- Lanza, S. T., Flaherty, B. P., & Collins, L. M. (2003). Latent class and latent transition analysis. *Handbook of Psychology, 2*, 663–685.
- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*, 315-340.
- Munter, C., Wilhelm, A. G., Cobb, P., & Cordray, D. S. (2014). Assessing fidelity of implementation of an unprescribed, diagnostic mathematics intervention. *Journal of Research on Educational Effectiveness, 7*: 83-113.
- Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services and Research, 39*(4): 374-96.
- O'Donnell, C. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*, 33-84.

Pas, E. T., and Bradshaw, C. P. (2012). Examining the association between implementation and outcomes: State-wide scale-up of school-wide positive behavior intervention and supports. *Journal of Behavioral Health Services and Research*, 39(4): 417-33.

Appendix B. Tables and Figures

Table 1. Descriptive Statistics for Indicators of Structural and Process Fidelity

	Year One						Year Two							
	Overall N	Overall P	NIC Class N	NIC Class P	OCR Class N	OCR Class P	Sig	Overall N	Overall P	NIC Class N	NIC Class P	OCR Class N	OCR Class P	Sig
Teacher Sample	157		84	54%	73	46%		147		80	54%	67	46%	
Random Assignment							***							***
Treatment	77	49%	4	5%	73	100%		73	50%	6	8%	67	100%	
Control	80	51%	80	95%	0	0%		74	50%	74	93%	0	0%	
Grade														
Lower	84	54%	45	54%	39	53%		79	54%	43	54%	36	54%	
Upper	73	46%	39	46%	34	47%		67	46%	37	46%	31	46%	
Teacher stability														
No	7	5%	5	6%	2	3%		2	1%	1	1%	1	2%	
Yes	150	96%	79	94%	71	97%		145	99%	79	99%	66	99%	
Minutes of reading														
1-59	5	3%	1	1%	4	6%		1	1%	1	1%	0	0%	
60 - 89	12	8%	4	5%	8	11%		23	16%	13	16%	10	15%	
90 - 119	51	33%	31	37%	20	27%		65	44%	40	50%	25	37%	
120 - 149	46	29%	22	26%	24	33%		38	26%	19	24%	19	28%	
150 or more	15	10%	9	11%	6	8%		18	12%	6	8%	12	18%	
Missing	28	18%	17	20%	11	15%		2	1%	1	1%	1	2%	
	Overall Mean	SD	NIC Class Mean	SD	OCR Class Mean	SD	Sig	Overall Mean	SD	NIC Class Mean	SD	OCR Class Mean	SD	Sig
OCR materials %	35%	0.40	1%	0.03	75%	0.21	***	25%	0.28	1%	0.05	55%	0.11	***
Teacher effectiveness	3.88	0.20	3.88	0.21	3.89	0.18		3.99	0.03	3.99	0.04	3.98	0.04	
Student Engagement	3.78	0.33	3.77	0.36	3.79	0.30		3.90	0.21	3.91	0.17	3.89	0.25	
Quality of delivery	1.75	1.89	0.01	0.12	3.75	0.32	***	1.81	1.96	0.02	0.15	3.92	0.17	***
Strands covered	0.89	1.03	0.01	0.07	1.91	0.56	***	0.75	0.87	0.01	0.08	1.62	0.50	***

Note: *** p < .001

Table 2. Process Fidelity and Student Reading Achievement Outcomes

	Year One			Year Two		
	Estimate		SE	Estimate		SE
Level-1						
Intercept	105.08	***	0.99	110.77	***	1.78
FRPL	-4.76	***	0.76	-2.86	**	0.93
Race/Ethnicity (White)						
Black	-5.64	***	1.39	-3.59	*	1.68
Hispanic	-1.64		1.15	-3.89	*	1.59
Level-2						
Pre-test	0.64	***	0.07	0.78	***	0.08
Grade 3/4	-3.97	***	0.98	-6.90	***	0.86
Process Fidelity	1.59	**	0.51	1.51	*	0.60
District (Derby)						
Muskogee	2.93		1.67	2.16		2.09
Nye	-0.34		1.30	3.95	*	1.72
Pike	2.33		1.34	-0.46		1.52
Pointe						
Coupee	2.69		2.00	8.41	**	2.51
Rapides	-2.05		1.21	-0.41		1.49
Sioux City	-1.96		1.03	2.53		1.39
ICC	0.12			0.16		
Sigma Squared	170.472			250.243		