Abstract Title Page
Not included in page count.

Title: Findings from a Multi-Year Scale-Up Effectiveness Trial of Open-Court Reading (Imagine It!)

Authors and Affiliations: Michael Vaden-Kiernan (AIR), Geoffrey Borman (University of Wisconsin-Madison), Sarah Caverly (AIR), Nance Bell (AIR), Veronica Ruiz de Castilla (AIR), Kate Sullivan (AIR) and Grace Fleming (AIR).

Abstract Body

Background / Context:
This study addresses the effectiveness of a widely used core reading program that reflects the research-based practices recommended by the National Reading Panel (2000). This and other similar programs are increasingly used to prevent reading difficulties and ensure that all children are reading at or above grade level by the end of third grade. Converging evidence from two decades of research suggests that with appropriate instruction, nearly all students can become competent readers (Denton & Mathes, 2003; Lyon, Fletcher, Fuchs, & Chhabra, 2006; Mathes & Denton, 2002; Snow, Burns, & Griffin, 1998). Yet statistics indicate that 65% of fourth grade students and 64% of eighth grade students fail to reach proficient-level reading scores (National Center for Educational Statistics, 2013). Recent initiatives emphasize the critical role of early reading instruction in preventing reading difficulties, recognizing that students who do not learn to read well by third grade are less likely to build vocabulary and interact with a wide variety of texts (Good, Simmons, & Kame'enui, 2001). Such failure can have a long-term impact on children's self-confidence, motivation to learn, performance in school, and success in life (Harris & Sipay, 1990; Juel, 1988; Stanovich, 1986, 2000), and reading difficulties are the most common reason for referral into special education (Donovan & Cross, 2002). Despite these concerns, however, only a few replicable beginning reading programs reviewed by the What Works Clearinghouse 2007) have more than a small amount of evidence suggesting that they have *potentially* positive effects: Success for All; Voyager; Reading Recovery; and Ladders to Literacy[*].

The *Open Court Reading* (OCR) program, published by SRA/McGraw-Hill and widely used since the 1960s, offers a phonics-based K-6 curriculum that shows promise for preventing reading difficulties. According to market research, OCR is among the top reading series (Educational Market Research, 2002). Findings from independent non-experimental evaluations suggest that, in comparison to other reading curricula, OCR is associated with significantly better reading outcomes and may be particularly effective with low-performing students (Edsource, 2006; McRae, 2002; Skindrud & Gersten, 2006). In addition, a recent cluster randomized efficacy trial documented the impact of the OCR program on reading achievement in grades 1 through 5 in 5 schools across the country. Results revealed one-year classroom-level impacts of treatment assignment of approximately one fifth of a standard deviation on the Comprehensive Test of Basic Skills (Borman, Dowling, & Schneck, 2007). Despite the program's widespread use and promising research findings, OCR has not been evaluated rigorously on a large scale as part of an objective, third-party evaluation.

Purpose / Objective / Research Question / Focus of Study:
Effective early reading instruction is critical for preventing later reading difficulties. With two-thirds of 4th grade students failing to achieve proficiency in reading on the National Assessment of Educational Progress in 2013, the need to implement and test effective early reading programs is relevant and pressing. The *Open Court Reading (OCR)* program published by SRA/McGraw-

[*] These programs were rated as having a "small" or "medium to large" amount of evidence which requires at least two studies that meet the WWC evidence screen with 2 schools and a total sample size of at least 350 students or 14 classrooms across the studies. "Potentially positive effects" is evidence of a positive effect in a domain with no overriding contrary evidence.

Hill and widely used for almost 40 years is a phonics-based core-reading program for students in kindergarten to 6th grade that incorporates many of the instructional practices related to phonemic awareness, phonics, fluency, vocabulary, and text comprehension recommended by the National Reading Panel (2000). In this study, an independent research team evaluated the effectiveness of the OCR program in a large national sample of elementary schools at scale, across diverse school populations and conditions, and with no more support than schools would have access to if they had selected OCR as their early reading curriculum apart from participation in a research project.

The results of this study contribute to an understanding of whether OCR is effective in promoting reading proficiency in the elementary grades when implemented "at scale" with typical "real world" levels of support. The study was designed to address the following research questions:

- **Overall Impacts**. Does school-level assignment to the OCR curriculum intervention produce stronger effects on reading achievement than assignment to the "business-as-usual" control condition?
- **Impacts by Subgroups**. Is there significant variation in the outcomes of OCR or do the effects reliably replicate across student subgroups, the sampled classrooms/teachers, schools, and districts?
- **Fidelity of Implementation**. To what extent was the intervention delivered as the curriculum developers indicated it should be implemented? Was there significant variation in implementation fidelity of OCR among the classrooms/teachers, schools, and districts? In what ways were OCR students' experiences similar or different to those of students in the control condition?
- **Proximal Outcomes as Mediators of Impacts**. Is there a significant relationship between proximal student and teacher outcomes, such as fidelity of implementation or student motivation/engagement, and student reading achievement outcomes and does this relationship vary by classrooms/teachers, schools, and districts?

Setting:

The study was conducted in a sample of 49 elementary schools (kindergarten through 5th grade) in 7 districts across the country. The first year focused on the impact of the program in kindergarten and third grade students and classrooms, while the second year focused on first and fourth grade students and classrooms.

Population / Participants / Subjects:

The study participants included approximately 4,500 elementary school students and 1,200 teachers across 49 schools in 7 districts each year of the study. Across the two years of implementation, the Implementation study sample includes 304 teachers, 147 teachers from year 1 and 157 teachers from year 2. Half of the teachers are from treatment schools (49.4%).

Intervention / Program / Practice:

The OCR program is widely used and incorporates the instructional practices recommended in the National Reading Panel (2000) related to phonemic awareness, phonics, fluency, vocabulary, and text comprehension. The OCR curriculum includes student materials, teacher manuals, diagnostic and assessment tools, and test preparation practice guides. The program includes a 2- or 3-day summer workshop to train teachers on program implementation and on-going support

by OCR reading consultants throughout the school year. In all grades (kindergarten through 5th grade), the instructional format is a three-part lesson with specific instruction on vocabulary, comprehension strategies, and comprehension skills. Both informal and formal assessments are used to monitor progress and inform instruction.

Research Design:
The evaluation of the OCR program involved two key elements: the multi-site cluster randomized trial (CRT) and the implementation study. The CRT includes 49 elementary schools across 7 districts that were randomized to receive training and delivery of the *Imagine It!* curriculum (treatment group) or to deliver the standard reading curriculum for the school (control group) blocking at the district level. Districts were recruited over a 3 year timeframe. Schools were randomly assigned to treatment (N=25) and comparison (N=24). The study focused on a grade cohort (grade K-3 and a grade 1-4) each year of the study, assessing approximately 50 students from the designated grade level for each cohort in the fall and spring of each school.

Data Collection and Analysis:
Data from teachers and students in two cohorts (grades K&3 and grades 1&4) were gathered over two school years. Each year of data was collected as cross-sectional and each year is analyzed independently. The pre- and post-test outcomes were assessed in the fall and spring using the *Group Reading Assessment and Diagnostic Evaluation* (GRADE) and Student Motivation Form (SMF). Fidelity of implementation was captured using classroom observations, interviews, and surveys with teachers and other key staff (e.g., curriculum trainers).

The main intent-to-treat (ITT) impact analyses uses a three-level model with school level fall pre-test scores on the GRADE as a covariate and spring post-test scores as the dependent variable, nested within schools, which in turn were nested within districts. Additional subgroup (moderator) analyses will be used to investigate the effects of the OCR program as a function of student baseline characteristics (e.g., age/grade, gender, baseline math proficiency, student engagement), teacher/classroom characteristics (e.g., class size, fidelity of implementation), and school characteristics (e.g., geographic region or locale).

The analysis of fidelity of implementation investigates the latent construct of fidelity of implementation underlying four separate components: dosage, adherence, quality of delivery, and student responsiveness.

Findings / Results:
Table 1 provides school level characteristics of students in the study schools at baseline for Year 1 and Year 2. The analytical sample for Year 1 is comprised of 4,486 students in 49 study schools and 4,392 students in 47[†] schools with valid scores on the Spring GRADE assessment used to measure students' reading proficiency. These data were pulled from the 2011-2012 CCD files and collected from schools. The sample is a heterogeneous group of students based on race/ethnicity and socioeconomic status. There is an even split by sex and grade. This table also provides a test of equivalence for students enrolled in schools randomized to the OCR treatment condition and students enrolled in business as usual schools. At baseline, there were no

[†] One treatment and one control school were consolidated by the district due to opening a new school, thus removing both from the school sample.

significant differences between the demographic characteristics of schools in the treatment and control conditions.

Results from the overall ITT impact analyses are presented in Table 2. The Intraclass Correlation (ICC) of student reading achievement was 0.082 for schools and 0.029 for the district level in year 1 and 0.104 for schools and 0.0 for the district level in Year 2. In both years, the pretest scores were equivalent across treatment and control schools.

The three level multilevel model (student, school, and district) includes grand-centered, school-mean GRADE pretest scores and an indicator for treatment condition. Both predictor variables of pretest and treatment condition are included in level 2 of the model. The fully specified level-1, or within-school model, will nest students within schools within randomization district blocks. The linear model for this level of the analysis is written as

Level-1:

$$Y_{ijk} = \pi_{0jk} + \sum_{n=1}^{x} \pi_{xjk} + e_{ijk}$$

Level 2:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} OCR_{jk} + \beta_{02k} Pretest_{jk} + \sum_{n=2}^{7} \beta_{0xk} + r_{ijk}$$

$$\pi_{xjk} = \beta_{0xk} + \sum_{n=1}^{x} \beta_{xjk}$$

Level 3:

$$\beta_{00k} = \gamma_{000} + u_{000}$$

$$\beta_{01k} = \gamma_{100} + u_{100}$$

$$\beta_{02k} = \gamma_{200}$$

The multilevel model results indicate that the schools mean pretest is predictive of the posttest achievement (see Table 2). A one standard deviation increase in school pretest score is associated with a 0.89 increase in outcome scores. ITT analysis indicates that the OCR program does not have a statistically significant impact on students' reading achievement compared to the business as usual curriculum in place in control schools across either year. However, subgroup analyses did find positive differential impacts of the OCR program by grade level (Kindergarten) and student ethnicity (Hispanic).

Implementation study results from the combined (professional development and classroom) OCR fidelity of implementation model indicated that the program was implemented with adequate to high levels of fidelity across the treatment schools in the first and second year and no evidence of contamination in the control schools in the study.

Conclusions:

This study provides preliminary evidence that the impacts of OCR are not significant on overall students' reading performance when implemented at scale in a large sample of schools after one or two years relative to other core reading curricula. However, there were positive differential impacts for Kindergarten students as well as Hispanic students. Future exploratory mediator analyses will explore other indirect relationships with reading outcomes. These findings are particularly important given the large number of students that are exposed to the program across the country and relatively small number of third-party evaluations of this reading program as well as others for elementary school students.

Appendices

Appendix A. References

Borman, G. D., Dowling, N. M, & Schneck, C. (2007). The national randomized field trial of Open Court Reading. Retrieved from the Interdisciplinary Training Program in Educational Sciences at the University of Wisconsin-Madison Web site: http://itp.wceruw.org/Fall%2007%20seminar/BormanFINAL%20OCR%20RFT9%20to%20EEPA_2_.pdf

Denton, C. A., & Mathes, P. G. (2003). Intervention for struggling readers: Possibilities and challenges. In B.R. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 229-251). Timonium, MD: York Press.

Donovan, M. S., & Cross, C. T. (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press. Available from http://www.nap.edu/catalog/10128.html

EdSource (2006). *Similar students, different results: Why do some schools do better?* Available from http://www.edsource.org/pub_abs_simstu05.cfm

Educational Market Research (2002). Elementary reading market-2002-2003 school year data. *The Complete K-12 Newsletter.*

Good, R. H. M., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high stakes outcomes. *Scientific Studies of Reading, 5*, 257-288.

Harris, A. J., & Sipay, E. R. (1990). *How to increase reading ability: A guide to developmental and remedial methods* (9th ed.). New York: Allyn & Bacon.

Juel, C. (1988). Learning to read and write: A longitudinal study of fifty-four children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447.

Lyon, G. R., Fletcher, J. M., Fuchs, L., & Chhabra, V. (2006). Learning disabilities. In E. Mash & R. Barkley (Eds.), *Treatment of childhood disorders* (3rd ed., pp. 512-591). New York: Guilford Press.

Mathes, P. G. & Denton, C. A. (2002). The prevention and identification of reading disability. *Seminars in Pediatric Neurology, 9*, 185-191.

McRae, D. J. (2002). *Test score gains for Open Court schools in California: Results from three cohorts of schools.* Retrieved from https://www.sraonline.com/download/OCR/Research/testscoresgain.pdf

National Center for Educational Statistics. (2013). *A First Look: 2013 Mathematics and Reading*

(National Assessment of Educational Progress at Grades 4 and 8). Washington, DC: U.S. Department of Education, Institute of Education Sciences.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Retrieved from https://www.nichd.nih.gov/publications/pubs/nrp/Documents/report.pdf

Skindrud, K. & Gersten, R. (2006). An evaluation of two contrasting approaches for improving reading achievement in a large urban district. *The Elementary School Journal, 106* (5*)*, 389-407.

Snow, C. E., Burnes, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

SRA/McGraw-Hill (J. Demery, Director of Research, SRA/McGraw-Hill, personal communication, June 16, 2008).

Stanovich, K. E. (1986). Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360-407.

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers.* New York: Guilford Press.

What Works Clearinghouse (2007, August). *What Works Clearinghouse topic report: Beginning reading.* Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved from: http://ies.ed.gov/ncee/wwc/pdf/BR_TR_08_13_07.pdf

Appendix B. Tables and Figures

Table 1. School Characteristics for Study Schools and by Treatment Assignment at Baseline

	Year One			Year Two		
	All Students	**T_x**	**C_x**	**All Students**	**T_x**	**C_x**
School Characteristics						
Urbanicity						
City	37%	40%	33%	34%	38%	30%
Rural	31%	36%	25%	32%	38%	26%
Suburban	25%	16%	33%	26%	17%	35%
Town	8%	8%	8%	9%	8%	9%
Region						
Midwest	51%	52%	50%	49%	50%	48%
South	41%	40%	42%	43%	42%	44%
West	8%	8%	8%	9%	8%	9%
Title I Eligible	78%	72%	83%	77%	71%	83%
Schoolwide Title I Eligible	78%	72%	83%	77%	71%	83%
Full time teachers (*M*)	30.66	30.521	30.809	31.26	31.111	31.409
Student teacher ratio (M)	15.75	15.784	15.72	15.86	15.847	15.864
School enrollment (M)	480.58	482.417	478.75	491.61	492.826	490.391
Student Demographics						
FRPL Eligible	59%	57%	60%	59%	58%	61%
Students in Grade 3 or 4	50%	51%	49%	50%	50%	50%
Male	51%	51%	52%	51%	50%	52%
Race/Ethnicity						
Non-Hispanic White	67%	69%	66%	67%	69%	65%
Non-Hispanic Black	14%	12%	15%	14%	13%	16%
Hispanic	12%	12%	13%	11%	10%	11%
Other	7%	7%	7%	8%	8%	8%
Special Education	9%	10%	8%	8%	9%	8%
ELL	11%	11%	12%	12%	11%	12%
School Sample Size	49	25	24	47	24	23

Note A: Data for these variables was pulled from the most recent CCD files available – 2011/2012. However, one study school was first operational in the 2013/2014 school year. This school is not represented in these data.

Table 2. Multilevel Model Estimates for Impact of OCR (Imagine It!) on Student Reading Achievement Year 1 and Year 2

	Year 1			Year 2		
	Estimate		**SE**	**Estimate**		**SE**
Level-1						
Intercept	102.35	***	0.72	104.02	***	1.06
Race/Ethnicity (White)						
Black	-4.64	***	0.73	-3.94	***	0.86
Hispanic	-4.07	***	0.73	-4.35	***	0.90
Other	-0.87		0.82	0.56		0.94
Free-Reduced Price Lunch	-3.56	***	0.46	-5.42	***	0.55
Level-2						
Treatment Effect	0.53		0.50	-1.04		0.80
Pretest Mean	0.51	***	0.05	0.62	***	0.08
District (Derby)						
Muskogee	1.82		1.38	1.94		2.16
Nye	-3.55	**	1.065	3.54	*	1.66
Pike	-2.46	**	0.891	0.29		1.38
Pointe Coupee	-1.35		1.191	4.27	*	1.91
Rapides	-1.65		0.922	-1.63		1.37
Sioux City	-0.53		0.744	0.67		1.17
ICC - School	0.10			0.10		
Variance (SD)						
School	0.96	(0.98)		4.49	(2.119)	
Residual	177.41	(13.32)		250.18	(15.817)	
Sigma Squared	177.408			250.178		
Effect Size	0.038			-0.06		
T_x Adjusted Mean	98.903			100.019		
C_x Adjusted Mean	98.37			101.026		
MDE	1.393			2.257		
MDE in SD units	0.099			0.134		

Table 3 Treatment Effects by Subgroup

	Year 1						Year 2					
	Estimate		SE	Effect Size		Variance	SD	Estimate		SE	Effect Size		Variance	SD
Subgroup: Grade														
Treatment	-0.74		0.64		School	0.96	0.98	-0.32		0.93		School	4.56	2.14
Low Grade	-4.18	***	0.56		Residual	174.92	13.23	6.66	***	0.66		Residual	241.21	15.53
Treatment * Low Grade	2.56	**	0.79	0.18				-1.44		0.94	-0.09			
Subgroup: Sex														
Treatment	0.19		0.65		School	1.03	1.02	-1.34		0.93		School	4.32	2.08
Male	-2.47	***	0.57		Residual	176.18	13.27	-3.50	***	0.67		Residual	247.66	15.74
Treatment * Male	0.65		0.79	0.05				0.59		0.95	0.04			
Subgroup: Race – Black														
Treatment	0.75		0.52		School	0.93	0.97	-0.90		0.83		School	4.41	2.10
Black	-3.99	***	0.89		Residual	177.36	13.32	-3.65	**	1.06		Residual	250.20	15.82
Treatment * Black	-1.71		1.34	-0.12				-0.77		1.62	-0.05			
Subgroup: Race – Hispanic														
Treatment	0.21		0.51		School	0.82	0.91	-1.12		0.82		School	4.43	2.02
Hispanic	-5.48	***	0.99		Residual	177.34	13.32	-4.86	***	1.22		Residual	250.18	15.82
Treatment * Hispanic	2.70	*	1.31	0.19				1.01		1.67	0.06			
Subgroup: FRPL														
Treatment	-0.03		0.71		School	0.97	0.98	-2.49	*	1.02		School	4.55	2.13
FRPL	-4.04	***	0.63		Residual	177.35	13.32	-6.65	***	0.76		Residual	249.83	15.81
Treatment * FRPL	0.97		0.88	0.07				2.53	*	1.07	0.15			
Subgroup: Pretest Continuous														
Treatment	0.55		0.50		School	0.93	0.97	-1.03		0.79		School	4.20	2.05
Pretest	0.56	***	0.10		Residual	177.41	13.32	0.50	***	0.12		Residual	250.17	15.82
Treatment * Pretest	-0.08		0.11	-0.01				0.21		0.15	0.01			