Abstract Title Page.

Title:

Comparison of the Reading Proficiency of Third Graders in Michigan's RF and Other Elementary Schools from 2005 to 2006

Author(s):

Joanne F. Carlisle, University of Michigan
Kai Schnabel Cortina, University of Michigan

Abstract Body

Background/context:
The goal of the RF program, Part B of the No Child Left Behind Act of 2001 (NCLB), is to have all children reading at grade level by the end of third grade. There is no legislative precedent for this program, focused as it is on preventing reading failure in the early elementary years (US Department of Education, April 2002). Because the design of the Reading First (RF) initiative was informed by research on effective reading instruction (National Reading Panel Report, 2000) and proposed measures to address the shortcomings of previous Title 1 and other federal initiatives, an important question is whether a program specifically focused on effective reading instruction will bring about significant improvements in students' reading achievement in the early elementary years.

The final report of the RF Impact study compared RF and other elementary schools (hereafter, NonRF schools) in a small number of districts that systematically rank ordered their schools in terms of need (e.g., poverty, underachievement) and assigned schools to RF by moving down the rank order until all available funds had been committed. The RF Impact study used a statistical method referred to as the regression discontinuity design (RDD). The results show that the RF schools did not make greater improvements in reading comprehension than the NonRF schools. Questions have been raised about the extent to which the results of this study would be replicated, were a broader array of states and districts included in the analysis. This depends on the extent to which RDD can be carried out in a given state—an issue that motivated the present study.

Purpose/objective/research question/focus of study:
This study, carried out within the context of the evaluation of RF in Michigan, examines districts' criteria for selecting (and ranking) schools to receive RF funding. We set out not only to determine whether RDD is an appropriate analytic method, but also to consider other analytic approaches, given districts' selection criteria and the profiles of poverty and underachievement in RF and NonRF schools in Michigan.

Setting:
Michigan started its RF program from 2002. Because of the tight timing of the approval of the state's RF application, only a small percent of the eligible schools applied for and were awarded funding in that year (Round 1 schools). For this reason, a second competition was run in 2003. In 2005-2006, the second phase of the states RF program began with a required application (or reapplication) for RF funding. Districts already participating in RF had to reapply for funding for schools with existing RF programs. The two conditions for a district to be eligible for RF funding were (1) significant indicators of economic disadvantage/poverty and (2) significant number of students underachieving in reading.

Population/Participants/Subjects:
Of the 44 districts in RF in Michigan in the fall of 2005, we excluded those with fewer than 6 schools and one district in which all 6 schools were assigned to RF. Table 1 provides information

about the ten districts with RF and NonRF schools, including % poverty (i.e., FRL) and % of students below the proficiency standard in reading (referred to as fail rate). All schools in the 10 districts are included in the analysis (n = 296).

Intervention/Program/Practice:
RF funding is intended to improve the quality of the teaching of reading in kindergarten through third grade classrooms in high poverty schools. The law provides principles that serve as the programmatic requirements; it is not a prescriptive program (US Dept of Education, April 2002). Each state was required to have its RF program approved by the US Department of Education. Key features of RF in Michigan are the professional development program, support for teachers provided by a literacy coach in every school, an uninterrupted 1 ½ hour block of reading instruction, the use of one of five comprehensive programs, and both classroom-based assessments (DIBELS) and a standardized test of reading (ITBS) (Michigan Department of Education, 2002). The professional development program is a combination of Language Essentials for Teachers of Reading and Spelling (Moats, 2003) and Texas Reading Academy. Preliminary analysis of the implementation of RF in certain districts indicates that a number of the required features of RF schools have been implemented in NonRF schools as well. For example, one district with 29 elementary schools supplied the NonRF schools with the same comprehensive program as the RF schools and required administration of DIBELS (Carlisle et al., 2006).

Research Design:
The research design involved specific steps to answer three research questions.
(1) Is it appropriate to use RDD to examine progress in reading in Michigan? We investigated the criteria by which districts selected schools to receive RF funding in their state applications in 2005. Having found that RDD could not be used to study reading progress in Michigan's RF schools, we carried out steps 2 and 3.

(2) What alternative analytic strategy or strategies would be viable in order to compare RF and NonRF schools in large districts in Michigan? We analyzed the basis for districts' selection of schools for RF funding in order to determine the extent to which these suggested reasonable approaches for comparing districts' decrease in underachievement in RF and Non RF schools.

(3) Making use of the analytic strategy best suited to the context of RF in Michigan, do we find a significant difference in the decrease in underachievement in RF and NonRF schools? We compared the RF and NonRF schools in three groups of districts, using analysis of variance and controlling for poverty and previous rates of underachievement in reading (those used in the schools selection process).

Data Collection and Analysis:
Timing of data collection. The second phase of RF in Michigan began with a process of selecting the districts to receive funding for 2005-2008. All eligible districts needed to apply or reapply for funding. In their application, the districts were required to rank order the schools that they had selected to receive RF funding. For districts reapplying for funding, schools that had been participating in RF as of 2002 or 2003 were among those in the rank ordering. We collected

poverty (% FRL for each school from the year the school entered RF) and achievement data (% students in 4th grade not meeting the state's standard for grade-level reading proficiency) in order to determine whether districts had used one or both of these to rank order schools to receive RF funding. Table 1 provides an overview of information about these districts.

Students' Reading Achievement: Michigan Educational Assessment Program (MEAP), English Language Arts (ELA) is a criterion-referenced test that consists of two components: Reading and Writing. Students are categorized into four groups based on their MEAP scores: Apprentice, Basic, Met Expectations, and Exceeded Expectations.
Students are considered not proficient when they score at the Apprentice and Basic levels. We used % of 3rd graders not proficient in reading on this test in fall 2005 as the initial assessment and the % of 4th graders not proficient in reading in fall 2006 as the outcome. Our measure of progress is the % decrease in students not proficient in reading.

School descriptors: We used 2005 data on % FRL and % not proficient on MEAP ELA as controls in our regression analyses. The FRL and MEAP data were provided by the state.

Data analysis: (1) Point biserial correlations were used to examine the relation of districts selected RF and NonRF schools and criteria used for selection (FRL and underachievement). (2) As results indicated that different criteria were used by different districts, we divided the 10 districts into three groups, based on the criteria they used to select schools for RF and examined the relation of these groups and district rankings for RF. (3) We used ANOVA to compare % decrease of underachievement testing the effects of RF and district group, controlling for % FRL and % underachievement.

Findings/Results:
Question 1: Appropriateness of RDD in Michigan. RDD assumes selection of schools into a program based on the criterion that the intervention is likely to influence. As the goal of RF is to increase the average reading skills in the participating schools, RDD implies a consistent selection, based on a reading-related achievement measure. Our analyses of Michigan's data showed that the correlation between selection as RF school and the schools 4th failing rate prior to the year of re-application for the sample as a whole is too small to make RDD a viable analytic strategy. The way the selection process is defined by RF, poverty is an additional criterion. This would lend itself to a bivariate selection mode, which would still allow a modified RDD analysis ("fuzzy" RDD) or suggest a propensity score analysis (D'Agostino 1998). However, in the case of RF in Michigan, the effect of district on the selection process is so pronounced that it is the most important predictor of RF participation in a logistic regression on district, poverty and achievement. The effect of district is compounded by the fact that the districts differed in the way they weighted both aspects in their selection decision. In a statistical model, this translates into specifying interaction effects. Given the limited number of degrees of freedom within most school districts, this adds too many parameters to the model and hence cannot be tested. In sum, the structure and complexity of the selection process into RF in Michigan rules out a RDD as statistical analytic model.

For schools in the 10 districts, the point-biserial correlations between being in RF or NonRF and % FRL was significant, as was the correlation of RF or NonRF and % not proficient on the MEAP ELA. However, the correlation between ranking of schools in the districts and % FRL

was not significant; this was true for % not proficient in reading as well. Thus, it was not appropriate to use RDD to analyze progress of reading in districts in Michigan.

Question 2: Development of an analytic strategy for Michigan. Districts differed in the way they used one or both selection criteria (poverty, low achievement). In order to avoid modeling complex interaction effects, we divided the districts in three groups. This was determined by running a logistic regression RF on poverty and low achievement separately for each district. Group 1 consists of those districts with strong empirical evidence that they used primarily (or exclusively) the poverty information of their schools. The second group was defined as those districts for which low achievement was the basis for selection of RF schools. The third group consists of those districts for which the analysis either had too little power (the smallest districts consists of 7 elementary schools, the largest of 146) or for which both predictors were significant and equal in size. Instead of district as independent variable we introduced district group as an additional predictor of reduction in the MEAP-ELA failing rate in a repeated measurement ANOVA design.

The picture changed dramatically once we ran the analyses separately for each district. Within each district, the higher of the two correlations is .54 with coefficients ranging from .3 to .8. The lack of overall association is due to the fact that the average for both indicators varies noticeably across districts. With an average of 85% of subsidized lunch NonRF, schools in district group 1 exceed the average of the RF schools in 7 out of the remaining 9 districts. With a 4^{th} grade failing rate on 60% the NonRF schools in district group 3 are more qualified for RF than the average RF school in 7 of the other districts.

Table 1 also indicates which criterion was more important for the district to select RF schools. For district group 3, for example, the mean score of RF and NonRF for subsidized lunch is almost identical while the percentage of failing students is about 8% apart. In this district, the failing rate was the driving selection criterion. In district 6 the opposite pattern suggests a selection by subsidized lunch.

Question 3: Comparison of RF and NonRF decrease in underachievement by district groups. Table 2 summarizes the WS-ANOVA results. Note that a reduction of the failing rate represents an improvement of the reading skills. There are three significant effects of the repeated measurement part of the analysis. A significant interaction of Year with the achievement covariate indicates that low achieving school had less decline in the failing rate than school with relatively high achievement scores in 2004. The interaction year x district group means that the change in failing rate from 2005 to 2006 was different for the three groups (averaging across RF and NonRF schools). The most relevant effect for understanding the effect of RF in Michigan is the significant three-way interaction Year x RF x district group, in conjunction with the insignificant interaction Year x RF. If RF had an effect in general (meaning across districts), the interaction Year X RF would have been significant. The three-way interaction means that the RF effect is different for the three district groups. RF was only effective in the predicted direction for group 2 (for whom selection was based on low achievement). All BS-effects were significant. As expected, both covariates (FRL and % underachievement) were associated with achievement in both years.

Conclusions:

The analyses have documented that RDD is not a viable model to analyze the impact of RF on the achievement development in Michigan schools. District turned out to be a key variable in three respects. First of all, the probability to get funding by RF is much larger in the small school districts than in large districts. In addition, the critical level of poverty and/or low achievement varies noticeably across districts. District also differed in the way they used the poverty and achievement information in deciding which schools would participate in RF.

The major finding with respect to the effects of RF is related to the different selection strategies within the district. As the analyses revealed, only those districts that based the selection on achievement criteria were able to reduce the failing rates significantly. This is remarkable, as this group includes the largest school district of Michigan, a district that in the past has stood out as having negative trends. It is likely that this school district was indirectly forced to utilize the achievement indicator to decide which school to include in RF because the poverty level is generally very high; over half of the elementary school have a poverty rate of at least 85%.

The fact that selection into RF based on low achievement turns out to be an important predictor of success is not surprising, once the nature of the intervention is taken into account. Low-achievement schools in Michigan are not automatically high poverty schools. This might suggest that many of the low-achieving school suffer from poor implementation of the RF program components and not from detrimental conditions that are associated with poverty (e.g., low parent involvement). Although our research design does not provide a sound basis for causal inferences, these results suggest that RF might be an effective intervention for the former condition but not for the latter (see also Cortina et al., 2008).

Appendixes

Appendix A. References

Carlisle, J.F., Schilling, S.G., Zeng, J., Cortina, K.S., & Kleyman, Y. (September 2006). *Progress in Reading First in a Michigan School District: A Study of Lansing Elementary Schools.* Technical Report #4.1. Evaluation of Reading First in Michigan. Ann Arbor, MI: University of Michigan.

Cortina, K.S., Carlisle, J.F., & Zeng, J. (January 2008). *Context effects on students' gains in reading comprehension in Reading First schools in Michigan. Technical* Report #6.1. Evaluation of Reading First in Michigan. University of Michigan, Ann Arbor.

D'Agostino, R. B. Jr. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17,* 2265-2281.

Michigan Department of Education. (2002). *Making Reading First in Michigan.* Application submitted to U.S. Department of Education. (Downloaded from www.mireadingfirst.org)

Moats, L. (2003). *Language essentials for teachers of reading and spelling*. Longmont, CO: Sopris West.

National Reading Panel Report. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development.

U.S. Department of Education. (April 2002). *Guidance for the Reading First Program*. Downloaded from www.ed.gov

U.S. Department of Education. (2008). *Final Reading First Impact Study*. Downloaded from http://ies.ed.gov/ncee/pubs/20094038.asp

Appendix B. Tables and Figures

Table 1: Districts FRL and % Fail Rate for RF and NonRF Schools

Report

	Name of School District	% Free/Reduced Price Lunch		%Fail Rate of 4th graders	
		Reading First		Reading First	
		0	1	0	1
Mean	1	85.4825	95.4768	40.6333	33.7000
	2	27.6020	67.0184	38.8071	44.1600
	3	73.5224	73.6543	54.5755	68.5929
	4	74.1658	87.4632	59.9913	69.1091
	5	63.6223	71.5551	48.4600	54.8556
	6	60.8406	71.8047	42.1200	39.6700
	7	62.1000	87.0000	13.8000	57.5833
	8	73.6739	80.0415	25.6571	60.9429
	9	64.5573	84.5993	31.6600	50.8375
	10	34.4443	51.9949	32.0455	56.3333

Note. Reading First 0 = NonRF; Reading First 1 = RF.

Table 2: Results of the Repeated measurement ANOVA design

Descriptive Statistics

	Reading First	distr3	Mean	Std. Deviation	N
School Level Percent Fail Rate of 3rd graders Fall 05 - did not meet MEAP Standards (Sum % Level 3 and Level 4)	0	1	28.8386	15.65184	44
		2	38.8553	16.72195	103
		3	22.7000	18.47293	23
		Total	34.0771	17.72562	170
	1	1	35.2758	13.71327	33
		2	53.0182	15.54412	33
		3	37.5360	15.06643	25
		Total	42.3308	16.72790	91
	Total	1	31.5974	15.10195	77
		2	42.2919	17.48367	136
		3	30.4271	18.21393	48
		Total	36.9548	17.79383	261
School Level Percent Fail Rate of 4th graders Fall 06- did not meet MEAP Standards (Sum % Level 3 and Level 4)	0	1	32.7273	15.23752	44
		2	38.2845	19.85992	103
		3	21.0478	14.03735	23
		Total	34.5141	18.89317	170
	1	1	37.2242	14.01609	33
		2	46.7909	13.28543	33
		3	45.3000	15.98708	25
		Total	42.9121	14.82312	91
	Total	1	34.6545	14.80202	77
		2	40.3485	18.79461	136
		3	33.6792	19.30433	48
		Total	37.4421	18.00431	261

Note. Reading First 0 = NonRF; Reading First 1 = RF.

Tables 3 and 4: Comparisons of RF and NonRF Schools Within District Groups, Controlling for FRL and Prior Underachievement in Reading

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
year	Wilks' Lambda	.999	.355[a]	1.000	253.000	.552
year * FRL	Wilks' Lambda	.997	.881[a]	1.000	253.000	.349
year * failrateW04_4	Wilks' Lambda	.975	6.373[a]	1.000	253.000	.012
year * RF	Wilks' Lambda	1.000	.004[a]	1.000	253.000	.948
year * distr3	Wilks' Lambda	.951	6.489[a]	2.000	253.000	.002
year * RF * distr3	Wilks' Lambda	.967	4.340[a]	2.000	253.000	.014

a. Exact statistic

b. Design: Intercept + FRL + failrateW04_4 + RF + distr3 + RF * distr3

Within Subjects Design: year

Tests of Between-Subjects Effects

Measure:MEASURE_1

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	1117.338	1	1117.338	3.587	.059
FRL	10954.367	1	10954.367	35.163	.000
failrateW04_4	12080.863	1	12080.863	38.779	.000
RF	1363.977	1	1363.977	4.378	.037
distr3	2963.536	2	1481.768	4.756	.009
RF * distr3	2172.536	2	1086.268	3.487	.032
Error	78817.880	253	311.533		

APA Reference Style Examples

Sample Citation: Journal Article

Hypericum Depression Trial Study Group. (2002). Effect of Hypericum perforatum (St John's Wort) in major depressive disorder: A randomized controlled trial. *JAMA, 287,* 1807–1814.

Sample Citation: Newsletter/Newspaper Article

Brown, L. S. (1993, Spring). My research with oranges. *The Psychology Department Newsletter*, *3*, 2.

Sample Citation: Book

American Psychiatric Association. (1990). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

Booth, W. C., Colomb, G. G., & Williams, J. M. (1995). *The craft of research.* Chicago: University of Chicago Press.

Sample Citation: Chapter or Section in a Book

Stephan, W. G. (1985). Intergroup relations. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (3rd ed., Vol. 2, pp. 599–658). New York: Random House.

Sample Citation: Web Page

Dewey, R. A. (2004). *APA Style Resources by Russ Dewey.* Retrieved September 8, 2004, from http://www.psywww.com/resource/apacrib.htm