

Outcomes in Oklahoma's Reading First Program –Longitudinal Data Analysis

Tania Jarosewich, Censeo Group

Elizabeth Oyer, Evaluation Solutions

Andrea Wheeler, Oklahoma State Department of Education

Paper presented at the annual meeting of the

American Educational Research Association

April 2009, San Diego, CA

Abstract

This paper presents the results of a mixed model analysis of a longitudinal database that tracked student outcomes on the ITBS test among Oklahoma's Reading First participants between 2004 and 2007 to examine the impact of grant implementation, poverty, race, and test year on the ITBS outcomes. The sample is restricted to students who have ITBS data for first, second, and third grades, 6,906 students from Cohort 1 schools and 2,405 students from Cohort 2 schools. For both cohorts test year was significant, with stronger performance in year 2 as compared to year 1. Students eligible for free and reduced lunch showed a weaker performance than those not eligible. Minority students, African American, Hispanic, and Native American, scored lower than Caucasian. For Cohort 1, the effect of a higher score on the Adherence to Reading First Practices scale was a lower score on the ITBS Reading Total score while the effect of a higher score on the Support and Management scale was a higher score on the ITBS Reading Total score. The implementation variables were not related to student outcomes for Cohort 2.

For full paper email Tania@CenseoGroup.com

Reading First is a federally funded effort to improve early reading skills among students in participating schools nationwide and provide participating schools with resources and guidance to help ensure that children read at or above grade level by the end of third grade. The three priorities of the Reading First program are to improve the quality of classroom instruction, to base instruction on scientifically proven methods, and to provide professional development for educators in reading instruction. In addition to these three components, screening and diagnosis of reading difficulties, as well as the monitoring of student progress are expected to contribute to higher reading achievement among Reading First grant participants.

An important difference between earlier reform efforts and those represented by *NCLB's* Reading First is that “schools are accountable not for *delivering* education to students, the historical norm, but for actually educating them – and to high academic standards” (Chubb, Linn, Haycock, & Wiener, 2005, p. 10). Although other grant programs required accountability for spending, none has required the level of accountability and monitoring that is integral to the Reading First program.

Research on Monitoring Fidelity of Program Implementation

Typically, state and federal agencies audit schools for meeting the fiscal aspects of a grant's expectations and have developed processes for schools to report financial information for review but examination of program implementation is less common (Melde, Esbense, & Tusinski, 2006; Zvoch, Letourneau, & Parker, 2007). Often funding agencies are satisfied with allowing an evaluation of grant-funded programs to describe schools' success in meeting goals and, implicitly then, assuming fidelity to the expectations of the grant. However, without measuring fidelity of implementation, the connection between grant activities and outcomes is

weak. Drawing the conclusion that a program has not been effective without adequately examining whether the program was fully implemented has been called Type III error (Kalafat, Illback, & Sanders, 2007) in which observed findings are erroneously attributed to the studied intervention (Dobson & Cook, 1980).

Many strategies to measure the performance of schools are based on research conducted outside of the field of education (see for example, Schmidt, Scanlan, & Bell, 1979; Wholey, 1998, 2001). Only in recent years has there been a concerted effort to measure the fidelity of the implementation of programs for the purpose of evaluating the program's impact with no consensus on the definition or methods of examining fidelity of implementation (Dusenbury, Brannigan, Falco & Hansen, 2003; Fullan & Pomfret, 1977; Scheirer & Rezmovic, 1983). Recent efforts are working to define fidelity of implementation especially as related to program implementation across multiple sites (see Ruiz-Primo, 2006; Zvoch, Letourneau, & Parker, 2007). Four components of fidelity that could be examined include: 1) whether the program was followed as it was designed; 2) whether all of the elements of, and resources necessary to carry out the program were in place; 3) whether the program was delivered consistently across sites; and 4) whether teachers and students received the expected-quality "dosage" of the program (Dane and Schneider, 1998). Each of these issues is critical in identifying elements of a program that may be making a difference in student achievement (Dusenbury, Brannigan, Falco, & Hansen, 2003; Melde, Esbense, & Tusinski, 2006).

Over the past 5 years, the federal government has provided states with nearly \$5 billion to implement Reading First. At least 80% of that money has gone directly to school districts to support their implementation of the program at the school level, which are occurring in

multiple school districts.

States engaged in Reading First were required to assess and evaluate the progress of local educational agencies in meeting the goals of the Reading First program in four areas, state level implementation, student achievement, state and school-level program effectiveness, and reduction in the number of students reading below grade level. States had autonomy in determining the best way to meet these expectations.

Monitoring Fidelity of Program Implementation in Reading First

Oklahoma's Reading First evaluation includes a number of data collection methods to examine grant implementation: a school self-assessment, annual site visits, annual staff surveys, and classroom observations. Using these different types of evidence (interview, documents, surveys, observations) to confirm the presence of desired practices allows staff members to provide input into the evaluation (Salzman, Jarosewich, Brown, Dorman, 2008). The specific grant elements that were examined in the implementation review included basic grant requirements (e.g., use of a 90-minute uninterrupted reading block), instructional strategies (e.g., differentiated and explicit instruction), instructional programs and materials (e.g., materials focused on the five components of reading and met student needs), resources, intervention strategies (e.g., frequency and appropriateness of intervention for struggling students), instructional assessments (e.g., use of assessment to guide instruction), instructional leadership (e.g., engagement of the administrator in the project), professional development (e.g., extent to which professional development is meeting teachers' needs), and evaluation strategies and sustainability (e.g., how the school is ensuring at the school-level that the program is being implemented).

Monitoring Student Outcomes in Reading First

Student outcomes are related to a myriad of student, school, and out-of-school variables. The documented achievement gap between the performance of Caucasian and minority students has been a concern for educators (McCoach, 2006; Meece & Kurtz-Costes, 2001) as has the relationship between a child's socioeconomic status and reaching achievement (Bowey, 1995; Entwisle, et al., 2005; Jimerson, Egeland, & Teo, 1999). Examining the progress of sub-groups based on SES status and race is a key requirement of NCLB (No Child Left Behind Act of 2001) in order to ensure that educators and stakeholders understand the progress that each group is making towards the goal of reading on grade-level by a child's third grade year.

The Oklahoma Reading First grant has analyzed student performance in Reading First schools on the Dynamic Indicators of Basic Literacy (DIBELS) and Iowa Test of Basic Skills (ITBS) tests over the course of the grant period. Overall, the majority of schools that have participated in the project have shown significant increases ITBS average scores and in the percentage of students who have met DIBELS benchmarks goals (Jarosewich, 2008).

Objectives of the Paper

The annual evaluation of the Oklahoma Reading First grant has collected information about fidelity of grant implementation and student outcomes. This paper presents the results of an analysis of a longitudinal database that tracked student outcomes on the ITBS test among Oklahoma's Reading First participants between 2004 and 2007. The analysis examined the impact of site implementation, poverty, race, and test year on the ITBS outcomes to identify variables that were related to positive student outcomes in participating schools.

Methods

Sample

The first cohort of schools joined the Oklahoma Reading First program in the 2003-2004 school year. A second cohort of schools was awarded funding in the 2004-2005 school year. The Oklahoma State Department of Education provided the evaluators with student-level ITBS Reading Total scores for the 2003-2004 through 2007-2008 school years. The evaluation team created a longitudinal database, which includes ITBS scores for 25,812 students. This study is restricted to those students who attended a Reading First school from first through third grade and who have ITBS data for each of these three years. The study includes data for 6,906 students from Cohort 1 schools and 2,405 students from Cohort 2 schools.

Instrumentation

Iowa Test of Basic Skills (ITBS). The ITBS is a norm-referenced achievement test that compares a student's performance to other students nationwide. First through third grade students in Oklahoma's Reading First schools took the ITBS Reading Vocabulary and Reading Comprehension subtests in each spring of the grant year. The ITBS Reading Total score was used in this study.

Student Variables. Students' race and eligibility for the National School Lunch Program, both obtained from the ITBS data set, were included in the mixed model.

School Variables. The 2008 spring implementation site visitor scores were included in the model to examine the extent to which each school was implementing the expected elements of the Reading First grant. The site visitors conducted classroom observations; interviewed teachers, coaches, and administrators; and examined extant data. Site visitors scored the extent to which each of a number of elements was evident. A cluster analysis of the

site visit scores suggested three clusters: Support and Management, Instructional Practices, and Adherence to Reading First Practices. Reading First Support and Management measures the extent of the leadership team's (not just the Reading Coach's) leadership of the Reading First grant, the Reading Coach's feedback and support, the district's support of grant implementation, effective monitoring of grant implementation and student progress, and teacher and principal engagement in grant implementation. The second cluster, Instructional Practices, measures the extent of explicit instruction, differentiated instruction, varied grouping practices, effective use of instructional time, classroom management, and student engagement. The final implementation component, Adherence to Reading First Practices, measures the consistent use of the 90 minute uninterrupted instructional timeframe, attention to the five components of reading, adherence to the scope and sequence of the core program, use of materials aligned with the core program, coordination of Reading First with other efforts, and provision of appropriate intervention for students. The sums of each of the elements in each of the three clusters were entered into the mixed model.

Statistical Analysis

SPSS was used to test a multi-level model with repeated measures of reading on ITBS Reading Total score NCE. The data set included the years of 2004-2008. Only students that had data for each grade, first through third, were included in the model. Increases in scores over the three years (differences within students) were modeled for students who were grouped in schools (between student differences). Student level (race and poverty) and school-level (program implementation) variables were tested. Separate models were tested for cohorts of schools starting in 2004 (cohort 1) and in 2005 (cohort 2). Different variances for the different

subject groups (students within schools) across time intervals were assumed with the diagonal covariance structure.

Results

Table 3 presents the descriptive statistics for students in Cohort 1 and Cohort 2 for each of the student variables. In general, Cohort 1 schools had higher ITBS mean scores for each variable under study. The mixed model results for both cohorts are presented in *Table 2* and the information criteria in *Table 3*.

Results for Cohort 1

The final model for Cohort 1 suggests that test year was a significant variable predicting reading total in the model ($p < .0001$). Students' scores increased approximately 4 points from wave 1 to wave 3. The change between test wave 2 and test wave 3 was not significant ($p = .399$).

Table 1
Descriptive Statistics

	Cohort 1 ITBS Mean Score (Standard Error)	Cohort 2 ITBS Mean Score (Standard Error)
Year		
Year 1	51.88 (0.63)	46.42 (0.90)
Year 2	56.14 (0.60)	50.67 (0.82)
Year 3	55.91 (0.59)	50.74 (0.78)
Race		
Asian, Multiracial, Other	55.56 (2.15)	53.78 (2.91)
Native American	56.66 (0.71)	53.47 (1.40)
African American	52.58 (1.21)	44.41 (1.07)
Hispanic	47.38 (1.09)	40.95 (1.14)
Caucasian	61.03 (0.49)	53.75 (1.03)
SES		
Eligible Free/Reduced	52.97 (0.57)	50.02 (0.77)

	Cohort 1 ITBS Mean Score (Standard Error)	Cohort 2 ITBS Mean Score (Standard Error)
Lunch		
Not Eligible	56.32 (0.66)	48.53 (0.886)
Free/Reduced Lunch		
Adherence to Reading	25.29 (1.38)	22.42 (1.84)
First Practices		
Reading First Support and	25.87 (2.70)	20.40 (2.68)
Management		
Instructional Practices	22.23 (2.55)	18.33 (2.33)

Adherence to Reading First Practices and Support and Management were significantly related to Reading Total Score for Cohort 1. The effect of a higher score on the Adherence to Reading First Practices scale was a lower score on the ITBS Reading Total score. However, the effect of a higher score on the Support and Management scale was a higher score on the ITBS Reading Total score. The model predicts lower Reading Totals for students qualifying for free or reduced lunch (approximately three points) after adjusting for the effects of the Adherence to Reading First Practices and Support and Management scores. Finally, race was a significant variable predicting Reading Total in the model ($p < .0001$). All categories of minority students were predicted to have lower Reading Total Scores than white students, with the largest gap evident between Caucasian and Hispanic students (about 15 points; $p < .0001$). African American students scored about 8.5 points below Caucasian students ($p < .0001$) and about five points above Hispanic students ($p = .001$). Native American students scored about four NCE points below Caucasian students ($p < .0001$), four NCE points higher than African American students ($p < .0001$), and about nine NCE points above Hispanic students ($p < .0001$).

Table 2
Mixed Model Results

	Cohort 1				Cohort 2			
	T	Sig	Estimate	SE	T	Sig	Estimate	SE
Intercept	11.34	.000	74.00	6.52	6.34	.000	46.49	7.33
Test Year (reference category = Year 3)								
Year 1	-11.54	.000	-4.03	0.35	-7.03	.000	-4.32	0.62
Year 2	0.84	.399	0.23	0.28	-0.17	.867	-0.08	0.48
Race (reference category = Caucasian)								
Asian, multiracial, other	-2.48	.013	-5.47	2.21	-0.15	.882	-0.46	3.10
Native American	-5.09	.000	-4.36	0.86	-0.15	.881	-0.26	1.74
African American	-6.48	.000	-8.45	1.30	-6.25	.000	-9.31	1.49
Hispanic	-11.41	.000	-13.65	1.20	-7.64	.000	-12.46	1.63
Free lunch (Yes vs. No)	7.26	.000	3.35	0.46	-2.51	.012	-1.52	0.61
Adherence to Reading First Practices	-2.82	.005	-0.88	0.31	1.48	.140	0.61	0.42
Reading First Support and Management	1.97	.049	0.34	0.17	-0.97	.333	-0.24	0.25

Table 3
Information Criteria for Cohort 1 and Cohort 2 Mixed Model

Information Criteria ^a	-2 Restricted Log Likelihood	Akaike's Information Criterion (AIC)	Schwarz's Bayesian Criterion (BIC)
Cohort 1-Test Wave	56431.468	56439.468	56466.768
Cohort 1-Test Wave, Race, Poverty	56166.986	56174.986	56202.283
Cohort 1-Test Wave, Race, Poverty, Reading First Implementation, Support and Management	56161.478	56169.478	56196.774
Cohort 2-Test Wave	19694.749	19702.749	19725.826
Cohort 2-Test Wave, Race, Poverty, Reading First Implementation (NS), Support and Management (NS)	19581.104	19589.104	19612.170
Cohort 2-Test Wave, Race, Poverty	19581.670	19589.670	19612.740

Note: Dependent Variable: Reading Total NCE.
Repeated effects - Diagonal Covariance Structure;
Random effects - Variance Covariance Structure;
Students nested in schools

Results for Cohort 2

For Cohort 2, test year was a significant variable predicting reading total in the model ($p < .0001$). Students' scores increased approximately 4 points from wave 1 to wave 3. The increase from test wave 2 to test wave 3 was not significant ($p = .081$).

The model predicts a Reading Total score of approximately one NCE point lower for students qualifying for free or reduced lunch ($p = .012$).

Race was a significant variable predicting Reading Total for the Cohort 2 model ($p < .0001$). All categories of minority students were predicted to have lower Reading Total Scores than Caucasian students, with the largest gap evident between Caucasian and Hispanic students (about 13 points; $p < .0001$). African American students scored about nine points below Caucasian students ($p < .0001$) and about three points above Hispanic students ($p = .048$). The performance of Native American students was about the same as the performance of Caucasian students ($p = .881$). Native American students scored nine NCE points higher than African American students ($p < .0001$), and about 12 NCE points above Hispanic students ($p < .0001$).

The implementation scores were not significantly related to Reading Total Score for Cohort 2.

Discussion

The goal of the Reading First program is to improve early reading skills among participating students. Across the five years of Reading First implementation in Oklahoma, the overall pattern of scores on the ITBS Total Reading Score has shown that a greater percentage of students has improved performance on this assessment measure. The successes are particularly evident among Cohort 1 schools, which have been implementing the program for

one year longer than Cohort 2 schools. The gains among Cohort 1 schools on the ITBS include improvement among all racial and ethnic groups, both genders, and all grade levels.

The goal of this study was to examine more closely the variables that impact student ITBS performance through a multi-level model of a longitudinal student-level dataset. The inclusion of implementation ratings in addition to the variables of test year, race, and poverty status added to the body of literature about implementation and impact of the federal Reading First program.

The mixed model hierarchical linear model analyses with students nested in school districts showed that for Cohort 1, student race, free lunch status, test year, and two of the school-level implementation had an effect on the ITBS Total Reading score. The effect of a higher score on the Adherence to Reading First Practices scale was a lower score on the ITBS Reading Total score. The effect of a higher score on the Support and Management scale was a higher score on the ITBS Reading Total score.

For students in Cohort 2 schools, race, free lunch status, and test year had an effect on the outcome measure. Implementation measures were not included in the final model.

Consistent with previous research, students eligible for free and reduced lunch and minority students scored lower than students who were not eligible for the school lunch program and Caucasian students in both cohorts. The differences among the participating Reading First schools in reading achievement were at least partially explained by differences in the demographics of their population. In order to be eligible to participate in the Reading First program, the district and school had to have a certain minimum percentage of students eligible for the free and reduced program. Even within this sample of schools, all of which had a high

percentage of students living in poverty, the differences in that percentage had an impact on student outcomes.

The program implementation information was intended not only to monitor levels of implementation but also to provide information for Reading First school teams to identify areas of strength in their implementation and areas in which they could improve. This was particularly important since self-assessments conducted by school teams suggested that the majority of schools thought that they were implementing areas of their Reading First grants with fidelity, results that were not always consistent with site visit information, and not surprising according to other researchers (Dusenbury, Brannigan, Falco, & Hansen, 2003; Lillehoj, Griffin, & Spoth, 2004). Two of the implementation variables were retained in the final model for Cohort 1. Schools that earned higher scores on the Adherence to Reading First Practices earned lower mean NCE scores on the ITBS Reading Total. This implementation variable measured the consistent use of the 90 minute uninterrupted instructional timeframe, attention to the five components of reading, adherence to the scope and sequence of the core program, use of materials aligned with the core program, coordination of Reading First with other efforts, and provision of appropriate intervention for students. The reason for the negative relationship between this factor and student outcomes is unclear. It is possible that the implementation measures did not accurately measure schools' functioning in this area. Alternately, it is possible that in the schools that adhered to the core program and ensured that materials were aligned with the core program, the core programs themselves were not meeting student needs. The sample of students in this study consisted of students who attended a Reading First school for the first through third grade. It might be that in this sample of students,

who had a stable early educational experience, adherence to the core program was not as beneficial if these students had experienced a more challenging or varied curriculum. Also, this element of implementation included a measure of intervention. However, it is possible that these students may have benefitted from acceleration or enrichment rather than intervention. Additional analyses may help to identify the relative contribution of each of the elements of this factor on student outcomes to help clarify which elements of implementation are supporting improved student outcomes.

Schools in Oklahoma participated in textbook adoption at the end of the 2007-2008 school year and the majority of Reading First schools chose an updated version of the core program that they were using or a different textbook series. The extent to which the finding from this current year's analysis continues in the next school year will be examined in the next evaluation cycle.

In contrast to the effect of the adherence variable, Cohort 1 schools with a higher score on the Support and Management scale had a higher score on the ITBS Reading Total score. That is, schools in which leadership of the Reading First grant was distributed among staff, the Reading Coach provided effective feedback and support, the district supported grant implementation, and the school had an effective system by which to monitor grant implementation and student progress, student outcomes were improved. In these schools, staff members were engaged in grant implementation and monitoring their success towards meeting expected outcomes.

The reason why these implementation variables were not included in the final model for Cohort 2 is unclear. The general level of implementation among Cohort 2 schools was lower

than the level of implementation among Cohort 1 schools. It could be that the Cohort 2 schools' implementation was too low to have an effect on student outcomes and with another year of implementation, the model might change. Alternately, perhaps differences in other aspects of the schools' culture or leadership systems affected implementation and its relationship with student outcomes.

While the literature suggests that higher implementation is related to higher student outcomes, examining different aspects of implementation and their relationship with student outcomes such as indicated in this study is recommended in future evaluation and research projects. As indicated in this project, accountability to clear standards may not be sufficient to ensure desired outcomes. In the schools under study in this project, as in other similar studies, there were a number of schools that showed high levels of fidelity of implementation, but not the expected student gains (Salzman, Jarosewich, Brown, Dorman, 2008). This may be due to the difference between ritually complying with grant expectations and authentically implementing a school reform (Hill, Maucione, & Hood, 2007).

Limitations

There are several limitations to the conclusions that can be drawn from this study. First, variables that were not included in the present models may be better predictors of reading growth or may better explain the observed relationships among the included variables and reading growth. Also, it is impossible to draw causal inferences from the results of this correlational study. The model proposed in this study is intended to describe performance in this sample of schools and may not have the same results as a model developed by a different data set. In particular, this study examined only students who attended a Reading First school

for first through third grade. Using a data set that includes students who had a less-regular educational experience might show a different pattern. Despite these limitations, this study provides valuable information about the relationships among student and school characteristics and students' reading growth and achievement for students who consistently participated in three years of the Reading First grant in Oklahoma.

References

- Bowey, J. A. (1995). Socioeconomic status difference in preschool phonological sensitivity and first-grade reading achievement. *Journal of Educational Psychology, 87*, 476–487.
- Chubb, J., Linn, R., Haycock, K., & Wiener, R. (2005). Do we need to repair the monument? Debating the future of No Child Left Behind. *Education Next, 5*(2), 9-19.
- 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.
- Dobson, D & Cook, T (1980). Avoiding type III error in program evaluation: Results from a field experiment. *Evaluation and Program Planning, 4*, pp 269-276.
- Dusenbury, L., Brannigan, R., Falcom, M. & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*, 237-256.
- Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1992). Summer setback: Race, poverty, school composition, and math achievement in the first two years of school. *American Sociological Review, 57*, 72–84.

Jarosewich, T. (2008). Oklahoma Reading First 2007-2008 External Evaluation Report. Censeo Group LLC.

Jimerson, S., Egeland, B., & Teo, A. (1999). A longitudinal study of achievement trajectories: Factors associated with change. *Journal of Educational Psychology, 91*, 116–126.

Kalafat, J., Illback, R. J., & Sanders, D. (2007). The relationship between implementation fidelity and educational outcomes in a school-based family support program: Development of a model for evaluating multidimensional full-service programs. *Evaluation and Program Planning, 30*, 136-148.

Melde, C., Esbense, F., & Tusinski, K. (2006). Addressing program fidelity using onsite observations and program provider descriptions of program delivery. *Evaluation Review, 30*, 714-740.

McCoach, D. B., O'Connell, A., A., Reis, S. M., Levitt, H. A. (2006). Growing Readers: A Hierarchical Linear Model of Children's Reading Growth during the First 2 Years of School. *Journal of Educational Psychology, 98*, 14-28.

Meece, J., & Kurtz-Costes, B. (2001). Introduction: The schooling of ethnic minority children. *Educational Psychologist, 36*, 1–7.

No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat 1425 (2002).

Salzman, J., Jarosewich, T., Brown, S., & Dorman A. (April, 2008). Measuring Program Implementation: A Tale of Two States. Paper Presented at the American Educational Research Association, New York, NY, 2008

Schmidt, R. E., Scanlon, J. W., & Bell, J. B. (1979). *Evaluability assessment: Making public programs work better*. Rockville, MD: U. S. Department of Health and Human Services, Project Share.

Zvoch, K., Letourneua, L. E., & Parker, R. P. (2007). A multilevel multisite outcomes-by-implementation evaluation of an early childhood literacy model. *American Journal of Evaluation, 28*, 132-150.

Wholey, J. S. (1998). Assessing performance measurement systems. *The Public Manager, 27*(3), 23.

Wholey, J. S. (2001). Managing for results: Roles for evaluators in a new management area. *American Journal of Evaluation, 22*(3), 343-347.

Zvoch, K., Letourneua, L. E., & Parker, R. P. (2007). A multilevel multisite outcomes-by-implementation evaluation of an early childhood literacy model. *American Journal of Evaluation, 28*, 132-150.