

## 2009 SREE Conference Abstract Template

### Abstract Title Page

**Title:**

Findings from the Reading First Impact Study

**Author(s):**

Beth C. Gamse, Ed.D, **Abt Associates**; Megan Horst, Ed.D., **Abt Associates**; Beth Boulay, Ed.D., **Abt Associates**, Fatih Unlu, Ph.D., **Abt Associates**

## Abstract Body

### **Background/context:**

This paper presents final findings from the Reading First Impact Study, a congressionally mandated evaluation of the federal government's \$1.0 billion-per-year initiative to help all children read at or above grade level by the end of third grade. The No Child Left Behind Act of 2001 (PL 107-110, Title I, Part B, Subpart 1) established Reading First (RF) and mandated its evaluation. The Reading First legislation requires programs and instruction to be based on scientific research in reading, and aims to ensure that all children can read at or above grade level by the end of third grade, thereby significantly reducing the number of students who experience difficulties in later years. The overarching goal of Reading First is to improve students' reading achievement. In September 2003, the US Department of Education contracted with Abt Associates to design and conduct the Reading First Impact Study (RFIS).

### **Purpose/objective/research question/focus of study:**

The Reading First Program is a central element of the No Child Left Behind legislation (No Child Left Behind Act, 2001). It builds on findings reported in a national consensus report (NICHD, 2000), about proven strategies to reduce the prevalence of reading difficulty in the early grades in order to help children read at or above grade level by the time they reach third grade. The program provides resources to schools for professional development, for purchasing reading programs/materials, for assessments of students, and for salaries of reading coaches who can help classroom teachers. Reading First departs from historical precedent in its reach into the content of instruction. Traditionally, the federal education agency has been focused on issues of access, and determination of actual curriculum content has been left to states and localities. The legislation contained a savvy blend of federal guidance for local decision-making; it did not endorse any reading materials or professional development or assessments, but rather articulated the procedures states and districts were required to use as they made local decisions about which districts and schools to fund as well as which specific tools those districts and schools selected help to implement the program. The Reading First Impact Study was designed to assess the impact of the Reading First Program on two core domains: classroom reading instruction and student reading achievement. This paper will summarize the methods used and the study's findings on the two questions of program impact.

### **Setting:**

The RF program targets low-income, low-performing elementary schools. The schools are situated in states (and districts) that prepared detailed plans for increasing the use of teachers' research-based instruction through intensive professional development for teachers, reading coaches, and administrators, with the explicit aim of reaching out to all eligible schools over time (No Child Left Behind Act, 2001).

### **Population/Participants/Subjects:**

The study sample includes low-income, low performing schools that were rated as eligible to receive RF funding; half of the study schools received RF funding and the other half did not. The sample was chosen purposefully to meet the requirements of the regression discontinuity design (described below). Of the 28 school districts and one state RF program that met the criteria, 16 districts and the state program participated in the regression discontinuity design. The final

study sample reflected wide variation in district characteristics and provided enough schools to meet the study's sample size requirements. One other school district agreed to randomly assign some of its eligible schools to RF or a control group. The final sample includes 248 schools across these 18 sites.

### **Intervention/Program/Practice:**

The Reading First program's overarching goal is to improve the quality of reading instruction—and thereby improve the reading skills and achievement of children in kindergarten through third grade—by implementing the use of research-based instruction and materials. Reading First aims to increase both the availability and quality of professional development for all K–3 teachers, including special education teachers, so that they have the necessary skills to teach research-based reading programs effectively. An important provision of the RF legislation is that professional development be available to staff in *all* schools, not only those with RF funding. Specifically, K–3 teachers are eligible to participate in professional development paid for by district RF funds, and K–12 Special Education teachers are eligible to participate in professional development paid for by state RF funds.

Reading programs and instruction methods should incorporate the five essential elements of effective primary grade reading instruction, specified in the legislation: 1) phonemic awareness; 2) phonics; 3) vocabulary development; 4) reading fluency, including oral reading skills; and 5) reading comprehension strategies. Reading First also emphasizes the use of assessments, both to monitor progress and to identify and address students' reading problems early, by helping classroom teachers to screen for, identify, and overcome barriers to students' ability to read at grade level by the end of third grade.

### **Research Design:**

The RFIS employs a regression discontinuity design that capitalizes on the systematic process used by a number of school districts to allocate their RF funds. A regression discontinuity design is the strongest quasi-experimental method that exists for estimating program impacts. Under the certain conditions outlined below, all of which are met by the present study, this method can produce unbiased estimates of program impacts: 1) Eligible schools were rank-ordered for funding based on a quantitative rating, such as an indicator of past student reading performance or poverty; 2) A cut-point in the rank-ordered priority list separated schools that did or did not receive RF grants, and this cut-point was set without knowing which schools would then receive funding; and 3) funding decisions were based only on whether a school's rating was above or below its local cut-point; nothing superseded these decisions.

Additionally, the functional form of the relationship between schools' ratings and outcomes can be modeled. Under these conditions, there should be no systematic differences between eligible schools that received a RF grant (RF schools) and eligible schools that did not (non-RF schools), except for the characteristics associated with the school rating used to determine the funding decision. By controlling for differences in schools' ratings, one can then control statistically for all systematic pre-existing differences between the two groups. This makes it possible to estimate the impact of RF by comparing the outcomes for RF schools and non-RF schools in the study sample, controlling for differences in their ratings.

### **Data Collection and Analysis:**

The RFIS is a multi-year study, with data collection over three school years, 2004-05, 2005-06, and 2006-07 (See Exhibit 1). The study collected several different types of data to address its evaluation questions, including direct assessment of students, classroom observations, and surveys of principals, teachers and reading coaches. These data were used to measure students' reading comprehension and decoding skills; teachers' instructional practices in reading; and aspects of scientifically based reading instruction such as professional development in scientifically based reading instruction, amount of reading instruction, supports for struggling readers, and use of assessments. This paper presents impact estimates for these measures, described more fully below.

*Student reading performance* was assessed with the reading comprehension subtest of the Stanford Achievement Test, 10th Edition (SAT 10, Harcourt Assessment, Inc., 2004). The SAT 10 was administered to students in grades one, two and three during fall 2004, spring 2005, spring 2006, and spring 2007. In the spring of 2007 only, first grade students were assessed with the Test of Silent Word Reading Fluency (TOSWRF, Mather et al., 2004), a measure designed to assess students' ability to decode words from among strings of letters.

*Classroom reading instruction* was assessed in first-grade and second-grade reading classes through an observation system developed by the study team called the Instructional Practice in Reading Inventory (IPRI). Observations were conducted during scheduled reading blocks in each sampled classroom on two consecutive days during each wave of data collection: spring 2005, fall 2005 and spring 2006, and fall 2006 and spring 2007. The IPRI, which is designed to record instructional behaviors in a series of three-minute intervals, can be used for observations of varying lengths, reflecting the fact that schools' defined reading blocks can and do vary. Most reading blocks are 90 minutes or more. Eight outcome measures of classroom reading instruction were created from IPRI data to represent the components of reading instruction emphasized by the RF legislation. Six of these measures are reported in terms of the amount of time spent on the various dimensions of instruction. Two of these measures are reported in terms of the proportion of the intervals within each observation.

*Professional development in scientifically based reading instruction, amount of reading instruction, supports for struggling readers, and use of assessments.* Within these four domains, eight outcome measures were created based on data from surveys of principals, reading coaches, and teachers about school and classroom resources. The eight outcome measures represent aspects of scientifically based reading instruction promoted in the RF legislation and guidance. Surveys were fielded in spring 2005 and again in spring 2007; findings presented here are from 2007 surveys only.

The RFIS is comprised of separate regression discontinuity designs for each of the 17 RDD sites, plus a group-randomized experiment for the experimental site; as a result, the impact estimates presented are averaged across the study's 18 sites. The average is weighted in proportion to the number of RF schools in the study sample from each site. Findings presented in this report therefore represent average impacts for the typical RF school in the sample. For each measure, an extension of the statistical model below was used to estimate the impacts of RF in the 17 RDD sites:

$$Y_k = \beta_0 + \beta_1 T_k + \beta_2 R_k + \mu_k \quad (1)$$

where:  $Y_k$  = the outcome measure for school k,  
 $T_k$  = one if school k is a RF school and zero otherwise,  
 $R_k$  = the value of the rating for school k,  
 $\mu_k$  = a random error for school k that is assumed to be independently and identically distributed.

The coefficient  $B_2$  summarizes the continuous relationship between outcomes and ratings that exists on either side of the cut-point. Controlling for this relationship controls for all systematic pre-existing differences between RF schools and non-RF schools. The coefficient,  $B_1$ , for the treatment indicator,  $T_k$ , provides an estimate of the impact of RF. To increase the precision of impact estimates a limited number of covariates (student background characteristics and/or school baseline test scores) were added to the estimation model. In addition, because of clustering (for example, students within classrooms, and classrooms within schools), multi-level models were used to estimate impacts on all outcomes.

### **Findings/Results:**

Exhibit 2 reports average impacts on classroom reading instruction and student reading comprehension pooled across school years 2004-05 and 2005-06 and 2006-07. Exhibit 3 reports average impacts on key components of scientifically based reading instruction from spring 2007. Exhibit 4 reports the average impact on first graders' decoding skills from spring 2007. Impacts were estimated for each study site and averaged across sites in proportion to their number of RF schools in the sample. Average impacts thus represent the typical study school. Key findings are as follows:

- Reading First produced a positive and statistically significant impact on amount of instructional time spent on the five essential components of reading instruction promoted by the program (phonemic awareness, phonics, vocabulary, fluency, and comprehension) in grades one and two. The impact was equivalent to an effect size of 0.33 standard deviations in grade one and 0.46 standard deviations in grade two.
- Reading First produced positive and statistically significant impacts on multiple practices that are promoted by the program, including professional development in scientifically based reading instruction (SBRI), support from full-time reading coaches, amount of reading instruction, and supports available for struggling readers.
- Reading First did not produce a statistically significant impact on student reading comprehension test scores in grades one, two or three.
- Reading First produced a positive and statistically significant impact on decoding among first grade students tested in one school year (spring 2007). The impact was equivalent to an effect size of 0.17 standard deviations. Because the test of students' decoding skills was only administered in a single grade and a single year, it is not possible to provide an estimate of Reading First's overall impact on decoding skills across multiple grades and across all three years of data collection, as was done for reading comprehension.

## Conclusions:

The findings presented in this report are generally consistent with findings presented in the study's Interim Report, which found statistically significant impacts on instructional time spent on the five essential components of reading instruction promoted by the program (phonemic awareness, phonics, vocabulary, fluency, and comprehension) in grades one and two, and which found no statistically significant impact on reading comprehension as measured by the SAT 10. In addition, analyses of the impact of Reading First on aspects of program implementation, as reported by teachers and reading coaches, revealed that the program had statistically significant impacts on several domains. The information obtained from the Test of Silent Word Reading Fluency indicates that Reading First had a positive and statistically significant impact on first grade students' decoding skill.

The \$6 billion investment in the Reading First Program is intended to address the persistent gap in reading achievement in the U.S. As a cornerstone of the No Child Left Behind Act, its focus on improving students' reading skills has significance not only for the funded schools, but also for other elementary schools that serve students who are learning to read. Specifically, the Reading First Program Guidance outlines how states and districts are to use their sub-grants to serve other schools than those directly funded by Reading First (see p. 17).<sup>1</sup> Consequently, the program's reach extends beyond the RF-funded districts and schools. Although the U.S. Congress decreased funding for RF by 61% in fiscal year 2008 and may eliminate funding in fiscal year 2009, many districts and schools believe RF practices have had positive effects on student achievement and are looking for ways to continue funding for RF practices.<sup>2</sup> The RFIS results will contribute to the discussion about the future of these practices in U.S. schools.

---

<sup>1</sup> Guidance for the Reading First program. (2002). Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education.

<sup>2</sup> See, for example, [http://blogs.edweek.org/edweek/curriculum/2008/07/theyre\\_singing\\_reading\\_firsts.html](http://blogs.edweek.org/edweek/curriculum/2008/07/theyre_singing_reading_firsts.html)

## Appendixes

### Appendix A. References

Harcourt Assessment, Inc. (2004). *Stanford Achievement Test series, Tenth Edition technical data report*. San Antonio, TX: Author.

Mather, N., Hammill, D.D., Allen, E.A., and Roberts, R. (2004). *Test of Silent Word Reading Fluency: Examiner's Manual*. Austin, TX: PRO-ED, Inc.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No 00-4769 and 00-4754). Washington, DC: U.S. Government Printing Office.

No Child Left Behind Act. (2001). PL 107-110, Washington DC.

U.S. Department of Education. (2002). *Guidance for the Reading First Program*. Washington DC: Office of Elementary and Secondary Education.

## Appendix B. Tables and Figures

**Exhibit 1: Data Collection Schedule for the Reading First Impact Study**

Data Collection Elements	2004-2005		2005-2006		2006-2007	
	Fall	Spring	Fall	Spring	Fall	Spring
Student Testing	✓	✓		✓		✓
Stanford Achievement Test, 10 <sup>th</sup> Edition (SAT 10)	✓	✓		✓		✓
Test of Silent Word Reading Fluency (TOSWRF)						✓
Classroom Observations		✓	✓	✓	✓	✓
Instructional Practice in Reading Inventory (IPRI)		✓	✓	✓	✓	✓
Student Time-on-Task and Engagement with Print (STEP)			✓	✓	✓	✓
Teacher, Principal, Reading Coach Surveys		✓				✓

**Exhibit 2: Estimated Impacts on Reading Instruction and Reading Comprehension: 2005, 2006, and 2007 (pooled)**

	<b>Actual Mean with Reading First</b>	<b>Estimated Mean without Reading First</b>	<b>Impact</b>	<b>Effect Size of Impact</b>	<b>Statistical Significance of Impact (p-value)</b>
<b>Instruction</b>					
<i>Number of minutes of instruction in the five components combined</i>					
Grade 1	59.23	52.31	6.92*	0.33*	(0.005)
Grade 2	59.08	49.30	9.79*	0.46*	(<0.001)
<i>Percentage of intervals in five components with Highly Explicit Instruction</i>					
Grade 1	29.39	26.10	3.29*	0.18*	(0.018)
Grade 2	30.95	27.95	3.00*	0.16*	(0.040)
<i>Percentage of intervals in five components with High Quality Student Practice</i>					
<i>High Quality Student Practice</i>					
Grade 1	18.44	17.61	0.82	0.05	(0.513)
Grade 2	17.82	14.88	2.94*	0.16*	(0.019)
<b>Reading Comprehension</b>					
<i>Scaled Score</i>					
Grade 1	543.8	539.1	4.7	0.10	(0.083)
Grade 2	584.4	582.8	1.7	0.04	(0.462)
Grade 3	609.1	608.8	0.3	0.01	(0.887)
<i>Percent Reading At or Above Grade Level</i>					
Grade 1	46.0	41.8	4.2	--	(0.104)
Grade 2	38.9	37.3	1.6	--	(0.504)
Grade 3	38.7	38.8	-0.1	--	(0.973)

**NOTES:**

The complete Reading First Impact Study (RFIS) sample includes 248 schools from 18 sites (17 school districts and 1 state) located in 13 states. 125 schools are Reading First schools and 123 are non-Reading First schools. For grade 2 in 2006, one non-RF school could not be included in the analysis because test score data were not available. For grade 3 in 2007, one RF school could not be included in the analysis because test score data were not available.

Impact estimates are statistically adjusted (e.g., take each school's rating, site-specific funding cut-point, and other covariates into account) to reflect the regression discontinuity design of the study.

Values in the "Actual Mean with Reading First" column are actual, unadjusted values for Reading First schools; values in the "Estimated Mean without Reading First" column represent the best estimates of what would have happened in RF schools absent RF funding and are calculated by subtracting the impact estimates from the RF schools' actual mean values.

A two-tailed test of significance was used; statistically significant findings at the  $p \leq .05$  level are indicated by \*.

**EXHIBIT READS:** The observed mean amount of time spent per daily reading block in instruction in the five components combined for first grade classrooms with Reading First was 59.23 minutes. The estimated mean amount of time without Reading First was 52.31 minutes. The impact of Reading First on the amount of time spent in instruction in the five components combined was 6.92 (or 0.33 standard deviations), which was statistically significant ( $p = .005$ ).

*SOURCES:* RFIS SAT 10 administrations in the spring of 2005, 2006, and 2007, as well as from state/district education agencies in those sites that already used the SAT 10 for their standardized testing (i.e., FL, KS, MD, OR); RFIS Instructional Practice in Reading Inventory, spring 2005, fall 2005, spring 2006, fall 2006, and spring 2007; RFIS Student Time-on-Task and Engagement with Print, fall 2005, spring 2006, fall 2006, and spring 2007.

**Exhibit 3: Estimated Impacts on Key Components of Scientifically Based Reading Instruction (SBRI): Spring 2007**

Domain	Actual Mean With Reading First	Estimated Mean Without Reading First	Impact	Effect Size of Impact	Statistical Significance of Impact (p-value)
<b>Professional Development (PD) in SBRI</b>					
Amount of PD in reading received by teachers (hours) <sup>a</sup>	25.84	13.71	12.13*	0.51*	(<0.001)
Teacher receipt of PD in the five essential components of reading instruction (0-5) <sup>a</sup>	4.30	3.75	0.55*	0.31*	(0.010)
Teacher receipt of coaching (proportion) <sup>a</sup>	0.83	0.63	0.20*	0.41*	(<0.001)
Amount of time dedicated to serving as K-3 reading coach (percent) <sup>b,c</sup>	91.06	57.57	33.49*	1.03*	(<0.001)
<b>Amount of Reading Instruction</b>					
Minutes of reading instruction per day <sup>a</sup>	105.71	87.24	18.47*	0.63*	(<0.001)
<b>Supports for Struggling Readers</b>					
Availability of differentiated instructional materials for struggling readers (proportion) <sup>b</sup>	0.98	0.97	0.01	0.15	(0.661)
Provision of extra classroom practice for struggling readers (0-4) <sup>a</sup>	3.79	3.59	0.19*	0.20*	(0.018)
<b>Use of Assessments</b>					
Use of assessments to inform classroom practice (0-3) <sup>a</sup>	2.63	2.45	0.18	0.19	(0.090)

**NOTES:**

<sup>a</sup> Classroom level outcome

<sup>b</sup> School level outcome

<sup>c</sup> The response rates for RF and nonRF reading coach surveys were statistically significantly different (p=0.037). Reading first schools were more likely to have had reading coaches and to have returned reading coach surveys.

<sup>d</sup> Missing data rates ranged from 0.1 to 3.3 percent for teacher survey outcomes (RF: 0.1 to 1.0 percent; non-RF: 0 to 4.9 percent) and 1.3 to 2.8 percent for reading coach and/or principal survey outcomes (RF: 0 to 1.6 percent; non-RF: 2.7 to 4.1 percent). Survey constructs (i.e., those outcomes comprised of more than one survey item) were computed only for observations with complete data, with one qualification: for the construct “minutes spent on reading instruction per day,” the mean was calculated as the total number of minutes reported for last week (over a maximum of 5 days) divided by the number of days with non-missing values. Only those teacher surveys with missing data for all 5 days were missing 0.9 percent).

The complete Reading First Impact Study sample includes 248 schools from 18 sites (17 districts and 1 state) located in 13 states. 125 schools are Reading First schools and 123 are non-Reading First schools.

The effect size of the impact is the impact divided by the actual standard deviation of the outcome for the non-Reading First Schools.

Values in the “Actual Mean with Reading First” column are actual, unadjusted values for Reading First schools; values in the “Estimated Mean without Reading First” column represent the best estimates of what would have happened in RF schools absent RF funding and are calculated by subtracting the impact estimates from the RF schools’ actual mean values.

A two-tailed test of significance was used; statistically significant findings at the p≤.05 level are indicated by \*.

**EXHIBIT READS: The observed mean amount of professional development in reading received by teachers with Reading First was 25.84 hours. The estimated mean amount of professional development in reading received by teachers without Reading First was 13.71 hours. This impact of 12.13 hours was statistically significantly (p<.001).**

*SOURCES: RFIS, Teacher, Reading Coach, and Principal Surveys, spring 2007*

---

**Exhibit 4: Estimated Impacts of Reading First on Decoding Skill: Grade One, Spring 2007**

---

	<b>Actual Mean with Reading First</b>	<b>Estimated Mean without Reading First</b>	<b>Impact</b>	<b>Effect Size of Impact</b>	<b>Statistical Significance of Impact (p-value)</b>
<b>Decoding Skill</b>					
Standard Score	96.9	94.4	2.5 *	0.17 *	(0.025)
<i>Corresponding Grade Equivalent</i>	1.7	1.4			
<i>Corresponding Percentile</i>	42	35			

---

**NOTES:**

The Test of Silent Word Reading Fluency (TOSWRF) sample includes first-graders in 248 schools from 18 sites (17 school districts and 1 state) located in 13 states. 125 schools are Reading First schools and 123 are non-Reading First schools.

The effect size of the impact is the impact divided by the actual standard deviation of the outcome for the non-Reading First Schools from spring 2007 TOSWRF test scores (1<sup>st</sup> grade).

The key metric for the TOSWRF analyses is the standard score, corresponding grade equivalents and percentiles are provided for reference. Although the publisher of the Test of Silent Word Reading Fluency states that straight comparisons between standard scores and grade equivalents will likely yield discrepancies due to the unreliability of the grade equivalents, they are provided because program criteria are sometimes based on grade equivalents (TOSWRF, Mather et al., 2004).

Values in the “Actual Mean with Reading First” column are actual, unadjusted values for Reading First schools; values in the “Estimated Mean without Reading First” column represent the best estimates of what would have happened in RF schools absent RF funding and are calculated by subtracting the impact estimates from the RF schools’ actual mean values.

A two-tailed test of significance was used; statistically significant findings at the  $p \leq .05$  level are indicated by \*.

**EXHIBIT READS: The observed mean silent word reading fluency standard score for first-graders with Reading First was 96.9 standard score points. The estimated mean without Reading First was 94.4 standard score points. The impact of Reading First was 2.5 standard score points (or 0.17 standard deviations), which was statistically significant (p=.025).**

*SOURCES: RFIS TOSWRF administration in spring 2007*

---