

Reading First Impact Study

Final Report

Executive Summary

Reading First Impact Study Final Report

Executive Summary

NOVEMBER 2008

Beth C. Gamse, Project Director, Abt Associates
Robin Tepper Jacob, Abt Associates/University of Michigan
Megan Horst, Abt Associates
Beth Boulay, Abt Associates
Fatih Unlu, Abt Associates

Laurie Bozzi
Linda Caswell
Chris Rodger
W. Carter Smith
Abt Associates

Nancy Brigham
Sheila Rosenblum
Rosenblum Brigham Associates

With the assistance of

Howard Bloom
Yequin He
Corinne Herlihy
James Kemple
Don Laliberty
Ken Lam
Kenyon Maree
Rachel McCormick
Rebecca Unterman
Pei Zhu

NCEE 2009-4039
U.S. DEPARTMENT OF EDUCATION



This report was prepared for the Institute of Education Sciences under Contract No. ED-01-CO-0093/0004. The project officer was Tracy Rimdzius in the National Center for Education Evaluation and Regional Assistance.

U.S. Department of Education

Margaret Spellings

Secretary

Institute of Education Sciences

Grover J. Whitehurst

Director

National Center for Education Evaluation and Regional Assistance

Phoebe Cottingham

Commissioner

November 2008

This report is in the public domain. Authorization to reproduce it in whole or in part is granted. While permission to reprint this publication is not necessary, the citation should be: Gamse, B.C., Jacob, R.T., Horst, M., Boulay, B., and Unlu, F. (2008). *Reading First Impact Study Final Report Executive Summary* (NCEE 2009-4039). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the reports.

To order copies of this report,

- Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 1398, Jessup, MD 20794-1398.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327 (800-USA-LEARN). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 800-437-0833.
- Fax your request to 301-470-1244.
- Order online at www.edpubs.org.

This report also is available on the IES website at <http://ncee.ed.gov>.

Alternate Formats

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

Acknowledgements

The Reading First Impact Study Team would like to thank the students, faculty, and staff in the study's participating schools and districts. Their contributions to the study (via assessments, observations, surveys, and more) are deeply appreciated. We are the beneficiaries of their generosity of time and spirit.

The listed authors of this report represent only a small part of the team involved in this project. We would like to acknowledge the support of staff from Computer Technology Services (for the study's data collection website), from DataStar (for data entry), from MDRC, from Retail Solutions at Work (and the hundreds of classroom observers who participated in intensive training and data collection activities), from Paladin Pictures (for developing training videos for classroom observations), from RMC Research (especially Chris Dwyer, for help on developing instruments and on training observers), from Rosenblum-Brigham Associates (for district site visits), from Westat (Sherry Sanborne and Alex Ratnofsky, for managing the student assessment, and the Student Assessment Coordinators and test administrators), and from Westover (Wanda Camper, LaKisha Dyson, and Pamela Wallace for helping with meeting logistics).

The study has also benefited from both external and internal technical advisors, including:

External Advisors

Josh Angrist
David Card
Robert Brennan
Thomas Cook*
Jack Fletcher*
David Francis
Larry Hedges*
Robinson Hollister*
Guido Imbens
Brian Jacob
David Lee
Sean Reardon
Tim Shanahan*
Judy Singer
Jeff Smith
Faith Stevens*
Petra Todd
Wilbert Van der Klaauw
Sharon Vaughn*

Internal Advisors

Steve Bell (A)
Gordon Berlin (M)
Nancy Burstein (A)
Fred Doolittle (M)
Barbara Goodson (A)
John Hutchins (M)
Jacob Klerman (A)
Marc Moss (A)
Chuck Michalopoulos (M)
Larry Orr (A)
Cris Price (A)
Janet Quint (M)
Howard Rolston (A)

(A—Abt Associates)
(M—MDRC)

* Individuals who have served on the study's Technical Work Group

We also want to recognize the steady contributions of Abt-SRBI staff, including Brenda Rodriguez, Fran Coffey, Kay Ely, Joanne Melton, Judy Meyer, Lynn Reneau, Davyd Roskilly, Jon Schmalz, Estella Sena, and Judy Walker, who were instrumental in completing multiple data collections, and Eileen Fahey, Katherine Linton, and Jan Nicholson for countless hours of production support. Finally, we want to acknowledge Diane Greene, whose wisdom helped us all.

Disclosure of Potential Conflicts of Interests¹

The research team for this evaluation consists of a prime contractor, Abt Associates, and two major subcontractors, MDRC and Westat. None of these organizations or their key staff has financial interests that could be affected by findings from the Reading First Impact Study. No one on the Technical Work Group, convened to provide advice and guidance, has financial interests that could be affected by findings from the evaluation.

¹ Contractors carrying out research and evaluation projects for IES frequently need to obtain expert advice and technical assistance from individuals and entities whose other professional work may not be entirely independent of or separable from the particular tasks they are carrying out for the IES contractor. Contractors endeavor not to put such individuals or entities in positions in which they could bias the analysis and reporting of results, and their potential conflicts of interest are disclosed.

Executive Summary

This report presents findings from the third and final year of the Reading First Impact Study (RFIS), 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. This evaluation is being conducted by Abt Associates and MDRC with collaboration from RMC Research, Rosenblum-Brigham Associates, Westat, Computer Technology Services, DataStar, Field Marketing Incorporated, and Westover Consulting, under the oversight of the U.S. Department of Education, Institute of Education Sciences (IES).

This report examines the impact of Reading First funding on 248 schools in 13 states and includes 17 school districts and one statewide program for a total of 18 sites. The study includes data from three school years: 2004-05, 2005-06 and 2006-07.

The Reading First Impact Study was commissioned to address the following questions:

- 1) What is the impact of Reading First on student reading achievement?
- 2) What is the impact of Reading First on classroom instruction?
- 3) What is the relationship between the degree of implementation of scientifically based reading instruction and student reading achievement?

The primary measure of student reading achievement was the Reading Comprehension subtest from the Stanford Achievement Test—10 (SAT 10), given to students in grades one, two, and three. A secondary measure of student reading achievement in decoding was given to students in first grade. The measure of classroom reading instruction was derived from direct observations of reading instruction, and measures of program implementation were derived from surveys of educational personnel. Findings related to the first two questions are based on results pooled across the study's three years of data collection (2004-05, 2005-06, and 2006-07) for classroom instruction and reading comprehension, results from first grade students in one school year (spring 2007) for decoding, and aspects of program implementation from spring 2007 surveys. 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.

Results are also presented from exploratory analyses that examine some hypotheses about factors that might account for the observed patterns of impacts. These analyses are considered exploratory because the study was not designed to provide a rigorous test of these hypotheses, and therefore the results must be considered as suggestive. Across different potential predictors of student outcomes, these exploratory analyses are based on different subgroups of students, schools, grade levels, and/or years of data collection. Key findings from these exploratory analyses are as follows:

- There was no consistent pattern of effects over time in the impact estimates for reading instruction in grade one or in reading comprehension in any grade. There appeared to be a systematic decline in reading instruction impacts in grade two over time.
- There was no relationship between reading comprehension and the number of years a student was exposed to RF.
- There is no statistically significant site-to-site variation in impacts, either by grade or overall, for classroom reading instruction or student reading comprehension.
- There is a positive association between time spent on the five essential components of reading instruction promoted by the program and reading comprehension measured by the SAT 10, but these findings are sensitive to both model specification and the sample used to estimate the relationship.

The Reading First Program

Reading First promotes instructional practices that have been validated by scientific research (No Child Left Behind Act, 2001). The legislation explicitly defines scientifically based reading research and outlines the specific activities state, district, and school grantees are to carry out based upon such research (No Child Left Behind Act, 2001). The Guidance for the Reading First Program provides further detail to states about the application of research-based approaches in reading (U.S. Department of Education, 2002). Reading First funding can be used for:

- *Reading curricula and materials* that focus on the five essential components of reading instruction as defined in the Reading First legislation: 1) phonemic awareness, 2) phonics, 3) vocabulary, 4) fluency, and 5) comprehension;
- *Professional development and coaching* for teachers on how to use scientifically based reading practices and how to work with struggling readers;
- *Diagnosis and prevention* of early reading difficulties through student screening, interventions for struggling readers, and monitoring of student progress.

Reading First is an ambitious federal program, yet it is also a funding stream that combines local flexibility and national commonalities. The commonalities are reflected in the guidelines to states and districts and schools about allowable uses of resources. The flexibility is reflected in two ways: one, states (and districts) could allocate resources to various categories within target ranges rather than on a strictly formulaic basis, and two, states could make local decisions about the specific choices within given categories (e.g., which materials, reading programs, assessments, professional development providers,

etc.). The activities, programs, and resources that were likely to be implemented across states and districts would therefore reflect both national priorities and local interpretations.

Reading First grants were made available to states between July 2002 and September 2003. By April 2007, states had awarded subgrants to 1,809 school districts, which had provided funds to 5,880 schools.² Districts and schools with the greatest demonstrated need, in terms of student reading proficiency and poverty status, were intended to have the highest funding priority (U.S. Department of Education, 2002). States could reserve up to 20 percent of their Reading First funds to support staff development, technical assistance to districts and schools, and planning, administration and reporting. According to the program guidance, this funding provided “States with the resources and opportunity...to improve instruction beyond the specific districts and schools that receive Reading First subgrants.” (U.S. Department of Education, 2002). Districts could reserve up to 3.5 percent of their Reading First funds for planning and administration (No Child Left Behind Act, 2001). For the purposes of this study, Reading First is defined as the receipt of Reading First funding at the school level.

The Reading First Impact Study

Research Design

The Reading First Impact Study uses a regression discontinuity design that capitalizes on the systematic processes some school districts used to allocate Reading First funds once their states had received RF grants.³ A regression discontinuity design is the strongest quasi-experimental method available to produce unbiased estimates of program impacts. Under certain conditions, all of which are met by the present study, this method can produce unbiased estimates of program impacts. Within each district or site:

- 1) Schools eligible for Reading First grants were rank-ordered for funding based on a quantitative rating, such as an indicator of past student reading performance or poverty;⁴
- 2) A cut-point in the rank-ordered priority list separated schools that did or did not receive Reading First 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.

Also, assuming that the shape of the relationship between schools’ ratings and outcomes is correctly modeled, once the above conditions have been met, there should be no systematic differences between eligible schools that did and did not receive Reading First grants (Reading First and non-Reading First schools respectively), *except* for the characteristics associated with the school ratings used to determine funding decisions. Controlling for differences in schools’ ratings allows one to control statistically for all systematic pre-existing differences between the two groups. One then can estimate the impact of Reading First by comparing the outcomes for Reading First schools and non-Reading First schools in the study

² Data were obtained from the SEDL website (www.sedl.org/readingfirst).

³ Appendix A in the full report indicates when study sites first received their Reading First grants.

⁴ Each study site could (and did) use different metrics to rate or rank schools; it is not necessary for all study sites to use the same metric.

sample, controlling for differences in their ratings. Non-Reading First schools in a regression discontinuity analysis thereby play the same role as do control schools in a randomized experiment—it is their regression-adjusted outcomes that represent the best indications of what outcomes would have been for the treatment group (in this instance, Reading First schools) in the absence of the program being evaluated.

Study Sample

The study sample was selected purposively to meet the requirements of the regression discontinuity design by selecting a sample of sites that had used a systematic rating or ranking process to select their Reading First school grantees. Within these sites, the selection of schools focused on schools as close to the site-specific cut-points as possible in order to obtain schools that were as comparable as possible in the treatment and comparison groups.

The study sample includes 18 study sites: 17 school districts and one state-wide program. Sixteen districts and one state-wide program were selected from among 28 districts and one state-wide program that had demonstrably met the three criteria listed above. One other school district agreed to randomly assign some of its eligible schools to Reading First or a control group. The final selection reflected wide variation in district characteristics and provided enough schools to meet the study's sample size requirements. The regression discontinuity sites provide 238 schools for the analysis, and the randomized experimental site provides 10 schools. Half the schools at each site are Reading First schools and half are non-Reading First schools: in three sites, the study sample includes all the RF schools (in that site), in the remaining 15 sites, the study sample includes some, but not all, of the RF schools (in that site).

At the same time, the study deliberately endeavored to obtain a sample that was geographically diverse and as similar as possible to the population of all RF schools. The final study sample of 248 schools, 125 of which are Reading First schools, represents 44 percent of the Reading First schools in their respective sites (at the time the study selected its sample in 2004). The study's sample of RF schools is large, is quite similar to the population of all RF schools, is geographically diverse, and represents states (and districts) that received their RF grants across the range of RF state award dates. The average Year 1 grant for RF schools in the study sample ranged from about \$81,790 to \$708,240, with a mean of \$188,782. This translates to an average of \$601 per RF student. For more detailed information about the selection process and the study sample, see the study's Interim Report (Gamse, Bloom, Kemple & Jacob, 2008).

Data Collection Schedule and Measures

Exhibit ES.1 summarizes the study's three-year, multi-source data collection plan. The present report is based on data for school years 2004-05, 2005-06, and 2006-07. Data collection included student assessments in reading comprehension and decoding, and classroom observations of teachers' instructional practices in reading, teachers' instructional organization and order, and students' engagement with print. Data were also collected through surveys of teachers, reading coaches, and principals, and interviews of district personnel.

Exhibit ES.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 th 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)			✓	✓	✓	✓
Global Appraisal of Teaching Strategies (GATS)			✓	✓	✓	✓
Teacher, Principal, Reading Coach Surveys		✓				✓
District Staff Interviews		✓				✓

Exhibit ES.2 lists the principal domains for the study, the outcome measures within each domain, and the data sources for each measure. These include:

Student reading performance, 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, with an average completion rate of 83 percent across all administrations. 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. The average completion rate was 86 percent. Three outcome measures of student reading performance were created from SAT 10 and TOSWRF data.

Classroom reading instruction, 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 average completion rate was 98 percent across all years. 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 Reading First legislation.⁵ Six of these measures are reported in terms of the amount of time spent on the

⁵ For ease of explication, the measures created from IPRI data are referred to as the five dimensions of reading instruction (or “the five dimensions”) throughout the report. References to the programmatic emphases as required by legislation are labeled as the five essential components of reading instruction.

Exhibit ES.2: Description of Domains, Outcome Measures, and Data Sources Utilized in the Reading First Impact Study

Domain	Outcome Measure and Description	Source
Student reading performance	Mean scaled scores for 1st, 2nd, and 3rd grade students , represented as a continuous measure of student reading comprehension. Because scaled scores are continuous across grade levels, values for all three grade levels can be shown on a single set of axes.	<i>Stanford Achievement Test, 10th Edition (SAT 10)</i>
	Percentage of 1st, 2nd, and 3rd grade students at or above grade level , based upon established test norms that correspond to grade level performance, by grade and month. The on or above grade level performance percentages were based on the start of the school year, date of the test and the scaled score, as well as the related grade equivalent.	<i>Stanford Achievement Test, 10th Edition (SAT 10)</i>
	Mean standard scores for 1st grade students , represented as a continuous measure of first grade students' decoding skill.	<i>Test of Silent Word Reading Fluency</i>
Classroom reading instruction	Minutes of instruction in phonemic awareness , or how much instructional time 1 st and 2 nd grade teachers spent on phonemic awareness.	<i>RFIS Instructional Practice in Reading Inventory</i>
	Minutes of instruction in phonics , or how much instructional time 1 st and 2 nd grade teachers spent on phonics.	<i>RFIS IPRI</i>
	Minutes of instruction in fluency building , or how much instructional time 1 st and 2 nd grade teachers spent on fluency building.	<i>RFIS IPRI</i>
	Minutes of instruction in vocabulary development , or how much instructional time 1 st and 2 nd grade teachers spent on vocabulary development.	<i>RFIS IPRI</i>
	Minutes of instruction in comprehension , or how much instructional time 1 st and 2 nd grade teachers spent on comprehension of connected text.	<i>RFIS IPRI</i>
	Minutes of instruction in all five dimensions combined , or how much instructional time 1 st and 2 nd grade teachers spent on all five dimensions combined.	<i>RFIS IPRI</i>
	Proportion of each observation with highly explicit instruction , or the proportion of time spent within the five dimensions when teachers used highly explicit instruction (e.g., instruction included teacher modeling, clear explanations, and the use of examples).	<i>RFIS IPRI</i>
	Proportion of each observation with high quality student practice , or the proportion of time spent within the five dimensions when teachers provided students with high quality student practice opportunities (e.g., teachers asked students to practice such word learning strategies as context, word structure, and meanings).	<i>RFIS IPRI</i>
Student engagement with print	Percentage of 1st and 2nd grade students engaged with print , represented as the per-classroom average of the percentage of students engaged with print across three sweeps in each classroom during observed reading instruction.	<i>RFIS Student Time-on-Task and Engagement with Print (STEP)</i>

Exhibit ES.2: Description of Domains, Outcome Measures, and Data Sources Utilized in the Reading First Impact Study (continued)

Domain	Outcome Measure and Description	Source
Professional development in scientifically based reading instruction	Amount of PD in reading received by teachers , or teachers' self-reported number of hours of professional development in reading during 2006-07.	<i>RFIS Teacher Survey</i>
	Teacher receipt of PD in the five essential components of reading instruction , or the number of essential components teachers reported were covered in professional development they received during 2006-07.	<i>RFIS Teacher Survey</i>
	Teacher receipt of coaching , or whether or not a teacher reported receiving coaching or mentoring from a reading coach in reading programs, materials, or strategies in 2006-07.	<i>RFIS Teacher Survey</i>
	Amount of time dedicated to serving as K-3 reading coach , or reading coaches' self-reported percentage of time spent as the K-3 reading coach for their school in 2006-07.	<i>RFIS Reading Coach Survey</i>
Amount of reading instruction	Minutes of reading instruction per day , or teachers' reported average amount of time devoted to reading instruction per day over the prior week.	<i>RFIS Teacher Survey</i>
Supports for struggling readers	Availability of differentiated instructional materials for struggling readers , or whether or not schools reported that specialized instructional materials beyond the core reading program were available for struggling readers.	<i>RFIS Reading Coach and Principal Surveys</i>
	Provision of extra classroom practice for struggling readers , or the number of dimensions in which teachers reported providing extra practice opportunities for struggling students in the past month.	<i>RFIS Teacher Survey</i>
Use of assessments	Use of assessments to inform classroom practice , or the number of instructional purposes for which teachers reported using assessment results.	<i>RFIS Teacher Survey</i>

various dimensions of instruction. Two of these measures are reported in terms of the proportion of the intervals within each observation .

Student engagement with print. Beginning in fall 2005, the study conducted classroom observations using the Student Time-on-Task and Engagement with Print (STEP) instrument to measure the percentage of students engaged in academic work who are reading or writing print. The STEP observation was completed by recording a time-sampled “snapshot” of student engagement three times in each observed classroom, for a total of three such “sweeps” during each STEP observation. The STEP was used to observe classrooms in fall 2005, spring 2006, fall 2006, and spring 2007, with an average completion rate of 98 percent across all years. One outcome measure was created using STEP data.

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 Reading First legislation and guidance. Surveys were fielded in spring 2005 and again in spring 2007 with an average completion rate across all respondents of 73 percent in spring 2005 and 86 percent in spring 2007. This final report includes findings from 2007 surveys only.

Additional data were collected by the study team in order to create measures used in correlational analyses. These data include:

The *Global Appraisal of Teaching Strategies (GATS)*, a 12-item checklist designed to measure teachers' instructional strategies related to overall instructional organization and order, is adapted from The Checklist of Teacher Competencies (Foorman and Schatschneider, 2003). Unlike the IPRI, which focuses on discrete teacher behaviors, the GATS was designed to capture global classroom management and environmental factors. Items covered topics such as the teacher's organization of materials, lesson delivery, responsiveness to students, and behavior management. The GATS was completed by the classroom observer immediately after each IPRI observation, meaning that each sampled classroom was rated on the GATS twice in the fall and twice in the spring in both the 2005-2006 school year and the 2006-2007 school year. The GATS was fielded in fall 2005, spring 2006, fall 2006, and spring 2007, with an average completion rate of over 99 percent. A single measure from the GATS data was created for use in correlational analyses.

Average Impacts on Classroom Reading Instruction, Key Components of Scientifically Based Reading Instruction, and Student Reading Achievement

Exhibit ES.3 reports average impacts on classroom reading instruction and student reading comprehension pooled across school years 2004-05 and 2005-06 and 2006-07.⁶ Exhibit ES.4 reports average impacts on key components of scientifically based reading instruction from spring 2007. Exhibit ES.5 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 Reading First schools in the sample. Average impacts thus represent the typical study school. On average:

- Reading First had a statistically significant impact on the total time that teachers spent on the five essential components of reading instruction promoted by the program in grades one and two.
- Reading First had a statistically significant impact on the use of highly explicit instruction in grades one and two and on the amount of high quality student practice in grade two. Its estimated impact on high quality student practice for grade one was not statistically significant.
- Reading First had no statistically significant impacts on student engagement with print.
- Reading First had a statistically significant impact on the amount of professional development in reading teachers reported receiving; teachers in RF schools reported receiving 25.8 hours of professional development compared to what would have been expected without Reading First (13.7 hours). The program also had a statistically significant impact on teachers' self-reported receipt of professional development in the five essential components of reading instruction; teachers in RF schools reported receiving professional development on an average of 4.3 of 5 components, compared to what would have been expected without Reading First (3.7 components).

⁶ Except for student engagement with print (STEP), which is pooled across the 2005-06 and 2006-07 school years only.

- A statistically significantly greater proportion (20 percent) of teachers in RF schools reported receiving coaching from a reading coach than would be expected without Reading First. The program also had a statistically significant impact on the amount of time reading coaches reported spending in their role as the school’s reading coach; coaches in RF schools reported spending 91.1 percent of their time in this role, 33.5 percentage points more than would be expected without Reading First (57.6 percent).
- Reading First had a statistically significant impact on the amount of time teachers reported spending on reading instruction per day. Teachers in RF schools reported an average of 105.7 minutes per day, 18.5 minutes more than the 87.2 minutes that would be expected without Reading First.
- Reading First had a statistically significant impact on teachers’ provision of extra classroom practice in the essential components of reading instruction in the past month; the impact was 0.2 components.
- There were no statistically significant impacts of Reading First on the availability of differentiated instructional materials for struggling readers or on teachers’ reported use of assessments to inform classroom practice for grouping, diagnostic, and progress monitoring purposes.
- Reading First had no statistically significant impact on students’ reading comprehension scaled scores or the percentages of students whose reading comprehension scores were at or above grade level in grades one, two or three. The average first, second, and third grade student in Reading First schools was reading at the 44th, 39th, and 39th percentile respectively on the end-of-the-year assessment (on average over the three years of data collection).
- Reading First had a positive and statistically significant impact on average scores on the TOSWRF, a measure of decoding skill, equivalent to 2.5 standard score points, or an effect size of 0.17 standard deviations (See Exhibit ES.5). 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.

Exploratory Analyses of Variations in Impacts and Relationships among Outcomes

This report also presents results from exploratory analyses that examine some hypotheses about factors that might account for the pattern of observed impacts presented above. These exploratory analyses are based on analyses of subgroups of students, schools, grade levels, and/or years of data collection. The information is provided as possible avenues for further exploration or for improving Reading First or programs like Reading First. However, the study was not designed to provide a rigorous test of these hypotheses, and therefore the results are only suggestive. Findings from these exploratory analyses include the following:

- Data collected during three school years (2004-05, 2005-06 and 2006-07) were used to examine variation over time in program impacts. No consistent pattern of differential impacts over time was established.
- No relationship was found between the number of years a student was exposed to RF and student reading achievement.

- There was no statistically significant variation in impacts across sites in the study, either by grade or overall, for reading instruction or for reading comprehension.
- Correlational analyses, which are outside the causal framework of the main impact analyses presented in the report, indicate a positive and statistically significant association between time spent on the five essential components of reading instruction promoted by the program and students' reading comprehension. A one-minute increase in time devoted to instruction in the five dimensions per daily reading block was associated with a 0.07 point increase in scaled score points in first grade, and a 0.06 point increase in second grade. This relationship does not hold for models that include other potential mediators of student achievement. However, due to data limitations, these latter models could only be run on a subset of the data; thus, we do not know whether the differences in the findings across models are due to changes in the sample or changes in the model specification itself.

Exhibit ES.3: Estimated Impacts on Reading Comprehension, Instruction, and Percentage of Students Engaged with Print: 2005, 2006, and 2007 (pooled)¹

	Actual Mean with Reading First	Estimated Mean without Reading First	Impact	Effect Size of Impact	Statistical Significance of Impact (p-value)
Instruction					
<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>					
Grade 1	18.44	17.61	0.82	0.05	(0.513)
Grade 2	17.82	14.88	2.94*	0.16*	(0.019)
Reading Comprehension					
<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)
Percentage of Students Engaged with Print					
Grade 1	47.84	42.52	5.33	0.18	(0.070)
Grade 2	50.53	55.27	-4.75	-0.17	(0.104)

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 *.

¹Except for STEP, which is pooled across 2006 and 2007 school years only.

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 ES.4: 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)
Professional Development (PD) in SBRI					
Amount of PD in reading received by teachers (hours) ^a	25.84	13.71	12.13*	0.51*	(<0.001)
Teacher receipt of PD in the five essential components of reading instruction (0-5) ^a	4.30	3.75	0.55*	0.31*	(0.010)
Teacher receipt of coaching (proportion) ^a	0.83	0.63	0.20*	0.41*	(<0.001)
Amount of time dedicated to serving as K-3 reading coach (percent) ^{b,c}	91.06	57.57	33.49*	1.03*	(<0.001)
Amount of Reading Instruction					
Minutes of reading instruction per day ^a	105.71	87.24	18.47*	0.63*	(<0.001)
Supports for Struggling Readers					
Availability of differentiated instructional materials for struggling readers (proportion) ^b	0.98	0.97	0.01	0.15	(0.661)
Provision of extra classroom practice for struggling readers (0-4) ^a	3.79	3.59	0.19*	0.20*	(0.018)
Use of Assessments					
Use of assessments to inform classroom practice (0-3) ^a	2.63	2.45	0.18	0.19	(0.090)

NOTES:

^a Classroom level outcome

^b School level outcome

^c 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.

^d 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 $p \leq .05$ 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 ES.5: Estimated Impacts of Reading First on Decoding Skill: Grade One, Spring 2007

	Actual Mean with Reading First	Estimated Mean without Reading First	Impact	Effect Size of Impact	Statistical Significance of Impact (p-value)
Decoding Skill					
Standard Score	96.9	94.4	2.5 *	0.17 *	(0.025)
Corresponding Grade Equivalent	1.7	1.4			
Corresponding Percentile	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 (1st 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

Summary

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 to data on the instructional and student achievement outcomes reported in the Interim Report, the final report also presents findings based upon information obtained during the study’s third year of data collection: data from a measure of first grade students’ decoding skill, and data from self-reported surveys of educational personnel in study schools.

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 final report also explored a number of hypotheses to explain the pattern of observed impacts. Analyses that explored the association between the length of implementation of Reading First in the study schools and reading comprehension scores, as well as between the number of years students had been exposed to Reading First instruction and reading comprehension scores were inconclusive. No statistically significant variation across sites in the pattern of impacts was found. Correlational analyses suggest that there is a positive association between time spent on the five essential components of reading instruction promoted by the program and reading comprehension measured by the SAT 10, but these findings appear to be sensitive to model specification and the sample used to estimate the relationship.

The study finds, on average, that after several years of funding the Reading First program, it has a consistent positive effect on reading instruction yet no statistically significant impact on student reading comprehension. Findings based on exploratory analyses do not provide consistent or systematic insight into the pattern of observed impacts.

References

- Foorman, B. and Schatschneider, C. (2003). Measuring teaching practice during reading/language arts instruction and its relation to student achievement. In S. R. Vaughn and K. L. Briggs (Eds.), *Reading in the classroom: Systems for observing teaching and learning*. Baltimore: Paul Brookes.
- Gamse, B.C., Bloom, H.S., Kemple, J.J., Jacob, R.T., (2008). Reading First Impact Study: Interim Report (NCEE 2008-4016). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- 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.
- No Child Left Behind Act of 2001, ESEA, 2001, Title 1, Part B, Subpart 1.
- Southwest Educational Development Laboratory. Summary Report: Awards. Retrieved April 20, 2007 from <http://www.sedl.org/readingfirst/report-awards.html>.
- U.S. Department of Education, Office of Elementary and Secondary Education. (2002). Guidance for the Reading First Program. Washington, DC.