

# Appendix G: Subgroup Analyses

This appendix updates analyses completed in the RFIS Interim Report on two subgroups of sites that were identified by the study team on the basis of when they received their initial Reading First grants.<sup>1,2</sup> The analyses below mirror those from the Interim Report for the two main outcomes of reading instruction and reading comprehension updated with an additional school year (2006-07) of data.<sup>3</sup>

The intent of assessing impacts separately for these two groups of sites was to explore whether sites with a longer implementation period (those that received Reading First funding earlier in the grant making period) produced larger impacts than sites with a shorter implementation period (those that received Reading First funding later). Early award sites (10 sites with 111 Reading First schools in the sample) received their initial Reading First grants between April and December 2003. Late award sites (8 sites with 137 Reading First schools in the sample) received their initial Reading First grants between January and August 2004. When the data collection period for the study ended (in June 2007), early award sites had been funded for an average of 46 months, and late award sites had been funded for an average of 37 months.

## Part 1: Subgroup Impacts over Time

Exhibits G.1-G.5 provide impact estimates separately by award group across follow-up years for reading comprehension and minutes in the five dimensions (impact estimates presented in the main body of the report are for the pooled full sample and not by award group).

The first set of analyses in this section explores the year-to-year changes in impacts as Reading First schools gained experience with the program (see Exhibits G.1 and G.2). The study's follow-up periods represent, roughly, years 1 to 3 of program funding for late award sites and years 2 to 4 for early award sites. By separating annualized findings for the two groups, one can see the findings in the context of a specific year of program implementation. Findings indicate that:

- In the early award sites, there is an overall pattern of year-to-year increases in impact estimates for reading comprehension scaled scores (Exhibit G.1). This pattern was statistically significant when grades were pooled, but not for any single grade (Exhibit G.2). In the late award sites, impact estimates increased less consistently over time, and the overall pattern was not statistically significant for the three grades pooled together or for any single grade.

---

<sup>1</sup> See pp. 51-63 in Gamse et al. (2008).

<sup>2</sup> Due to inconsistent availability of information on the date of receipt of RF funding at the school level, all schools within a site were assigned the date at which their site received RF funding from the state.

<sup>3</sup> For this report, analyses were not conducted for the Student Engagement with Print measure.

**Exhibit G.1: Estimated Impacts on Reading Comprehension and Minutes in the Five Dimensions, by Implementation Year, Calendar Year, and Award Status**

	Implementation Year							
	Year 1		Year 2		Year 3		Year 4	
	Impact	(p-value)	Impact	(p-value)	Impact	(p-value)	Impact	(p-value)
<b>Panel 1</b>								
<b>Early Award Sites</b>	<b>2004</b>		<b>2005</b>		<b>2006</b>		<b>2007</b>	
Grade 1								
Percent reading at or above grade level (%)	N/A	N/A	-2.6	(0.708)	-1.9	(0.751)	6.8	(0.324)
Instruction in five dimensions (minutes)	N/A	N/A	5.49	(0.376)	4.26	(0.448)	1.00	(0.850)
Grade 2								
Percent reading at or above grade level (%)	N/A	N/A	-8.2	(0.163)	-6.8	(0.303)	-0.2	(0.970)
Instruction in five dimensions (minutes)	N/A	N/A	10.90	(0.079)	3.88	(0.478)	2.13	(0.685)
Grade 3								
Percent reading at or above grade level (%)	N/A	N/A	-9.9	(0.110)	-7.7	(0.225)	3.6	(0.588)
Instruction in five dimensions (minutes)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
<b>Panel 2</b>								
<b>Late Award Sites</b>	<b>2005</b>		<b>2006</b>		<b>2007</b>			
Grade 1								
Percent reading at or above grade level (%)	6.3	(0.077)	9.4*	(0.024)	5.8	(0.156)	N/A	N/A
Instruction in five dimensions (minutes)	11.51*	(0.001)	12.21*	(0.003)	9.79*	(0.006)	N/A	N/A
Grade 2								
Percent reading at or above grade level (%)	6.3*	(0.028)	5.7	(0.155)	7.3*	(0.049)	N/A	N/A
Instruction in five dimensions (minutes)	14.84*	(<0.001)	16.26*	(<0.001)	9.94*	(0.004)	N/A	N/A
Grade 3								
Percent reading at or above grade level (%)	1.7	(0.537)	4.2	(0.269)	7.4*	(0.035)	N/A	N/A
Instruction in five dimensions (minutes)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

**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. There are 8 late award sites, with 137 schools, and 10 early award sites, with 111 schools.

A two-tailed test of significance was used; statistically significant findings at the  $p \leq 0.05$  level are indicated by \*.

**EXHIBIT READS: The estimated impact of Reading First on percent of students reading at or above grade level in grade 1, early award sites in 2005 was -2.6 percentage points, which was not statistically significant ( $p = .708$ ).**

*SOURCES: RFIS SAT 10 administration in the spring of 2005, 2006, and 2007 as well as from state/district education agencies in those sites that already use the SAT 10 for their standardized testing (i.e., FL, KS, MD, OR); and RFIS Instructional Practice in Reading Inventory, spring 2005, fall 2005, spring 2006, fall 2006 and spring 2007.*

**Exhibit G.2: Change Over Time in Program Impact on Reading Comprehension and Instruction, By Award Status**

<b>Panel 1 Early Award Sites</b>		<b>Reading Comprehension (SAT 10 Scaled Score)</b>	<b>Reading Instruction (min. in 5 Dimensions)</b>
Grade 1	Linear Year-to-Year Change	5.55	-3.98
	SE	3.61	2.17
	p-value	(0.123)	(0.066)
	F-test for overall variation across years	1.537	0.19
	p-value	(0.215)	(0.660)
Grade 2	Linear Year-to-Year Change	0.21	-3.02
	SE	3.28	2.12
	p-value	(0.948)	(0.154)
	F-test for overall variation across years	0.013	2.52
	p-value	(0.987)	(0.113)
Grade 3	Linear Year-to-Year Change	5.94	n.a.
	SE	3.33	n.a.
	p-value	(0.074)	n.a.
	F-test for overall variation across years	1.834	n.a.
	p-value	(0.160)	n.a.
All Available Grades <sup>a</sup>	Linear Year-to-Year Change	3.96*	-3.05*
	SE	1.98	1.52
	p-value	(0.045)	(0.045)
	F-test for overall variation across years	2.708	1.64
	p-value	(0.067)	(0.201)

**Exhibit G.2: Change Over Time in Program Impact on Reading Comprehension and Instruction, By Award Status (continued)**

Panel 2 Late Award Sites		Reading Comprehension (SAT 10 Scaled Score)	Reading Instruction (min. in 5 Dimensions)
Grade 1	Linear Year-to-Year Change	-0.25	-1.50
	SE	2.34	1.38
	p-value	(0.913)	(0.278)
	F-test for overall variation across years	0.909	1.46
	p-value	(0.403)	(0.228)
Grade 2	Linear Year-to-Year Change	0.28	-2.78
	SE	1.92	1.49
	p-value	(0.884)	(0.063)
	F-test for overall variation across years	0.044	2.49
	p-value	(0.957)	(0.115)
Grade 3	Linear Year-to-Year Change	2.19	n.a.
	SE	1.80	n.a.
	p-value	(0.223)	n.a.
	F-test for overall variation across years	0.833	n.a.
	p-value	(0.435)	n.a.
All Available Grades <sup>a</sup>	Linear Year-to-Year Change	0.87	-1.83
	SE	1.17	1.02
	p-value	(0.458)	(0.073)
	F-test for overall variation across years	0.590	2.80
	p-value	(0.554)	(0.094)

**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.

<sup>a</sup> For Reading Comprehension, grades 1-3 were included in the analysis. For Reading Instruction, only grades 1 and 2 were included in the analysis because instructional data were only available for these two grades.

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.

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

**EXHIBIT READS: For grade 1, the program impact on reading comprehension in early award sites increases by 5.55 scaled score points per year between 2005 and 2007. This change was not statistically significant ( $p=.123$ ). The program impact on instruction in the five dimensions of reading instruction decreases by -3.98 minutes per year. This change was not statistically significant ( $p=.066$ ).**

*SOURCES: RFIS SAT 10 administration 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.*

- In the early award sites, there is an overall pattern of year-to-year decreases in the impact estimates for instructional time in the five dimensions (Exhibit G.1). This pattern was statistically significant for the pooled sample of first and second grade teachers, but not for either grade alone (Exhibit G.2). In the late award sites, there was no consistent pattern over time and the year-to-year variation in impacts on instructional time was not statistically significant for both grades pooled together or for any single grade.

The second set of analyses examines the three-year average impacts of Reading First on instructional time in the five dimensions and reading comprehension scaled scores (see Exhibits G.3, G.4 and G.5, which are identical to Exhibits 4.8, 4.9, and 4.10). These analyses explore whether the differences in impacts between the two groups of sites were statistically significant.

- In the early award sites, estimated impacts on instruction in the five dimensions were not statistically significant. In late award sites, estimated impacts on instructional time in the five dimensions were positive and statistically significant. Differences between early and late award sites in their estimates of impacts on instructional time in the five dimensions were not statistically significant.
- In the early award sites, estimated impacts on reading comprehension were not statistically significant. In late award sites, the estimated impacts on reading comprehension scaled scores are positive and statistically significant only for grade two. Differences between early and late award sites in their estimates of impacts on reading comprehension scaled scores were statistically significant for grade two, but not for grade one or grade three.

## **Part 2: Linear Interactions between Program Impacts and Site Characteristics**

Exhibit G.6 presents updated results regarding the change in impact associated with one unit change in three characteristics that distinguish the early award sites from the late award sites: (1) the number of months each site had access to its Reading First grant; (2) the amount of Reading First funding allocated per K-3 student in Reading First schools; and (3) the levels of reading comprehension exhibited by students in non-Reading First schools in fall 2004. Relevant information is presented in the next section below about the site characteristics used in these analyses.

The information in Exhibit G.7 indicates that:

- On average, late award sites allocated more Reading First funding per school and per student than did early award sites. Hence, there may have been a greater concentration of resources to produce change in the late award sites.
- On average, third grade students from schools without Reading First in the late award sites were less likely to be reading at grade level than those from the early award sites. There may have been a greater margin for improvement in the late award sites (since the study does not have data from early award sites from before they began their implementation of RF, it is not possible to know definitively that early award sites had more or less room for improvement).

**Exhibit G.3: Estimated Impacts on Classroom Instruction: 2005, 2006, and 2007 (pooled), by Award Status**

	Actual Mean with Reading First	Estimated Mean without Reading First	Impact	Effect Size of Impact	Statistical Significance of Impact (p-value)
<b>Early Award Sites</b>					
<b>Number of minutes of instruction in the five dimensions combined</b>					
Grade 1	62.02	60.00	2.02	0.10	0.640
Grade 2	63.04	57.49	5.55	0.26	0.223
<b>Percentage of intervals in five dimensions with highly explicit instruction</b>					
Grade 1	29.90	26.12	3.78	0.21	0.067
Grade 2	31.34	31.38	-0.04	0.00	0.987
<b>Percentage of intervals in five dimensions with High Quality Student Practice</b>					
Grade 1	18.18	20.06	-1.88	-0.11	0.336
Grade 2	17.66	14.14	3.53	0.20	0.073
<b>Late Award Sites</b>					
<b>Number of minutes of instruction in the five dimensions combined</b>					
Grade 1	57.04	46.30	10.74*	0.52*	<0.001
Grade 2	55.98	42.90	13.08*	0.62*	<0.001
<b>Percentage of intervals in five dimensions with highly explicit instruction</b>					
Grade 1	28.98	25.98	3.01	0.17	0.109
Grade 2	30.65	25.25	5.40*	0.28*	0.004
<b>Percentage of intervals in five dimensions with High Quality Student Practice</b>					
Grade 1	18.63	15.70	2.93	0.17	0.073
Grade 2	17.95	15.41	2.54	0.14	0.113

**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. There are 8 late award sites, with 137 schools, and 10 early award sites, with 111 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 pooled across the spring 2005, fall 2005, and spring 2006 IPRI data (by grade).

Impact estimates are statistically adjusted 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 in instruction in the five dimensions (phonemic awareness, phonics, vocabulary, fluency, and comprehension) in first grade classrooms with Reading First in early award sites was 62.02 minutes. The estimated mean amount of time without Reading First was 60.00 minutes. The impact of Reading First on the amount of time spent in instruction in the five dimensions was 2.02 minutes (or 0.10 standard deviations), which was not statistically significant (p=.640).**

*SOURCES: RFIS Instructional Practice in Reading Inventory, spring 2005, fall 2005, spring 2006, fall 2006 and spring 2007*

**Exhibit G.4: Estimated Impacts on Reading Comprehension: Spring 2005, 2006, and 2007 (pooled), by Award Status**

	<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>Early Award Sites</b>					
Reading Comprehension					
Grade 1: Scaled Score	546.6	543.8	2.9	0.06	(0.569)
Corresponding Grade Equivalent	1.8	1.7			
Corresponding Percentile	47	44			
Grade 2: Scaled Score	587.4	591.8	-4.4	-0.10	(0.287)
Corresponding Grade Equivalent	2.6	2.7			
Corresponding Percentile	41	45			
Grade 3: Scaled Score	613.1	617.0	-3.9	-0.10	(0.343)
Corresponding Grade Equivalent	3.5	3.6			
Corresponding Percentile	43	46			
<b>Late Award Sites</b>					
Reading Comprehension					
Grade 1: Scaled Score	541.6	536.0	5.6	0.11	(0.061)
Corresponding Grade Equivalent	1.7	1.6			
Corresponding Percentile	43	39			
Grade 2: Scaled Score	582.1	576.1	6.0 *	0.14 *	(0.021)
Corresponding Grade Equivalent	2.4	2.3			
Corresponding Percentile	38	33			
Grade 3: Scaled Score	606.0	602.4	3.5	0.09	(0.108)
Corresponding Grade Equivalent	3.1	3.0			
Corresponding Percentile	36	34			

**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. Among them, there are 8 late award sites, with 137 schools, and 10 early award sites, with 111 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.

The effect size of the impact is the impact divided by the actual standard deviation of the outcome for the non-Reading First Schools pooled across the spring 2005 and 2006 SAT 10 test scores (by grade).

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 reading comprehension score for first-graders with Reading First in the late award sites was 541.6 scaled score points. The estimated mean without Reading First was 536.0 scaled score points. The impact of Reading First was 5.6 scaled score points (or 0.11 standard deviations), which was not statistically significant ( $p = .061$ ).**

*SOURCES: RFIS SAT 10 administration 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).*

**Exhibit G.5: Award Group Differences in Estimated Impacts on Reading Comprehension and Classroom Instruction: 2005, 2006, and 2007 (pooled)**

	Difference in Impact (Early - Late)	Effect Size of Difference	Statistical Significance of Differences (p-value)
<b>Average Scaled Score</b>			
Grade 1	-2.8	-0.06	(0.636)
Grade 2	-10.4*	-0.25*	(0.032)
Grade 3	-7.4	-0.19	(0.110)
<b>Number of minutes spent in instruction in five dimensions combined</b>			
Grade 1	-8.72	-0.42	(0.092)
Grade 2	-7.53	-0.35	(0.155)
<b>Percentage of observation intervals in five dimensions with Highly Explicit Instruction</b>			
Grade 1	0.78	0.04	(0.779)
Grade 2	-5.44	-0.28	(0.068)
<b>High Quality Student Practice</b>			
Grade 1	-4.81	-0.29	(0.059)
Grade 2	0.98	0.05	(0.696)

**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. There are 8 late award sites, with 137 schools, and 10 early award sites, with 111 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 pooled across the spring 2005 and 2006 data (by grade).

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.

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

**EXHIBIT READS: The estimated difference in impact between early and late award sites in grade 1 was -2.8 scaled score points. The effect size of the difference was -0.06 standard deviations. The estimated difference was not statistically significant ( $p=.636$ ).**

*SOURCES: RFIS SAT 10 administration in the spring of 2005, 2006, and 2007 as well as from state/district education agencies in those sites that already use 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.*

**Exhibit G.6: Change in Impact Associated with One Unit of Change In Continuous Dimensions**

		Reading Comprehension (SAT 10 scaled score)	Reading Instruction (min. in 5 Dimensions)
<b>Award Date</b>			
Grade 1	Impact	0.12	-0.14
	SE	0.13	0.09
	p-value	(0.375)	(0.096)
Grade 2	Impact	-0.02	-0.20*
	SE	0.11	0.09
	p-value	(0.859)	(0.032)
Grade 3	Impact	0.12	n.a.
	SE	0.11	n.a.
	p-value	(0.281)	n.a.
<b>Fall 2004 Reading Performance of non-RF Schools</b>			
Grade 1	Impact	0.22	0.18
	SE	0.24	0.23
	p-value	(0.348)	(0.431)
Grade 2	Impact	-0.20	0.28
	SE	0.20	0.23
	p-value	(0.303)	(0.226)
Grade 3	Impact	-0.10	n.a.
	SE	0.18	n.a.
	p-value	(0.590)	n.a.
<b>Reading First Funding Per Student</b>			
Grade 1	Impact	0.02*	0.01
	SE	0.01	0.01
	p-value	(0.039)	(0.390)
Grade 2	Impact	0.02*	0.01
	SE	0.01	0.01
	p-value	(0.011)	(0.191)
Grade 3	Impact	0.01	n.a.
	SE	0.01	n.a.
	p-value	(0.275)	n.a.

**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, 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. There are 8 late award sites totaling 137 schools and 10 early award sites totaling 111 schools. There are 10 high performance non-RF school sites totaling 120 schools, and 8 low performance non-RF sites totaling 128 schools. There are 8 low RF funding sites totaling 126 schools and 10 high RF funding sites totaling 122 schools.

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.

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

**EXHIBIT READS: An increase of one month in Reading First award date in grade 1 is associated with an increase of 0.12 scaled score points in reading comprehension and a decrease of -0.14 minutes of instruction in the five dimensions. Neither of these impacts was statistically significant.**

*SOURCES: RFIS SAT 10 administration in the spring of 2005, 2006, 2007 as well as from state/district education agencies in those sites that already use the SAT 10 for their standardized testing (i.e., FL, KS, MD, OR); and RFIS Instructional Practice in Reading Inventory, spring 2005, fall 2005, spring 2006, fall 2006, and spring 2007.*

---

**Exhibit G.7: Characteristics of Early and Late Award Sites**

---

Characteristic	Early Award Sites	Late Award Sites
Average number of months of Reading First funding (current as of May 2006)	34 months	25 months
Percent of schools in LEA receiving a Reading First grant	35 percent	16 percent
Average Reading First grant amount (per school)	\$97,776	\$143,850
Average Reading First grant amount (per student)	\$432	\$574
Fall 2004 reading performance of comparison schools (percent of students at or above grade level—grades 1, 2, and 3) <sup>a</sup>	54 percent	43 percent

---

**Notes:**

The complete RF study 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. There are 10 early award sites, with 111 schools, and 8 late award sites, with 137 schools.

<sup>a</sup> The RFIS SAT 10 administration in fall 2004 occurred an average of 15 months after Reading First funds were made available in early award sites and an average of 5 months after Reading First funds were made available in the late award sites.

**EXHIBIT READS: Schools in early award sites had received Reading First funding for an average of 34 months (as of May 2006).**

Sources: RFIS SAT 10 administration in fall 2004, <http://www.seidl.org/readingfirst/welcome.html>, <http://www.ed.gov/programs/readingfirst/awards.html>

---

## Part 3: Impact Estimates for Subgroups Defined by Site Characteristics

### Award Date

The award date information was obtained from Reading First district coordinators in November 2005. District coordinators were asked to provide the month and year that Reading First money was made available to schools in their respective districts. The continuous award variable was then calculated as the number of months between the month/year the funds became available to each site and January 2003. For example, if the funds became available to a site in April of 2003, the continuous award variable for that site would be 3. When dividing the sample into two groups, the study grouped those sites that received funding April and December 2003 as early award sites, and those sites funded between January and August 2004 as late award sites.

### Fall 2004 Reading Performance of the non-Reading First Schools

Fall 2004 reading performance for students in the non-RF schools represents the best approximation of existing student reading proficiency in each site. This variable draws on test score data from fall 2004,

which is up to 16 months after the RF award date in early award sites, and prior to the RF award date in all late award sites. The percent of students in grades 1-3 at or above grade level variable was constructed using students' fall 2004 SAT 10 scaled scores,<sup>4</sup> as well as the test date at each school. Each student's scaled score was compared to corresponding grade equivalency norms to determine whether the student was at or above grade level. The percent of non-Reading First students at or above grade level was created by taking the mean of the student-level at or above grade level variable, across all grades within a school, and averaging across all schools within a site. To create two subgroups of sites, the study team ordered the sites according to non-RF school performance, taking into account the number of RF schools in each site, and then creating two subgroups as equivalent as possible with respect to the number of schools. This resulted in 120 schools in the high performance group (at or above 48.57 percent of students at grade level) and 128 schools in the low performance group (at or below 48.02 percent of students at grade level).

### **Reading First Funding Per Student**

The amount of the Reading First funding per student was constructed using data from the SEDL database<sup>5</sup> (as of October 2004) about award amounts for each site, and the Common Core Data that provided the number of K-3 students within each school. The Reading First funding per pupil was calculated separately for the 2002-2003 and 2003-2004 school years. Since a portion of the funds were set aside for district and state Reading First activities, and therefore not used to directly fund schools, the grant amounts were adjusted to reflect the proportion of funding set aside (3.5 percent of the Reading First grant to each district). The award amount was then divided by the number of students in grades K-3 in all Reading First schools per site. The Reading First funding per pupil for the two school years was then averaged by site to create the Reading First per pupil expenditure variable used in analysis. To create two subgroups of sites, the study team ordered the sites according to funding, taking into account the number of RF schools in each site, and then creating two subgroups as equivalent as possible with respect to number of schools. This resulted in 126 schools in the high funding (sites with a per-pupil Reading First grant amount at or above \$635), and 122 schools in the low funding group (sites with a per-pupil Reading First grant amount at or below \$513).

For each of the three site characteristics described above, study sites were ordered and then separated into two subgroups of sites that were as balanced as possible, with respect to the number schools in each group. For each characteristic, the order of sites was slightly different. Therefore, the composition of the two subgroups for each moderating factor differed both in the actual sites included and in the total number of schools included. Program impacts were then estimated for one key outcome measure from each of the three domains for the two subgroups. These outcomes included (a) the SAT 10 scaled score for reading comprehension and (b) total minutes in the five dimensions of reading instruction. First, analyses tested the difference between impacts for the two subgroups. Then, to test whether the conclusions were sensitive to the specific cut-point chosen to define the subgroups, average impacts were re-estimated for each subgroup after dropping the two sites closest to the cut-point between the two

---

<sup>4</sup> In the fall of 2004, students' SAT 10 scores were unavailable. For those sites scores from the spring of 2005 were substituted and adjusted by the mean difference of all other students' spring and fall SAT 10 scores, by grade.

<sup>5</sup> Southwest Educational Development Laboratory (SEDL) was contracted to maintain the Reading First Awards database available online at <http://www.sedl.org/readingfirst/welcome.html>. SEDL lists the amount awarded to each Reading First district in the first year. State Reading First Coordinators are responsible for providing this information to SEDL.

subgroups. This was repeated again after dropping the next two sites closest to the cut-point between the two subgroups.

As reported in the study's Interim Report, over the first two of the study's three data collection years (2004-05 and 2005-06), there were statistically significant differences in impacts between early and late award sites for some outcome variables, but not others. There were not statistically significant differences in estimated impacts between early and late award sites for classroom instruction or student engagement with print. However, there were statistically significant differences in impacts between early and late award sites on reading comprehension for grades two and three (but not for grade one). For more detailed information, see Appendix H of the Interim Report (Gamse, Bloom, Kemple & Jacob, 2008). The final set of exploratory analyses presented below provide a follow up to Interim Report analyses conducted to test possible explanations for the award date differences found in the earlier report (see Exhibits G.8 - G.13).

- The relationship between the numbers of months of access to funding and impact on student achievement are not statistically significant in any grade. The relationship between months of funding access and impact on instruction in the five dimensions of reading was not significant in grade 1 and was significant in grade 2.
- The relationship between fall 2004 reading achievement of students in non-Reading First schools and impacts on student achievement or reading instruction are not statistically significant.
- The results indicate that sites with higher allocations of Reading First funds per K-3 student had larger program impacts on student achievement than did sites with lower allocations. This relationship was statistically significant for grades one and two. The relationship between funding and impacts on reading instruction was not statistically significant.

---

**Exhibit G.8: Estimated Impacts on Reading Comprehension, by Award Status**


---

SAT 10 Scaled Scores		Full	Drop 1 Pair	Drop 2 Pairs
<b>Early Award Sites</b>				
Grade 1	Impact	2.87	2.52	6.37
	SE	5.01	5.23	5.39
	p-value	(0.567)	(0.630)	(0.237)
Grade 2	Impact	-4.36	-5.37	-3.93
	SE	4.05	4.25	4.02
	p-value	(0.282)	(0.207)	(0.328)
Grade 3	Impact	-3.89	-5.09	-3.03
	SE	4.09	4.30	4.11
	p-value	(0.342)	(0.236)	(0.462)
<b>Late Award Sites</b>				
Grade 1	Impact	5.64	5.08	2.24
	SE	2.98	3.20	3.58
	p-value	(0.058)	(0.112)	(0.531)
Grade 2	Impact	6.05*	5.00	5.72
	SE	2.59	2.75	3.32
	p-value	(0.019)	(0.070)	(0.085)
Grade 3	Impact	3.52	2.75	2.75
	SE	2.17	2.25	2.47
	p-value	(0.105)	(0.221)	(0.265)
<b>Difference</b>				
Grade 1	Impact	-2.77	-2.57	4.12
	SE	5.83	6.13	6.47
	p-value	(0.636)	(0.676)	(0.525)
Grade 2	Impact	-10.41*	-10.36*	-9.66
	SE	4.81	5.07	5.21
	p-value	(0.032)	(0.042)	(0.066)
Grade 3	Impact	-7.41	-7.85	-5.78
	SE	4.63	4.85	4.80
	p-value	(0.111)	(0.108)	(0.231)

**Notes:**

The complete RF study 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 scores 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. There are 8 late award sites totaling 137 schools and 10 early award sites totaling 111 schools.

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.

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

**EXHIBIT READS: The impact of the Reading First program in early award sites for grade 1 on reading comprehension was**

**2.87 scaled score points, on average, for the full sample of 10 early award sites. The impact was not statistically significant ( $p=.567$ ). The impact of the Reading First program in early award sites for grade 1 on reading comprehension was 2.52 scaled score points, on average, for the sample of 9 early award sites remaining after one pair of sites closest to the cut-point was dropped. The impact was not statistically significant ( $p=.630$ ). The impact of the Reading First program in early award sites for grade 1 on reading comprehension scaled score was 6.37 scaled score points, on average, for the sample of 8 early award sites remaining after two pairs of sites closest to the cut-point were dropped. The impact was not statistically significant ( $p=.237$ ).**

*Sources: RFIS SAT 10 administration in the spring of 2005, 2006 and 2007, as well as from state/district education agencies in those sites that already use the SAT 10 for their standardized testing (i.e., FL, KS, MD, OR).*

---

---

**Exhibit G.9: Estimated Impacts on Reading Instruction, by Award Status**


---

Minutes in Five Dimensions		Full	Drop 1 Pair	Drop 2 Pairs
<b>Early Award Sites</b>				
Grade 1	Impact	2.02	0.82	0.02
	SE	4.30	4.52	4.76
	p-value	(0.640)	(0.856)	(0.996)
Grade 2	Impact	5.55	5.45	6.28
	SE	4.52	4.77	5.07
	p-value	(0.223)	(0.256)	(0.219)
<b>Late Award Sites</b>				
Grade 1	Impact	10.74*	9.65*	6.70
	SE	2.85	3.07	3.40
	p-value	(<0.001)	(0.002)	(0.052)
Grade 2	Impact	13.08*	11.25*	9.99*
	SE	2.73	2.83	3.27
	p-value	(<0.001)	(<0.001)	(0.003)
<b>Difference</b>				
Grade 1	Impact	-8.72	-8.83	-6.68
	SE	5.16	5.46	5.85
	p-value	(0.092)	(0.107)	(0.256)
Grade 2	Impact	-7.53	-5.79	-3.71
	SE	5.28	5.55	6.03
	p-value	(0.155)	(0.297)	(0.540)

**NOTES:**

The complete RF study 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. There are 8 late award sites totaling 137 schools and 10 early award sites totaling 111 schools.

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

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

**EXHIBIT READS:** The impact of the Reading First program in early award sites for grade 1 on the number of minutes of instruction in the five dimensions was 2.02 minutes on average for the full sample of 10 early award sites. The impact was not statistically significant ( $p=.640$ ). The impact of the Reading First program in early award sites for grade 1 on the number of minutes of instruction in the five dimensions was 0.82 minutes on average for the sample of 9 early award sites remaining after one pair of sites closest to the cut-point was dropped. The impact was not statistically significant ( $p=.856$ ). The impact of the Reading First program in early award sites for grade 1 on the number of minutes of instruction in the five dimensions was 0.02 minutes on average for the sample of 80 early award sites remaining after two pairs of sites closest to the cut-point were dropped. The impact was not statistically significant ( $p=.996$ ).

*SOURCE: RFIS Instructional Practice in Reading Inventory, spring 2005, fall 2005, spring 2006, fall 2006, and spring 2007.*

---

**Exhibit G.10: Estimated Impacts on Reading Comprehension, by Fall 2004 Reading Performance of the non-Reading First Schools**

SAT 10 Scaled Scores		Full	Drop 1 Pair	Drop 2 Pairs
<b>High Non-RF School Performance</b>				
Grade 1	Impact	8.26	6.44	9.93*
	SE	4.84	4.00	3.91
	p-value	(0.088)	(0.108)	(0.011)
Grade 2	Impact	-1.85	0.73	2.69
	SE	3.72	3.04	2.72
	p-value	(0.619)	(0.810)	(0.324)
Grade 3	Impact	-0.81	0.20	2.61
	SE	3.73	2.98	2.76
	p-value	(0.828)	(0.946)	(0.344)
<b>Low Non-RF School Performance</b>				
Grade 1	Impact	0.98	2.64	3.27
	SE	3.03	3.32	3.35
	p-value	(0.747)	(0.428)	(0.331)
Grade 2	Impact	5.13	5.54	5.27
	SE	2.76	3.08	3.14
	p-value	(0.063)	(0.075)	(0.096)
Grade 3	Impact	1.21	1.16	3.01
	SE	2.35	2.70	2.75
	p-value	(0.607)	(0.670)	(0.277)
<b>Difference</b>				
Grade 1	Impact	-7.28	-3.79	-6.66
	SE	5.71	5.20	5.15
	p-value	(0.204)	(0.467)	(0.198)
Grade 2	Impact	6.98	4.81	2.59
	SE	4.63	4.33	4.16
	p-value	(0.133)	(0.268)	(0.535)
Grade 3	Impact	2.02	0.95	0.40
	SE	4.41	4.02	3.90
	p-value	(0.648)	(0.813)	(0.919)

**NOTES:**

The complete RF study 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 scores 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. There are 10 high non-RF comparison school sites totaling 120 schools and 8 low performance non-RF school sites totaling 128 schools.

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.

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

**EXHIBIT READS:** The impact of the Reading First program in high performance non-RF school sites for grade 1 on reading comprehension was 8.26 scaled score points on average for the full sample of 10 high non-RF comparison school sites. The impact was not statistically significant ( $p=.088$ ). The impact of the Reading First program in high performance non-RF school sites for grade 1 on reading comprehension was 6.44 scaled score points on average for the sample of 9 high non-RF comparison school sites remaining after one pair of sites closest to the cut-point was dropped. The impact was not statistically significant ( $p=.108$ ). The impact of the Reading First program in high performance non-RF school sites for grade 1 on average reading comprehension scaled score was 9.93 scaled score points on average for the sample of 8 high non-RF comparison school sites remaining after two pairs of sites closest to the cut-point were dropped. The impact was statistically significant ( $p=.011$ ).

*SOURCES: RFIS SAT 10 administration in the spring of 2005, 2006 and 2007, as well as from state/district education agencies in those sites that already use the SAT 10 for their standardized testing (i.e., FL, KS, MD, OR).*

**Exhibit G.11: Estimated Impacts on Reading Instruction, by Fall 2004 Reading Performance of the Non-Reading First Schools**

Minutes in Five Dimensions		Full	Drop 1 Pair	Drop 2 Pairs
<b>High Non-RF School Performance</b>				
Grade 1	Impact	9.76*	12.96*	13.60*
	SE	3.93	3.59	3.64
	p-value	(0.015)	(0.001)	(<0.001)
Grade 2	Impact	10.14*	13.16*	14.84*
	SE	3.87	3.41	3.45
	p-value	(0.010)	(<0.001)	(<0.001)
<b>Low Non-RF School Performance</b>				
Grade 1	Impact	4.18	5.23	6.14
	SE	3.05	3.29	3.32
	p-value	(0.173)	(0.115)	(0.068)
Grade 2	Impact	9.36*	8.65*	9.86*
	SE	3.12	3.47	3.55
	p-value	(0.003)	(0.015)	(0.007)
<b>Difference</b>				
Grade 1	Impact	-5.57	-7.73	-7.46
	SE	4.97	4.87	4.92
	p-value	(0.264)	(0.114)	(0.131)
Grade 2	Impact	-0.78	-4.51	-4.98
	SE	4.98	4.87	4.95
	p-value	(0.875)	(0.355)	(0.316)

**NOTES:**

The complete RF study 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. There are 10 high performance non-RF school sites totaling 120 schools, and 8 low performance non-RF school sites totaling 128 schools.

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

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

**EXHIBIT READS:** The impact of the Reading First program in high performance non-RF school sites for grade 1 on the number of minutes of instruction in the five dimensions was 9.76 minutes on average for the full sample of 10 high non-RF comparison school sites. The impact was statistically significant ( $p=.015$ ). The impact of the Reading First program in high performance non-RF school sites for grade 1 on the number of minutes of instruction in the five dimensions was 12.96 minutes on average for the sample of 9 high non-RF comparison school sites remaining after one pair of sites closest to the cut-point was dropped. The impact was statistically significant ( $p=.001$ ). The impact of the Reading First program in high performance non-RF school sites for grade 1 on the number of minutes of instruction in the five dimensions was 13.60 minutes on average for the sample of 8 high non-RF comparison school sites remaining after two pairs of sites closest to the cut-point were dropped. The impact was statistically significant ( $p<.001$ ).

*SOURCE: RFIS Instructional Practice in Reading Inventory, spring 2005, fall 2005, spring 2006, fall 2006, and spring 2007.*

**Exhibit G.12: Estimated Impacts on Reading Comprehension, by Reading First Funds Per Student**

SAT 10 Scaled Score		Full	Drop 1 Pair	Drop 2 Pairs
<b>Low RF Funding</b>				
Grade 1	Impact	0.05	-1.60	-3.22
	SE	3.87	3.87	3.96
	p-value	(0.990)	(0.680)	(0.416)
Grade 2	Impact	-4.42	-2.33	-3.19
	SE	3.12	3.08	3.28
	p-value	(0.156)	(0.449)	(0.329)
Grade 3	Impact	-2.53	-0.80	-1.89
	SE	3.03	2.97	3.10
	p-value	(0.404)	(0.788)	(0.543)
<b>High RF Funding</b>				
Grade 1	Impact	8.76*	8.95*	5.92
	SE	4.22	4.30	4.73
	p-value	(0.038)	(0.037)	(0.211)
Grade 2	Impact	9.11*	10.06*	9.98*
	SE	3.97	3.96	4.47
	p-value	(0.022)	(0.011)	(0.026)
Grade 3	Impact	0.77	0.72	-0.92
	SE	3.06	2.97	3.27
	p-value	(0.800)	(0.808)	(0.779)
<b>Difference</b>				
Grade 1	Impact	-8.71	-10.54	-9.14
	SE	5.72	5.78	6.17
	p-value	(0.130)	(0.070)	(0.141)
Grade 2	Impact	-13.53*	-12.40*	-13.17*
	SE	5.05	5.02	5.54
	p-value	(0.008)	(0.015)	(0.019)
Grade 3	Impact	-3.30	-1.52	-0.97
	SE	4.30	4.20	4.51
	p-value	(0.444)	(0.718)	(0.830)

**NOTES:**

The complete RF study 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 scores 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. There are 8 low RF funding sites totaling 126 schools and 10 high RF funding sites totaling 122 schools. 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.

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

**EXHIBIT READS:** The impact of the Reading First program in low Reading First funding sites for grade 1 on reading comprehension was 0.05 scaled score points on average for the full sample of 8 low RF funding sites. The impact was not statistically significant ( $p=.990$ ). The impact of the Reading First program in low Reading First funding sites for grade 1 reading comprehension scaled score was -1.60 scaled score points on average for the sample of 7 low RF funding sites remaining after one pair of sites closest to the cut-point was dropped. The impact was not statistically significant ( $p=.680$ ). The impact of the Reading First program in low Reading First funding sites for grade 1 on average reading comprehension scaled score was -3.22 scaled score points on average for the sample of 6 low RF funding sites remaining after two pairs of sites closest to the cut-point were dropped. The impact was not statistically significant ( $p=.416$ ).

*SOURCES: RFIS SAT 10 administration in the spring of 2005, 2006 and 2007, as well as from state/district education agencies in those sites that already use the SAT 10 for their standardized testing (i.e., FL, KS, MD, OR).*

**Exhibit G.13: Estimated Impacts on Reading Instruction, by Reading First Funds Per Student**

Minutes in Five Dimensions		Full	Drop 1 Pair	Drop 2 Pairs
<b>Low RF Funding</b>				
Grade 1	Impact	3.28	5.82	1.39
	SE	3.56	3.26	3.77
	p-value	(0.359)	(0.077)	(0.714)
Grade 2	Impact	4.86	7.23*	4.51
	SE	3.57	3.15	3.82
	p-value	(0.177)	(0.024)	(0.241)
<b>High RF Funding</b>				
Grade 1	Impact	10.66*	9.35*	10.15*
	SE	3.33	3.70	3.73
	p-value	(0.002)	(0.013)	(0.008)
Grade 2	Impact	14.95*	13.16*	12.36*
	SE	3.37	3.66	3.75
	p-value	(<.001)	(<.001)	(0.001)
<b>Difference</b>				
Grade 1	Impact	-7.38	-3.52	-8.76
	SE	4.88	4.94	5.31
	p-value	(0.132)	(0.476)	(0.101)
Grade 2	Impact	-10.10*	-5.93	-7.84
	SE	4.91	4.83	5.36
	p-value	(0.041)	(0.221)	(0.145)

**NOTES:**

The complete RF study 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. There are 8 low RF funding sites totaling 126 schools and 10 high RF funding sites totaling 122 schools.

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

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

**EXHIBIT READS:** The impact of the Reading First program in low Reading First funding sites for grade 1 on the number of minutes of instruction in the five dimensions was 3.28 minutes on average for the full sample of 8 low RF funding sites. The impact was not statistically significant ( $p=.359$ ). The impact of the Reading First program in low Reading First funding sites for grade 1 on the number of minutes of instruction in the five dimensions was 5.82 minutes on average for the sample of 7 low RF funding sites remaining after one pair of sites closest to the cut-point was dropped. The impact was not statistically significant ( $p=.077$ ). The impact of the Reading First program in low Reading First funding sites for grade 1 on the number of minutes of instruction in the five dimensions was 1.39 minutes on average for the sample of 6 low RF funding sites remaining after two pairs of sites closest to the cut-point were dropped. The impact was not statistically significant ( $p=.714$ ).

*SOURCES: RFIS Instructional Practice in Reading Inventory, spring 2005, fall 2005, spring 2006, fall 2006, and spring 2007.*

## References

- Abt Associates and RMC Research (2002). Classroom Observation Record. Unpublished instrument.
- Armbruster, B.B., Lehr, F., and Osborn, J. (2003). *Put reading first: The research building blocks for teaching children to read* (2nd Ed.). Developed by the Center for the Improvement of Early Reading Achievement (CIERA). Washington, DC: The National Institute for Literacy (NIFL).
- Ball, E.W., and Blachman, B.A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly*, 26, 49–66.
- Beck, I.L., Perfetti, C.A., and McKeown, M.G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74, 506-521.
- Bloom, H., Kemple, J.K., and Gamse, B.C. (2004). Using regression discontinuity analysis to measure the impacts of Reading First. Paper prepared for the Institute of Education Sciences.
- Bloom, H.S. (2001). *Measuring the impacts of whole-school reforms: Methodological lessons from an evaluation of accelerated schools*. New York: MDRC.
- Bloom, H.S. (1995). “Minimum Detectable Effects: A Simple Way to Report the Statistical Power of Experimental Designs,” *Evaluation Review*, Vol. 19, No. 5, pp. 547–556.
- Brennan, R.L. (2001). *Generalizability theory*. New York, NY: Springer-Verlag.
- Brett, A., Rothlein, L., and Hurley, M. (1996). Vocabulary acquisition from listening to stories and explanations of target words. *Elementary School Journal*, 96, 415–422.
- Bus, A.G., and van Ijzendoorn, M.H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, 91, 403–414.
- Cain, G. (1975). Regression and Selection Models to Improve Nonexperimental Comparisons. In C. A. Bennet and A.A. Lumsdaine (Eds.), *Evaluation and Experiment* (pp.297-317). New York: Academic Press. 297-317.
- Cappelleri, J.C.; Trochim, W.M.K.; Stanley, T.D.; Reichardt, C.S. (1991). Random measurement error does not bias the treatment effect estimate in the regression-discontinuity design: I. The case of no interaction. *Evaluation Review*, 15:395-419.
- Carlisle, J., and Scott, S. (2003). Teachers’ Instructional Practice. Unpublished instrument.
- Cook, T.D. (2008). “Waiting for life to arrive:” A history of the regression-discontinuity design in Psychology, Statistics and Economics. *Journal of Econometrics*, 142(2), 636-654.
- Crowe, L.K. (2005). Comparison of two oral reading feedback strategies in improving reading comprehension of school-age children with low reading ability. *Remedial and Special Education*, 26, 32–42.

- CTB/McGraw-Hill. (2003). *Terranova second edition: California Achievement Tests technical report*. Monterey, CA: Author.
- Cullen, J., Jacob, B. and Levitt, S. (2006). The Effect of School Choice on Student Outcomes: Evidence from Randomized Lotteries. *Econometrica*. 74(5): 1191-1230.
- Dixon, L.Q., Smith, W.C., Dwyer, M.C., Peabody, B.K., and Gamse, B.C. (2007, April). Training observers to use the Instructional Practice in Reading Inventory. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL
- Dole, E., Nelson, D., Fox, D., and Gardner, J. (2001). Utah's Profile of Scientifically based Reading Instruction. Salt Lake City, Utah: The Institute for Behavioral Research in Creativity.
- Dwyer, M.C., Smith, W.C., Dixon, L.Q., and Gamse, B.C. (2007, April). The development of the Instructional Practice in Reading Inventory. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.
- Edmonds, M.S., and Briggs, K.L. (2003). The instructional content emphasis instrument: Observations of reading instruction. In S. Vaughn and K. L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 31–52). Baltimore: Paul Brookes.
- Foorman, B.R., Francis, D.J., Fletcher, J.M., Schatschneider, C. and Mehta, P. (1998). “The Role of Instruction in Learning to Read: Preventing Reading Failure in At-Risk Children”, *Journal of Educational Psychology*, Vol. 90, No. 1, pp. 37-55.
- Foorman, B.R., and Torgesen, J. (2001). Critical Elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice*, 16(4), 203–212.
- 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.
- Foorman, B.R., Schatschneider, C., Eakin, M.N., Fletcher, J.M., Moats, L.C., & Francis, D.J. (2006). The impact of instructional practices in grades 1 and 2 on reading and spelling achievement in high poverty schools. *Contemporary Educational Psychology*, 31, 1-29.
- Gamse, B.C., Bloom, H., Celebuski, C., Dwyer, M.C., Gersten, R., Leos-Urbel, J., et al. (2004). *Reading First Impact Study: Revised study design*. Cambridge, MA: Abt Associates Inc.
- 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.
- Gelman, A. and Stern, H. (2006). The difference between “Significant” and “Not Significant” is not itself statistically significant. *The American Statistician*, 60(4), 328—331.
- Goldberger, A.S. (1972). *Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations*. Institute for Research on Poverty: Madison, WI, Discussion Paper 123.

- Goodson, B., Layzer, J., Smith, C., and Rimdzius, T. (2004). *Observation Measure of Language and Literacy Instruction (OMLIT)*. Developed as part of the Even Start Classroom Literacy Interventions and Outcomes (CLIO) Study, under contract ED-01-CO-0120 as administered by the Institute of Education Sciences, U.S. Department of Education.
- Graves, A.W., Gerston, R., and Haager, D. (2004). Literacy instruction in multiple-language first-grade classrooms: Linking student outcomes to observed instructional practice. *Learning Disabilities Research & Practice, 19*(4), 262–272.
- Gunn, B., Smolkowski, K., Biglan, A., and Black, C. (2002). Supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school: A follow-up study. *Journal of Special Education, 36*(2), 69–79.
- Haager, D., Gersten, R., Baker, S., and Graves, A. (2003). The English-Language Learner Classroom Observation Instrument: Observations of beginning reading instruction in urban schools. In S. R. Vaughn and K. L. Briggs (Eds.), *Reading in the Classroom: Systems for Observing Teaching and Learning*. Baltimore: Paul Brookes.
- Hahn, H., Todd, P., and van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica, 69*(3), 201–209.
- Harcourt Assessment, Inc. (2004). *Stanford Achievement Test series, Tenth Edition technical data report*. San Antonio, TX: Author.
- Hatcher, P. J., Hulme, C. & Snowling, M.J. (2004). Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *Journal of Child Psychology & Psychiatry, 45*, 338-358.
- Hedges, L.V., and Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Hernandez, A.V., Boersma, E., Murray, G.D., Habbema, J. D, Steyerberg, E.W. (2006). Subgroup analyses in therapeutic cardiovascular clinical trials: Are most of them misleading? *American Heart Journal. 151*(2):257-264.
- Hoover, H.D., Dunbar, S. B., Frisbie, D. A., Oberley, K. R., Ordman, V. L., Naylor, R. J., Lewis, J.C., Qualls, A. L., Mengeling, M.A., and Shannon, G.P. (2003). The Iowa Tests guide to research and development, Forms A and B. Riverside Publishing, Itasca, IL.
- Kamil, M.L. (2004). Vocabulary and comprehension instruction: Summary and implications of national reading panel findings. In P. McCardle and V. Chhabra (Eds.), *The voice of evidence in reading research*. Baltimore: Paul Brookes.
- Kling, J.R., Liebman, J.B., and Katz, L.F. (2007). Experimental analysis of neighborhood effects. *Econometrica, 75*(1), 83–119.
- Logan, A.C. and Tamhane, B. (2002). Accurate critical constants for the one-sided approximate likelihood ratio test of a normal mean vector when the covariance matrix is estimated. *Biometrics, 58*(3), 650–656.

- MacGinitie, W. H., MacGinitie, R. K., Maria, K., and Dreyer, L. G. (2000). *Gates-MacGinitie Reading Tests—Manual for scoring and interpretation*. Itasca, IL: Riverside.
- Mason, L.H. (2004). Explicit self-regulated strategy development versus reciprocal questioning: Effects on expository reading comprehension among struggling readers. *Journal of Educational Psychology*, *96*, 283–296.
- 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.
- McCutchen, D., Abbott, R.D., Green, L.B., Beretvas, S. N., Cox, S., Potter, N.S., Quiroga, T., & Gray, A. L. (2002). Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities*, *35*, 69-86.
- McKeown, M.G., Beck, I.L., and Omanson, R.C. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly*, *20*, 522–535.
- McKeown, M.G., Beck, I.L., and Omanson, R.C., and Perfetti, C.A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Reading Behavior*, *15*, 3-18.
- Mohr, L.B. (1995). *Impact Analysis for Program Evaluation*. Second Edition. Thousand Oaks, CA: Sage Publications.
- 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 of 2001, ESEA, 2001, Title 1, Part B, Subpart 1.
- O'Brien, P.C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics*, *40*(4), 1079–1087.
- O'Connor, R.E., Bell, K.M., Harty, K.R., Larkin, L.K., Sackor, S.M., and Zigmond, N. (2002). Teaching reading to poor readers in the intermediate grades: A comparison of text difficulty. *Journal of Educational Psychology*, *94*, 474–485.
- Rasinski, T. and Oswald, R. (2005). Making and writing words: Constructivist word learning in a second-grade classroom. *Reading & Writing Quarterly*, *21*, 151–163.
- Reading First Advisory Committee. (2008) *Advisory Committee's Response to the Reading First Impact Study Interim Report*, retrieved 10/27/2008 from U.S. Department of Education website: <http://www.ed.gov/programs/readingfirst/statement.pdf>.
- Reichardt, C.S., Trochim, W.M.K., and Cappelleri, J.C. (1995). Reports of the death of regression-discontinuity analysis are greatly exaggerated. *Evaluation Review*, *19*:39-63.
- Reutzel, D.R. and Hollingsworth, P.M. (1993). Effects of fluency training on second graders' reading comprehension. *Journal of Educational Research*, *86*, 325-331.

- Reutzel, D.R. and Hollingsworth, P.M. (1991). Using literature webbing for books with predictable narrative: Improving young readers' prediction, comprehension, and story structure knowledge. *Reading Psychology, 12*(4), 319–333.
- Rosenshine, B., Meister, C., and Chapman, S. (1996). Teaching students to generate questions: A review of the intervention strategies. *Review of Educational Research, 66*(2), 181–221.
- Rothwell, P.M (2005). Subgroup analysis in randomized controlled trials: importance, indications, and interpretation. *Lancet, 365*:176-86.
- Schochet, Peter Z. (2008). *Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations* (NCEE 2008-4018). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Shadish, W.R., Cook, T.D., and Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shaffer, J.P. (1995). Multiple hypothesis testing. *Annual Review of Psychology, 46*, 561-84.
- Shavelson, R.J., & Webb, N.M. (1991). *Generalizability Theory: A Primer*. Newbury Park, CA: Sage Publications.
- Snow, C.E., Burns, M.S., and Griffin, P. (Eds.) (1998). Preventing reading difficulties in young children (Report of the Committee on the Prevention of Reading Difficulties in Young Children, National Research Council). Washington, DC: National Academies Press.
- Southwest Educational Development Laboratory. Summary Report: Awards. Retrieved April 20, 2007 from <http://www.sedl.org/readingfirst/report-awards.html>.
- Stahl, S.A. (2004). What do we know about fluency? Findings of the National Reading Panel. In P. McCardle and V. Chhabra (Eds.), *The Voice of Evidence in Reading Research* (pp. 187–210). Baltimore: Paul Brookes.
- Suen, H.K., and Ary, D. (1989). Reliability: Conventional methods. In H. K. Suen and D. Ary (Eds.), *Analyzing Quantitative behavioral observation data* (pp. 99–129). Hillsdale, NJ: Lawrence Erlbaum.
- Taylor, L.K., Alber, S.R., and Walker, D.W. (2002). The comparative effects of a modified self-questioning strategy and story mapping on the reading comprehension of elementary students with learning disabilities. *Journal of Behavioral Education, 11*(2), 69–87.
- Therrien, W.J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education, 25*, 252–261.
- Thistlethwaite, D.L. and Campbell, D. T. (1960). Regression discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology, 51*(6): 309–317.
- Tomesen, M. and Aarnoutse, C. (1998). Effects of an instructional program for deriving word meanings. *Educational Studies, 24*, 107–128.

- Torgesen, J.K., Wagner, R.K., Rashotte, C.A., Rose, E., Lindamood, P., Conway, T., et al. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91*, 579–593.
- Trochim, W.M.K. (1990). The Regression Discontinuity Design. In L. Sechrest, P. Perrin and J. Bunker (Eds.), *Research methodology: Strengthening causal interpretations of nonexperimental data* (pp. 119-130). Rockville, MD: Agency for Health Care Policy and Research, U.S. Public Health Service.
- U.S. Department of Education, Office of Elementary and Secondary Education. (2002). *Guidance for the Reading First Program*. Washington, DC.
- U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. (2006) *Reading First Implementation Evaluation: Interim Report*, Washington, D.C.
- Vaughn, S. and Briggs, K.L. (2003). *Reading in the classroom: System for the observation of teaching and learning*. Baltimore: Brookes Publishing.
- Wang, R., Lagakos, S.W., Ware, J.H., Hunter, D.J., & Drazen, J.M. (2007). Statistics in medicine—Reporting of subgroup analyses in clinical trials. *New England Journal of Medicine, 357*, 2189-2194.
- Williams, K.T. (2001). *Group Reading Assessment and Diagnostic Evaluation: Technical Manual*. Circle Pines, MN: American Guidance Service, Inc.